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.

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.

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 CO_2 growth rate variation and trend.
- We have summarised the overview of global gridded anthropogenic CO_2 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 CO_2 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

• 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.

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.

• *Why the EDGAR sector uncertainty is not purely additive? Please expand on the exemptions*.

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.

• *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?*

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 CO_2. 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.

• *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)*.

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)[1].

3 Technical corrections

• *line 41: Since the early 2002s -> 2000s*

Corrected.

• *line 86: Presence of observations may should better say availability of observations and emission information*.

Corrected.

• *line 150: lower case S in Savannah*

Corrected.

• *line 159: What is an autoproducer? Is this an automobile manufacturer?*

Autoproducers is the energy generated and used specifically for industrial purposes and manufacturing. We added this explanation to the main text.

• *line 235: You repeat "per activity" several times*

Corrected.

• *Table 3 and table 7- why are lower bounds with larger uncertainty than upper bounds*

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.

• *Better description of ensuring log-normal distribution*

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.

---

[1] https://www.nature.com/articles/s41597-020-0462-2

• *Table S5: why '*', which indicates for residential sector only according to the table caption, on fuel types aviation fuel, motor gasoline, etc*?

Unfortunate misprint. Corrected.

• *I find too many acronyms difficult to follow, make text confusing: AD, NIR, TFI, EF, LDS, WDS, GLB, L, U etc*.

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.

• *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*.

We complemented the main text with an explanation and adopted a more transparent background colours for the figures.

• *Indenting or centering of equations to distinguish them better from normal text*.

Unfortunately, ESSD template does not allow changes to the current format of the equations.

• *Could section 4.1 be largely substituted by a table and map*?

We have substituted this section by Table 9.

• *Please consider adding section S3 to main text as it makes the log-normal distribution more clear*.

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

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.

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?

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 $kg \cdot m^{-2} \cdot s^{-1}$ 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?

The only source of internationally accepted anthropogenic CO_2 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 development 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.

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

We added a description of other global CO_2 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)[2].

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.

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.

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.

---

[2] https://arxiv.org/abs/1910.11727

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"?

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?

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 CO_2 point sources, generating CO_2 plumes that can be directly observed by in-situ or space borne measurements and these CO_2 "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?

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?

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"?

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.

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.

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.

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?

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.

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.

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.

An extra column has been added to the table that includes the main references Andrew (2020)[3], 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.

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 0.1◦×0.1◦ 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.

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 CO_2 emissions at 0.1◦×0.1◦ 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 CO_2 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.

---

[3] Andrew, R.M: A comparison of estimates of global carbon dioxide emissions from fossil carbon sources, Earth Syst. Sci. Data, 12, 1437–1465, https://doi.org/10.5194/essd-12-1437-2020, 2020.

This has been corrected and the big folder has been split in more clear subfolders. As such, a new Zenodo link is introduced.

<u>*Specific comments from the first few pages*</u>

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"?

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.

This has been rephrased as "*For an increased understanding of the CO_2 emission sources, patterns and trends, a link between the emission inventories and observed CO_2 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.

This has been rephrased, avoiding the term "prior".

L15 Are power and energy different? If not, the same word should be used.

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.

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.

These CO_2 emissions and uncertainties have been used in the ECMWF IFS CO_2 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 CO_2 emissions and uncertainties will be used by the CAMS inversion system (currently under development).

L21 How large are these changes to EDGAR emissions?

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 CO_2 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 describe the sensitivity tests a bit more.

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".

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.

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"?

This has been corrected.

L43-8. Sentence needs to be revised or deleted.

This has been revised.

L51 Observation not Observatory
This has been corrected.

L61 emissions not concentrations
This has been corrected.

L63 What is the Mitchell 1984 reference, and why is it cited when referring to the year 2018?
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
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.
Table 1 was updated with more precise wording.

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

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

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

L80 delete "with long carbon cycle"
This has been deleted.

L81-3 Uncertainties on a 0.1degree grid? What about your revised estimates of emissions?
National sectoral uncertainties and revised emissions are both uniformly mapped onto a regular latitude/longitude 0.1°×0.1° resolution grid.

L86, 90 Incomplete sentences, should start with "the"
Theses have been corrected.

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

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

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

L165-6 Needs reference
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.

Global anthropogenic CO_2 emissions and uncertainties as prior for Earth system modelling and data assimilation

Margarita Choulga[1], Greet Janssens-Maenhout[2], Ingrid Super[3], Efisio Solazzo[2], Anna Agusti-Panareda[1], Gianpaolo Balsamo[1], Nicolas Bousserez[1], Monica Crippa[2], Hugo Denier van der Gon[3], Richard Engelen[1], Diego Guizzardi[2], Jeroen Kuenen[3], Joe McNorton[1], Gabriel Oreggioni[2], ~~Efisio Solazzo[2],~~ and Antoon Visschedijk[3]

[1]Research Department, ECMWF, Reading, RG2 9AX, United Kingdom
~~[2]Joint~~[2]European Commission, Joint Research Centre~~ of the European Commission, EC~~, JRC, Ispra, 21027, Italy
[3]TNO, Department of Climate, Air and Sustainability, Utrecht, 3584 CB, The Netherlands

Correspondence to: Margarita Choulga (margarita.choulga@ecmwf.int)

Abstract. ~~Anthropogenic~~The growth in anthropogenic carbon dioxide (CO_2) emissions ~~and their observed growing trends raise awareness in~~ acts as a major climate-change driver, which has widespread implications across society, influencing the scientific, political and public sectors~~ of the society as the major driver of climate-change~~. For an increased understanding of the CO_2 emission sources, patterns and trends, a link between the emission inventories and observed CO_2 concentrations is best established via Earth system modelling and data assimilation. In this study anthropogenic CO_2 emission inventories are processed into gridded maps to provide an estimate of ~~prior~~ CO_2 emissions for 7 main emissions groups: 1) ~~power generation~~ energy production super-emitters~~ and~~, 2) energy production ~~average~~standard-emitters, 3) manufacturing, 4) settlements, 5) aviation, 6) other transport at ground level and 7) others, with estimation of their uncertainty and covariance to be included in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS). The emission inventories are sourced from the Intergovernmental Panel on Climate Change (IPCC) 2006 Guidelines for National Greenhouse Gas Inventories and revised information from its 2019 Refinements, and the global grid-maps of Emissions Database for Global Atmospheric Research (EDGAR) inventory. The anthropogenic CO_2 emissions for 2012 and 2015, (EDGAR versions 4.3.2 and 4.3.2_FT2015 respectively) are considered, updated with improved apportionment of the energy sector~~,~~ (decreased by 8 %) and the energy usage for manufacturing ~~and~~(increased by 18 %), and with newly generated diffusive CO_2 emissions from coal mines. These emissions aggregated into 7 ECMWF groups with their emission uncertainties are calculated per country (considering its statistical infrastructure development level) and sector (considering the most typical fuel type) and use the IPCC recommended error propagation method assuming fully uncorrelated emissions ~~to generate covariance matrices of parsimonious dimension (7×7).~~. While the uncertainty of most groups remains relatively small~~,~~ (5-20 %), the largest contribution (usually over 40 %) to the total uncertainty is determined by the OTHER group with usually the smallest budget, consisting of oil refineries and transformation industry, fuel exploitation, coal production, agricultural soils and solvents and products use emissions~~.~~, with uncertainties more than 100 %. Several sensitivity studies are performed: 1) for country type ~~(with~~ – by analysing the impact of assuming either a well~~/~~ or less~~-~~well-developed

statistical infrastructure~~),~~, 2) for fuel type specification – by adding explicit information for each fuel type used per each IPCC activity, and 3) for national emission source distribution ~~(~~– by aggregating all emission sources and evenly redistributing them over the country – highlights the importance of ~~accurate point source~~spatial mapping~~).~~. Uncertainties are compared with United Nations Framework Convention on Climate Change (UNFCCC) and the Netherlands Organisation for Applied Scientific Research (TNO) data. Upgraded anthropogenic CO_2 emission maps with their yearly and monthly uncertainties are combined into the CHE_EDGAR-ECMWF_2015 dataset (Choulga et al., 2020) available from ~~doi:10.5281/zenodo.3712339.~~https://doi.org/10.5281/zenodo.3967439. CHE_EDGAR-ECMWF_2015 consists of 11 global NetCDF files with gridded yearly and monthly upper and lower bounds of uncertainties in % and $kg \cdot m^{-2} \cdot s^{-1}$ for 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).

1 Introduction

Carbon dioxide (CO_2) is the most abundant greenhouse gas (GHG) (NOAA, 2019) contributing to the ~~Earth's radiative balance and~~ climate ~~stability~~change. This study focuses on anthropogenic (man-made) long-cycle carbon ~~cycle (period between carbon release to and capture from the atmosphere is longer than a year)~~ CO_2 emissions~~,~~ (i.e. emissions from fossil fuel use and industrial processes: cement production, carbonate use of limestone and dolomite, non-energy use of fuels and other combustion, chemical and metal processes, solvents, agricultural liming and urea, waste and fossil fuel fires (Janssens-Maenhout et al., 2019)), that occur on top of an active natural carbon cycle, and generation of a reliable uncertainty band globally for different emission types that can be used in Earth system modelling and data assimilation.

The CO_2 growth rate varies from year to year with a tendency toward higher growth rates since the early ~~2002s~~2000s. The added CO_2 has a long life-time and only a portion of it transfers each year from the atmosphere to the oceans and to vegetation on land. The atmosphere exchanges carbon mainly between: (i) the terrestrial biosphere – ~~is influenced~~impact on growth rate through deforestation and other forms of land management; (ii) the oceans – impact on growth rate through marine ecosystems~~ have~~ implications due to CO_2 in the form of carbonic acid absorption in surface waters and their mix with deep ocean waters; (iii) the fossil fuels and cement and other CO_2 process emissions – when around 1920 fossil fuel burning became the dominant source of anthropogenic emissions to the atmosphere~~ and there is~~, with a clear increase of 91 ppm ~~since 1959 (~~in the past six decades (from 316 ppm~~)~~ in 1959 till ~~2018 (~~407.4 ± 0.1 ppm in 2018), according to NOAA (2019).

Accurate assessment of anthropogenic CO_2 emissions is important to better understand the global carbon cycle. Efforts towards a global anthropogenic CO_2 monitoring and verification support capacity as described by Janssens-Maenhout et al. (2020), rely on atmospheric modelling and atmospheric observations (in-situ from e.g. the Integrated Carbon ~~Observatory~~Observation System, ICOS, air-borne from e.g. aircraft campaigns, or space-borne from e.g. the Orbiting Carbon Observatory, OCO-2, and the Greenhouse gases Observing Satellite, GOSAT). All measurements are assimilated by global tracer transport models to infer atmospheric CO_2 changes or by flux inversion systems to estimate the large-scale surface

CO_2 fluxes. ECMWF for example applies both inverse modelling and direct modelling of global concentrations of CO_2 in the atmosphere assimilating several types of observations.

The global transport models require an initial best estimate of the CO_2 emission fields with uncertainties, the so-called prior information. The intensity of the emission fields is corrected through minimization of the difference between the modelled and measured concentration values for CO_2. The uncertainty of these corrected CO_2 fluxes based on inverse modelling will be lower with the increase of CO_2 observations and its accuracy. The disentanglement of the fossil CO_2 emissions from the total atmospheric CO_2 ~~concentration~~emissions remains challenging, e.g. in 2018 total anthropogenic CO_2 concentrations (42.5 ± 3.3 Gt CO_2) represented only 1.3 % of the global atmospheric CO_2 concentration (407.4 ± 0.1 ppm) (Friedlingstein et al., 2019~~; Mitchell, 1984~~), which states the need for high accuracy of measurements (≥ 1.0 %).

~~Global tracer transport models also require input of emission data, which is~~ Emission fields are often supplied through emission inventories. Bottom-up emission inventories start from human activity statistics and emission factors ~~(EF)~~ are defined for each activity and provided at international or country level (e.g. National greenhouse gas Inventory Report, NIR). Such bottom-up inventories need to be gridded and characterised with uncertainties in order to represent a prior data set useful for numerical modelling. Table 1 shows ~~some~~ examples of most used global gridded CO_2 emission datasets, for more details see Andrew (2020), Janssens-Maenhout et al. (2019, Table 3) and Cong et al. (2018, Table 1).

There is strong evidence to suggest fixed annual emissions do not represent sufficient temporal variability, for example, natural gas consumption has two seasonal peaks, with consumption patterns predominantly driven by weather – the largest peak occurs during winter due to low temperatures when natural gas is used more to heat residential and commercial spaces, the smaller peak occurs during summer due to high temperatures when natural gas, coal or petroleum-fired generators are used to generate more electric power for air conditioning (Bradley, 2015; Comstock, 2020). New version of EDGAR v5.0 is addressing high temporal disaggregation of emissions (Crippa et al., 2020).

Table 1: ~~Some examples~~Examples of global gridded anthropogenic CO_2 emission bottom-up datasets

Name	Resolution	Period	~~Note~~Main assumptions, uncertainties	Source
Carbon Dioxide Information Analysis Center (CDIAC)	Spatial: 1.0º×1.0º Temporal: annual, monthly Sectoral: 1	1751-2013	Use population density to disaggregate emissions, the mass-emissions data based on fossil-fuel consumption estimates. Provide gridded annual and monthly uncertainty estimates for 1950-2013	Andres et al., 1996; Andres et al., 2016
Open-Data Inventory for Anthropogenic Carbon dioxide (ODIAC)	Spatial: 1×1 km^2, 0.1º×0.1º Temporal: monthly Sectoral: 6	1979-2018	First introduced the combined use of nightlight data and individual power plant emission/location profiles	Oda and Maksyutov, 2011; Oda et al. 2018; ODIAC, 2020
Emissions Database for Global Atmospheric Research (EDGAR)	Spatial: 0.1º×0.1º Temporal: annual, monthly Sectoral: 26	1970-(year-1)	Based on international statistics, covers all IPCC (2006) reporting categories, consistent methodology applied to all the world countries	Janssens-Maenhout et al., 2019
Fossil Fuel Data Assimilation System (FFDAS)	Spatial: 0.1º×0.1º Temporal: annual Sectoral: 2	1997-2012	Provide gridded posterior uncertainty (version 2.2); in addition, provide monthly, weekly, and hourly fractions from annual CO_2 emissions	Asefi-Najafabady et al., 2014
Community Emissions Data System (CEDS)	Spatial: 0.1º×0.1º Temporal: annual, monthly	1750-2014	Provide emissions of CO_2 and other GHGs and pollutants	Hoesly et al., 2018

	Sectoral: 55			
Peking University Fuel combustion inventory (PKU-FUEL)	Spatial: 0.1º×0.1º Temporal: monthly Sectoral: 6	1960-2014	By request provide daily emissions and the results of Monte Carlo simulation-based uncertainty analyses	Chen et al., 2016; Liu et al., 2015

~~Global emission budget values from different datasets are never the same, therefore it is important to identify why estimates differ between datasets (e.g. differences in sources and methods used or emission double counting and omissions). Though there are global anthropogenic emission gridded datasets, most of them have scarce evaluation of uncertainties, which needs enhancement with the relative errors for sector-specific country totals and the uncertainties in trends with the appropriate probability density functions. Global gridded uncertainties used in an independent atmospheric inversion method might also increase level of confidence in a certain emission dataset (Andrew, 2020).~~

Though there are global anthropogenic emission gridded datasets, most of them have scarce evaluation of uncertainties, which needs enhancement with the relative errors for sector-specific country totals and the uncertainties in trends with the appropriate probability density functions. Only 3 datasets from Table 1 provide uncertainty estimates, namely CDIAC, FFDAS and PKU-FUEL. CDIAC uncertainties have no sectors and include contributions from the tabular fossil fuel CO_2 emissions (assigned per 7 country types, values are constant over time), geography map (power plant location), and population map (has details both in time and space, is used to distribute fossil fuel CO_2 emissions). Population map uncertainty strongly dominates in the generated gridded fossil fuel CO_2 uncertainties (Andres et al., 2016). CDIAC uncertainties have no sectoral distribution and are presented on 1.0º×1.0º grid. FFDAS provides only posterior uncertainties, which are based on a model inversion. These posterior uncertainties could be used as prior uncertainties for separate inversion systems, however this would make the characterisation of uncertainty more complex if there were similarities in the model and observations used. PKU-FUEL uncertainty estimates of CO_2 emission maps associated with uncertain fuel data and uncertain activity data in the spatial disaggregation process are based on Monte Carlo ensemble simulations. Input data was randomly sampled 1000 times from an a priori normal uncertainty distribution with a certain coefficient of variation: for fuel consumptions from ships/aviation sector coefficient of variation is set to be 20 %, for wildfires sector – 18 %, for all other fuel data – 10 %, for combustion rates – 20 %. Additional coefficient of variation was assigned for each country or subnational unit based on its size to consider uncertainty of spatial fuel data disaggregation (e.g. coefficient of variation for the largest subnational unit of the world Asian part of the Russian Federation is 1000 %). Emission factor coefficient of variation was constant value of 5 % (Wang et al., 2013). PKU-FUEL uncertainties were heavily based on subjective assumptions and rather detailed information of fuel type, which makes is difficult to use for IPCC (2006) reporting categories.

In this study, we focus on fossil emissions (from fossil fuel combustion, use and production, and process emissions from cement production and others such as glass, chemicals, urea) ~~with long carbon cycle~~ and we distinguish between point sources and sources with wider spatial distribution. The scope of this research is to generate a reliable uncertainty band on 0.1º×0.1º grid with global coverage based on emission type for the yearly and monthly emission budgets, that are the composite of anthropogenic fossil fluxes~~.~~, and that are aligned with updated IPCC requirements. Uncertainty characterisation is key for optimally combining the bottom-up inventories with the top-down data assimilation.

In this study 2015 is chosen as a base year to analyse anthropogenic CO_2 budgets (i.e. global, regional, national) from different sources (i.e. global statistics, national reports~~). Main reason for this choice is~~), benefitting the ~~presence~~availability of observations (both in-situ and space-borne~~),~~) as well as reported and ~~that all available information is already~~ verified ~~and reported~~emission inventories. Global CO_2 emissions from fossil fuel and industrial processes such as cement production reached a total of 36.2 Pg CO_2 in 2015 according to EDGAR inventory version 4.3.2_FT2015 (Olivier et al., 2016a). ~~This result shows a stagnation of the fossil emissions growth, also thanks to the curbing of China's emissions. Largest contribution to this global total originates from China (with a 29 % share in the global total), the United States (14 %), the European Union (28 members till end of 2019) (10 %), India (7 %), the Russian Federation (5 %).~~ The use of energy represents by far the largest source of emissions (89 % share globally~~). The~~), and in particular the energy industry sector ~~includes emissions from fuel~~(38 % share) (including both combustion ~~(the large majority, with 38 % share)~~ and fugitive ~~emissions, which are intentional or unintentional~~ gas releases ~~of gas~~ from use but also production, processes, transmission~~,~~ and storage ~~and use~~ of fuels~~. Other sectors manufacturing, transport and buildings show a share of 22 %, 20 %~~ for energy and ~~9 % respectively in 2015. More details are given in Olivier et al. (2016b~~heat generation). Another reason for choosing 2015 is that it's the year of the Paris Agreement and the reference year for several Nationally Determined Contributions (NDCs~~) (most countries in their NDCs also mention years 1990, 2005, 2025 and 2030). On 12th December 2015 at the twenty-first session of the Conference of the Parties to the United Nations Framework Convention on Climate Change the Paris Agreement was agreed, and currently it is ratified by 189 countries (CarbonBrief, 2020; Paris Agreement - Status of Ratification, 2020). It aims to limit the increase in global average temperature to 1.5 ºC, since this would significantly reduce risks and the impacts of climate change (Paris Agreement, 2020).~~). Countries have submitted their pledges to the United Nations (UN), setting out how far they plan to reduce their GHG emissions – NDCs (CarbonBrief, 2020). ~~For example, the European Union's NDC under the Paris Agreement is to reduce GHG emissions by at least 40 % by 2030 compared to 1990 (Paris Agreement, 2020).~~ Yet concentrations are still growing. In 2015, the average concentration of CO_2 (399 ppm) was about 40 % higher than in the mid-1800s, with an average growth of 2 ppm/yr in the last ten years. ~~CO_2 resulting from the oxidation of carbon in fuels during combustion dominates total GHG emissions.~~ Furthermore, according to JRC 2019 Report (Crippa et al., 2019) between 2015 till 2018, just in three years global CO_2 emissions have raised by 4.3 % (1575.2 Mt CO_2/yr), ~~while~~of which the international component of CO_2 emissions (shipping and aviation ~~CO_2 emissions have~~bunker fuel) has even raised by 6.3 % (75.1 ~~% and 6.6 % (40.2 and 34.9~~ Mt CO_2/yr~~) respectively.~~).

Following the Intergovernmental Panel on Climate Change (IPCC) 2006 Guidelines for National Greenhouse Gas Inventories and revised information from its 2019 Refinements (IPCC-TFI, 2019) we start from the global fossil CO_2 grid-maps of EDGAR inventory versions 4.3.2 (Janssens-Maenhout et al., 2019) and 4.3.2_FT2015 (Olivier et al., 2016a), for 2012 and 2015 respectively, and derive an updated emission dataset as prior input to the ECMWF model: CHE_EDGAR-ECMWF_2015 (CHE stands for the CO_2 Human Emissions project (CHE, 2020)). We improve the apportionmentEDGARv4.3.2 dataset by correcting the allocation of the autoproducers (autoproducers are defined by International Energy Agency (IEA) and include the energy sector and the energy used(electricity and heat) generated by an industry for its own use, mostly for the manufacturing, add) to the manufacturing sector instead of the energy sector and by adding the diffusive CO_2 emissions from coal mines and. We then aggregate the sectors in 7 emission groups while tracking 232 countries separately. Uncertainties are calculated per country and sector considering the most typical fuel type using the error propagation method of the IPCC (2006) guidelines. According to the IPCC (2006) guidance all emissions are considered to be fully uncorrelated; this assumption is further used to calculate uncertainty and covariance matrices. The country-based uncertainties and the share to the total uncertainty are presented for the 7 ECMWF emission groups, with calculations based on 20 EDGAR sectors for two distinct country types with well- and less well-developed statistical infrastructure.infrastructures. While the uncertainty of most groups (i.e. power industry, combustion for manufacturing, and road transport) remains small, (5-20 %), the largest contribution (over 40 %) to the total uncertainty is determined by rather small but relative uncertain (more than 100 %) sectors (i.e. non energy use of fuels, chemical processes, fuel exploitation, and coal production) emissions.

This paper is organised as follows. Section 2 describes the data sources and includes the description of the anthropogenic CO_2 emission datasets used to calculate emission uncertainties, data pre-processing, emission sectors and groups, and geographical treatment of emissions. Section 3 discusses the uncertainty calculation methodology applied to the datasets, to calculate both yearly and monthly uncertainties. Emission country and sector budgets comparison with other institute data and discussion of the results and further developments are covered in Section 4 dedicated to comparison and discussion.Section 4 provides details on the newly generated dataset. National sectorial emission budgets are compared in Section 5. The main results, a discussion and further research guidance are covered in the conclusion in Section 56. This paper also has Supplementary Information with details on methods and assumptions used.

2 Data

2.1 Update of fossil CO_2 emissions as input for the ECMWF model

Main requirements for datasets in order to be used in global numerical models are being global and gridded, and preferably with continuous update. In this study it was decided to use EDGARv4.3.2 (and EDGARv4.3.2_FT2015) because it is based on international statistics, mainly International Energy Agency (IEA) data, has a unique global geo-coverage with 228 countries/regions and continuous updates withof the time-series. The most relevant activity data for both EDGARv4.3.2 and CHE_EDGAR-ECMWF_2015 are the energy statistics from 1970 onwards, till the year 1IEA (2014), which has been corrected for few outliers and for the revised Chinese coal statistics of 2015. For the update from 2012 to 2015 we used the fast track approach of Olivier et al. (2016b), with IEA (2016) energy statistics and BP (2017) statistics. EDGAR distributes anthropogenic emissions for each source category over a uniform, global 0.1º×0.1º grid defined with lower left coordinates and provides annual and monthly global emissions grid-maps. In emission inventories the emissions can be emitted either

~~from a single point source (e.g. power plants, factories) or distributed over a linear source (e.g. roads) or over an area source (e.g. agricultural fields), depending on the source sector or subsector.~~ The bottom-up emissions calculation methodology and ~~(mainly default) EFs~~emission factors, either defaults recommended by IPCC (2006) guidelines or region-specific ones justified by scientific evidence, are consistently applied to all countries in order to achieve comparability and full transparency. ~~Region-specific EFs are selected, when these are recommended by IPCC (2006) guidelines or when these are justified by robust information on significant differences in economic activities, in customs or in geographical ambient conditions and proven to be more representative than the global average. All sectors based on fuel or product consumption statistics are considered.~~

We focus on long-cycle carbon~~-cycle~~ CO_2 and therefore consider the CO_2 from fossil fuel use (combustion and other use of 42 fossil fuels) and from industrial processes (cement production, carbonate use of limestone and dolomite, non-energy use of fuels and other combustion, chemical and metal processes, solvents, agricultural liming and urea, waste and fossil fuel fires). Excluded are consumption of biofuels and short-cycle biomass burning (such as agricultural waste burning), large-scale biomass burning (such as forest fires, ~~Savannah~~savannah burning, woodland and peatland fires) and carbon emissions/removals of land-use, land-use change and forestry (LULUCF)[1]. Based on the Global Carbon Budget 2018 findings this sector showed no significant trend since 1960s, only high year-to-year variability and high uncertainty (Bastos et al., 2020; Le Quéré et al., 2018; Arneth et al., 2017). We excluded also the fossil fuel fires, because we ~~do not~~ focus ~~on historical time series but~~only on 2015~~ (so~~, for which the Kuwait oil fires of 1991 are of no importance~~)~~ and the coal mine fires data are considered to be very uncertain~~. The most relevant activity data (AD) for our CHE_EDGAR-ECMWF_2015 are the energy statistics from IEA (2014), which has been corrected for few outliers and for the revised Chinese coal statistics of 2015~~.

~~While EDGARv4.3.2 provides emissions~~Starting from EDGARv4.3.2_FT2015, the following updates were considered necessary for the derivation of ~~150 activities, and 42 fossil fuels~~the CHE_EDGAR-ECMWF_2015 dataset. Firstly, there was a need to ~~re-attribute~~ reallocate the part of autoproducers in the energy sector to the manufacturing industry in ~~order to match the United Nations Framework Convention on Climate Change (~~line with UNFCCC~~)~~ reporting. ~~EDGARv4.3.2_FT2015~~The autoproducers' energy ~~sector emissions were divided into autoproducers and the rest. The autoproducing energy part~~generated and used for industrial manufacturing was added to the ~~industry~~ manufacturing sector (causing an increase of 18 %) and taken away from the energy sector ~~as it is generated purposely for manufacturing, and not for power generation in general.~~(leading to a decrease of 8 %). The reallocation of the autoproducers part ~~reported in~~was done using the energy statistics reported by every country separately (IEA, 2016) ~~was reattributed to the manufacturing in CHE_EDGAR-ECMWF_2015~~ but the correction remained limited to 30 % of the national total energy sector. More details are given in the Supplementary Information, section S.1.

[1] Following the UNFCCC national inventory reporting guidelines, emissions of biofuel combustion are only a memo item and have to be reported under the LULUCF sector. Together with all short-cycle carbon emissions they are excluded from this study.

~~Another update resulted in the expansion of the emissions with the~~Secondly, super power plants were considered to be treated separately, because they are expected to operate at full capacity with maximum availability. Super power plants are defined in this study as a large power plant or a group of closely located power plants causing CO_2 plumes from a single grid cell with a CO_2 flux $\geq 7.9 \cdot 10^{-6}$ kg·m^{-2}·s^{-1}. According to expert knowledge the upper bound of uncertainty for super power plants is not larger than +3.0 %, whereas for small plants which operate based on day-to-day needs, this can reach up to +15.0 %. Currently 30 grid-cells from 12 countries represent energy generated by the super power plants (7.1 % or 896.7 Mton of the remaining energy sector after autoproducers part separation 12705.5 Mton). Top 3 countries that produce energy using super power plants are China, Russia and India. For the detailed ranking of the power plant sites in function of their emission intensity, we refer to the Supplementary Information, section S.1.

Finally, an extra emission source of fugitive CO_2 from coal mines was added, following the recommendations from IPCC-TFI (2019). Even though this emission source is not that large globally, ~~it is a highly uncertain emission source that was detected by space-borne images over the United States of America.~~usually the coalseam gas is composed dominantly from methane, but in some coalmines (in Australia, and also in Brazil) seam gas consists predominantly (> 95 %) from CO_2 (Beamish and Vance, 1992), leading to significant atmospheric CO_2 concentration increases. An additional map for CHE_EDGAR-ECMWF_2015 with coal mining emissions from underground mines has been generated, following the IPCC-TFI (2019) default values and the coal mining activity of the methane (CH_4) emission grid-maps from hard and brown coal production of EDGARv4.3.2. More details are given in the Supplementary Information, section S.2, in which Table S3 lists all differences between EDGARv4.3.2_FT2015 and CHE_EDGAR-ECMWF_2015.

The detailed EDGARv4.3.2 spatial distribution is used for mapping the updated 2015 emission values~~. For the update from 2012 to 2015 we used the fast track approach of Olivier~~ (Janssens-Maenhout et al. ~~(2016b), with IEA (2016) energy statistics~~(2019) provide all special details on how emissions are spatially distributed and ~~BP (2017) statistics.~~what proxies are used for that in EDGARv4.3.2). The relative changes per sector, fuel type and country from 2012 to 2015 are then applied on the EDGARv4.3.2 reference maps to obtain EDGARv4.3.2_FT2015.

For non-energy use of fuels, chemical processes, and solvents and products use we used directly the EDGARv4.3.2 maps. Also, the CO_2 emission maps from coal production are based on the 2012 maps of CH_4 from EDGARv4.3.2. Gridded monthly multiplication factors are obtained from 2010 monthly gridded emissions and applied to the final set of yearly emission maps of CHE_EDGAR-ECMWF_2015.

~~For the full list of differences between EDGARv4.3.2_FT2015 and CHE_EDGAR-ECMWF_2015, we refer to the Supplementary Information, section S.2 Table S3.~~

2.2 Aggregation of CO_2 emission groups for the ECMWF model

EDGARv4.3.2_FT2015 (as well as EDGARv4.3.2) has 20 global maps with anthropogenic long-cycle carbon~~-cycle~~ CO_2 flux values for energy, fugitives, industrial processes, solvents and products use, agriculture and waste involved sectors. In this study these sectors had to be grouped for the use of global flux inversion and ensemble perturbation systems. Grouping was done keeping in mind possible future evolution of present systems and sector common features: activity type (point sources, 3D field, etc.), amount of knowledge for the activity (uncertainty value), geographical distribution (e.g. over urban areas only), size of sector covariance matrix~~. An adequate~~ (computationally affordable size for the inversion system of the ECMWF model is ~~less than 50 and a~~ covariance matrix of 7×7~~ has been chosen.~~). Table 2 shows additional grouping of 20 EDGAR sectors into 7 ECMWF groups~~. The remaining energy sector (after~~ , and emission budget difference between EDGARv4.3.2_FT2015 and CHE_EDGAR-ECMWF_2015 datasets due to reallocation of the autoproducers ~~part separation) was divided into one produced by super power plants, and one produced by average (non-super) power plants. As super power plants are considered grid-cells with annual flux 7.9·10^-6 kg·m^-2·s^-1 and higher. In total there are 30 super power plant grid-cells, all the remaining energy sector grid-cells are assumed to have emissions~~ from the ~~average power plants. For the detailed ranking of the power plant sites in function of their emission intensity, we refer~~energy sector (-8 %) to the ~~Supplementary Information, section S.1~~manufacturing sector (+18 %), and due to the extra emission source of diffusive coal mine CO_2.

Table 2: Grouping of anthropogenic long-cycle carbon CO_2 emission EDGAR sectors into ECMWF groups, note provides main information and typical fuel type, global emission budgets for 2015 in Mton provides values for EDGARv4.3.2_FT2015 and CHE_EDGAR-ECMWF_2015; *italics* – values with biggest differences, * – values that were replaced from EDGARv4.3.2

№	ECMWF group	IPCC (2006) activities per EDGAR sector	Note	Emission budget 2015, Mton	
				EDGARv4.3.2_FT2015	CHE_EDGAR-ECMWF_2015
1	ENERGY_S	1.A.1.a (subset)	Power industry (without autoproducers): super emitting power plants	*13704.0*	*896.7*
2	ENERGY_A	1.A.1.a (rest)	Power industry (without autoproducers): standard emitting power plants		*11671.6*
		4.C	Solid waste incineration	137.2	137.2
3	MANUFACTURING	1.A.2	Combustion for manufacturing (including autoproducers)	*6182.8*	*7320.4*
		2.C.1, 2.C.2	Iron and steel production	233.6	233.6
		2.C.3, 2.C.4, 2.C.5, 2.C.6, 2.C.7	Non-ferrous metals production	91.4	91.4
		2.D.1, 2.D.2, 2.D.4	Non energy use of fuels	24.7*	24.6
		2.A.1, 2.A.2, 2.A.3, 2.A.4	Non-metallic minerals production	1748.8	1749.0
		2.B.1, 2.B.2, 2.B.3, 2.B.4, 2.B.5, 2.B.6, 2.B.8	Chemical processes	678.8*	677.0
4	SETTLEMENTS	1.A.4, 1.A.5.a, 1.A.5.b.i, 1.A.5.b.ii	Energy for buildings	3321.9	3322.7
5	AVIATION	1.A.3.a_CRS	Aviation cruise; typical fuel: jet kerosene	412.2	412.2
		1.A.3.a_CDS	Aviation climbing & descent; typical fuel: jet kerosene	305.5	305.5
		1.A.3.a_LTO	Aviation landing & take off; typical fuel: jet kerosene	97.7	97.7
6	TRANSPORT	1.A.3.b	Road transportation; typical fuel: most typical emission factor uncertainty	5530.2	5530.6
		1.A.3.d	Shipping; typical fuel: composition of 80 % diesel and 20 % residual fuel oil	819.0	819.1
		1.A.3.c, 1.A.3.e	Railways, pipelines, off-road transport; typical fuel: railways – diesel, off-road	255.2	255.2

			transport – most typical emission factor uncertainty		
7	OTHER	1.A.1.b, 1.A.1.c, 1.A.5.b.iii, 1.B.1.c, 1.B.2.a.iii.4, 1.B.2.a.iii.6, 1.B.2.b.iii.3	Oil refineries and Transformation industry	1917.4	1917.8
		1.B.2.a.ii, 1.B.2.a.iii.2, 1.B.2.a.iii.3, 1.B.2.b.ii, 1.B.2.b.iii.2, 1.B.2.b.iii.4, 1.B.2.b.iii.5, 1.C	Fuel exploitation	258.4	258.4
		1.B.1.a	Coal production	*0.0*	*7.0*
		3.C.2, 3.C.3, 3.C.4, 3.C.7	Agricultural soils	99.0	99.1
		2.D.3, 2.B.9, 2.E, 2.F, 2.G	Solvents and products use	168.7*	168.3

3 Uncertainty calculation methodology

3.1 Overview

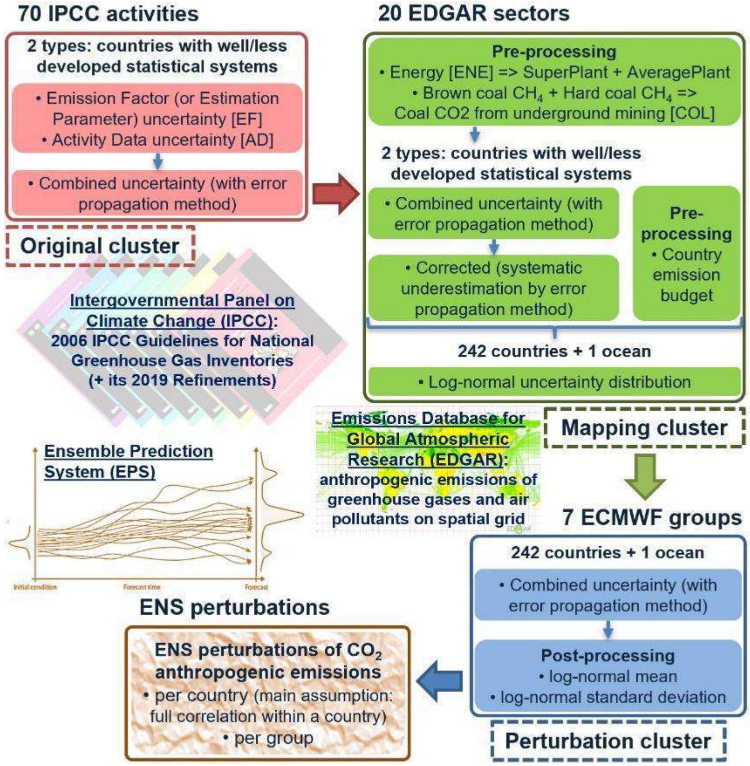
The IPCC (2006) Guidelines for NIR for fossil CO_2 uncertainty calculations and updated IPCC-TFI (2019) provide vast information about numerous human activities emitting CO_2 and how certain these values are. Use of the IPCC-TFI (2019) permitted to consider the 2019 ~~EF~~emission factor and ~~AD~~activity data uncertainties for petroleum refining, solid fuel manufacturing, transformation, processing and transport and oil and gas production, which differed significantly from the 2006 defaults. In order to use the same methodology globally and because CO_2 emissions are not technologically dependant, it was decided to omit regional (e.g. Europe) detailed information and use only information required for the most basic and simplest (Tier 1) approach for emission reporting. The Tier 1 methodology to estimate CO_2 emissions from fossil fuel combustion follows the concept of carbon conservation (from the fuel combusted into CO_2). Uncertainties for all emission activities, sectors and groups can be derived following two different approaches of IPCC (2006): (Approach 1) propagation of error – gives informative results even if the criterion "standard deviation divided by the mean value is less than 0.3" is not strictly met and data still have some correlation. The advantages are that it only needs uncertainty ranges for ~~AD~~activity data and ~~EF~~emission factors, that are provided by IPCC and that it is relatively easy to improve in case of large and asymmetric uncertainties; (Approach 2) Monte Carlo simulation or similar techniques – suitable only if detailed category-by-category uncertainty information is available and complex calculations can be done. In order to use the same methodology for all world countries/geographical entities (i.e. not needing detailed information for each emission activity) it was decided to use the error propagation method (Approach 1).

~~**Table 2: Additional grouping of anthropogenic long carbon cycle CO_2 emission EDGAR sectors (with global emission budgets for 2015 in Mton) into ECMWF groups**~~

~~№~~	~~ECMWF group~~	~~IPCC (2006) activities per EDGAR sector~~	~~Note~~	~~Emission budget, Mton~~
~~1~~	~~ENERGY_S~~	~~1.A.1.a (subset)~~	~~Power industry (without autoproducers):~~	~~896.7~~

			super emitting power plants	
2	ENERGY_A	1.A.1.a (rest)	Power industry (without autoproducers): average emitting power plants	11671.6
		4.C	Solid waste incineration	137.2
3	MANUFACTURING	1.A.2	Combustion for manufacturing (including autoproducers)	7320.4
		2.C.1, 2.C.2	Iron and steel production	233.6
		2.C.3, 2.C.4, 2.C.5, 2.C.6, 2.C.7	Non-ferrous metals production	91.4
		2.D.1, 2.D.2, 2.D.4	Non energy use of fuels	24.6
		2.A.1, 2.A.2, 2.A.3, 2.A.4	Non-metallic minerals production	1749.0
		2.B.1, 2.B.2, 2.B.3, 2.B.4, 2.B.5, 2.B.6, 2.B.8	Chemical processes	677.0
4	SETTLEMENTS	1.A.4, 1.A.5.a, 1.A.5.b.i, 1.A.5.b.ii	Energy for buildings	3322.7
5	AVIATION	1.A.3.a_CRS	Aviation cruise	412.2
		1.A.3.a_CDS	Aviation climbing & descent	305.5
		1.A.3.a_LTO	Aviation landing & take off	97.7
6	TRANSPORT	1.A.3.b	Road transportation	5530.6
		1.A.3.d	Shipping	819.1
		1.A.3.c, 1.A.3.e	Railways, pipelines, off-road transport	255.2
7	OTHER	1.A.1.b, 1.A.1.c, 1.A.5.b.iii, 1.B.1.c, 1.B.2.a.iii.4, 1.B.2.a.iii.6, 1.B.2.b.iii.3	Oil refineries and Transformation industry	1917.8
		1.B.2.a.ii, 1.B.2.a.iii.2, 1.B.2.a.iii.3, 1.B.2.b.ii, 1.B.2.b.iii.2, 1.B.2.b.iii.4, 1.B.2.b.iii.5, 1.C	Fuel exploitation	258.4
		1.B.1.a	Coal production	7.0
		3.C.2, 3.C.3, 3.C.4, 3.C.7	Agricultural soils	99.1
		2.D.3, 2.B.9, 2.E, 2.F, 2.G	Solvents and products use	168.3

To summarize, the final uncertainties per geographical entity per ECMWF fossil CO_2 emission group are based on: emission budgets calculated from CHE_EDGAR-ECMWF_2015 maps (upgraded combination of EDGARv4.3.2 and EDGARv4.3.2_FT2015), uncertainty default values from IPCC (2006) and IPCC-TFI (2019), Tier 1 approach (error propagation method) and the definition of a log-normal distribution (needed for non-negative anthropogenic CO_2 emissions). It should be noted that all uncertainty calculations were done per country (geographical entity) and only then for comparison purposes aggregated to Europe (28 members till end 2019) or global values assuming no correlation following IPCC (2006). Figure 1 shows a simplified scheme of the uncertainty calculation roadmap, followed by a detailed description below on how exactly yearly and monthly uncertainties are calculated.


70 IPCC activities
2 types: countries with well/less developed statistical systems
• Emission Factor (or Estimation Parameter) uncertainty [EF]
• Activity Data uncertainty [AD]
• Combined uncertainty (with error propagation method)
Original cluster
Intergovernmental Panel on Climate Change (IPCC):
2006 IPCC Guidelines for National Greenhouse Gas Inventories (+ its 2019 Refinements)
20 EDGAR sectors
Pre-processing
• Energy [ENE] => SuperPlant + AveragePlant
• Brown coal CH_4 + Hard coal CH_4 => Coal CO2 from underground mining [COL]
2 types: countries with well/less developed statistical systems
• Combined uncertainty (with error propagation method)
• Corrected (systematic underestimation by error propagation method)
Pre-processing
• Country emission budget
242 countries + 1 ocean
• Log-normal uncertainty distribution
Emissions Database for Global Atmospheric Research (EDGAR):
anthropogenic emissions of greenhouse gases and air pollutants on spatial grid
Mapping cluster
7 ECMWF groups
242 countries + 1 ocean
• Combined uncertainty (with error propagation method)
Post-processing
• log-normal mean
• log-normal standard deviation
Perturbation cluster
Ensemble Prediction System (EPS)
ENS perturbations
ENS perturbations of CO_2 anthropogenic emissions
• per country (main assumption: full correlation within a country)
• per group

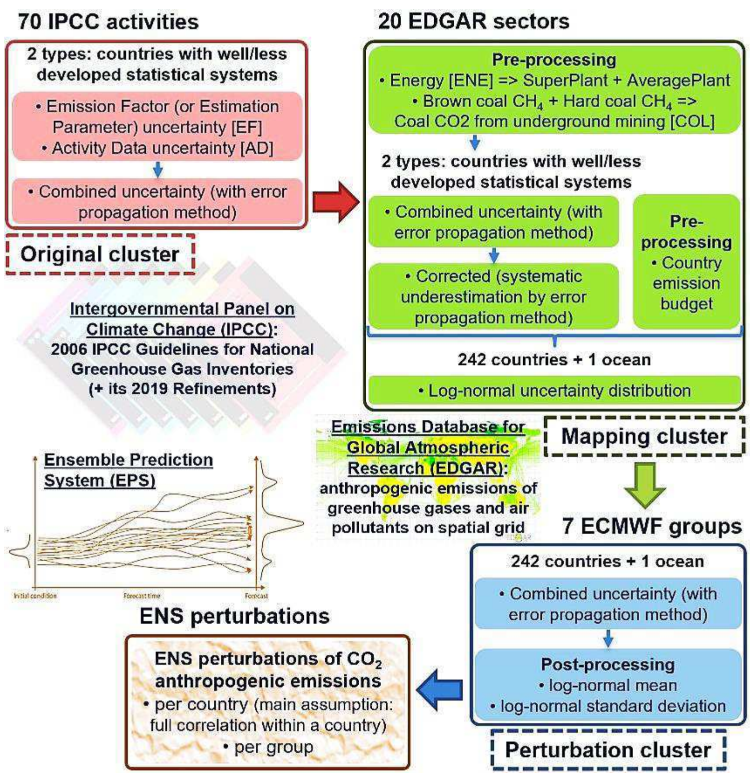

70 IPCC activities
2 types: countries with well/less developed statistical systems
• Emission Factor (or Estimation Parameter) uncertainty [EF]
• Activity Data uncertainty [AD]
• Combined uncertainty (with error propagation method)
Original cluster
Intergovernmental Panel on Climate Change (IPCC):
2006 IPCC Guidelines for National Greenhouse Gas Inventories (+ its 2019 Refinements)
20 EDGAR sectors
Pre-processing
• Energy [ENE] => SuperPlant + AveragePlant
• Brown coal CH_4 + Hard coal CH_4 => Coal CO2 from underground mining [COL]
2 types: countries with well/less developed statistical systems
• Combined uncertainty (with error propagation method)
• Corrected (systematic underestimation by error propagation method)
Pre-processing
• Country emission budget
242 countries + 1 ocean
• Log-normal uncertainty distribution
Emissions Database for Global Atmospheric Research (EDGAR):
anthropogenic emissions of greenhouse gases and air pollutants on spatial grid
Mapping cluster
7 ECMWF groups
242 countries + 1 ocean
• Combined uncertainty (with error propagation method)
Post-processing
• log-normal mean
• log-normal standard deviation
Perturbation cluster
Ensemble Prediction System (EPS)
ENS perturbations
ENS perturbations of CO_2 anthropogenic emissions
• per country (main assumption: full correlation within a country)
• per group

Figure 1: Simplified roadmap for yearly uncertainty calculation

3.2 Yearly uncertainties

3.2.1 Calculating uncertainty per each IPCC activity

Uncertainties in the emissions per IPCC activity from Table 2 – Combined Uncertainties UC_{IPCCi} – were calculated using uncertainties for emission factors EF_{IPCCi} and activity data AD_{IPCCi} in % provided in IPCC (2006) and IPCC-TFI (2019) following Eq. (1):

$$UC_{IPCCi} = \sqrt{EF_{IPCCi}^2 + AD_{IPCCi}^2}. \quad (1)$$

It should be noted that IPCC (2006) and IPCC-TFI (2019) provide upper and lower limits of ~~EF~~emission factor and ~~AD~~activity data, which are not always symmetrical. In order to preserve as much initial information as possible (and not to inflate artificially lower or upper limits of log-normal emission distributions) all calculations were performed for upper and lower uncertainty limits separately although it is not required by the Approach 1 methodology. Moreover, IPCC (2006) provide default ~~EF~~emission factor values for different fuels in transport-related activities (e.g. railways, aviation, etc.). Detailed fuel consumption information per each IPCC activity that result in long-cycle carbon was not available and it was decided to use the most typical and consumed (common) fuel type (its ~~EF~~emission factor value~~) per each activity. The following~~). Table 2 shows the most typical fuels ~~were assumed as most typical ones:~~ for ~~aviation – jet kerosene, for railways – diesel, and for shipping (or water-borne navigation) – composition of 80 % diesel and 20 % residual fuel oil. Following IPCC (2006) recommendations for road and off-road~~each transport ~~the most typical EF uncertainty was used (instead of the typical fuel type EF).~~related sector.

3.2.2 Calculating uncertainty for each EDGAR sector

Uncertainties for each of the 70 IPCC activities from Table 2 are calculated with the error propagation method and combined into the 20 EDGAR sectors, following Eq. (2):

$$UC_{EDGARj} = \sqrt{UC_{IPCC1}^2 + UC_{IPCC2}^2 + \dots + UC_{IPCCn}^2}, \quad (2)$$

where *EDGARj* – combined uncertainty per sector *j*, and *1,2,...,n* – IPCC activities that are taken into account in a particular EDGAR sector; UC_{IPCC1}, UC_{IPCC2},..., UC_{IPCCn} used in %.

3.2.3 Correction of EDGAR sector uncertainty due to underestimation by the chosen method

The EDGAR sector uncertainty had to be corrected, as the error propagation method of Approach 1 systematically underestimates the uncertainty unless the model is purely additive, which was not the case~~. Here, uncertainty calculations~~ as EDGAR emissions are estimated based on the sum of several product terms. To fix this underestimation IPCC (2006) advises using a correction factor. One example of a correction factor is proposed in Frey (2003), where the performance of an analytical approach for combining uncertainty in comparison to a Monte Carlo simulation with large sample sizes for many cases involving different ranges of uncertainty for additive, multiplicative, and quotient models are evaluated. Frey found that error propagation and Monte Carlo simulated estimates of the uncertainty half-range of the model output agreed well for values of less than 100 %, but with the increase of the uncertainty a systematic underestimation of uncertainty in the total inventory by the error propagation approach appeared. The relationship between the simulated and propagated error estimates was found to be well-behaved, which led to a correction factor development for the large (i.e. greater than 100 %) total inventory uncertainties. This correction factor will not necessarily be reliable for very large uncertainties (i.e. greater than 230 %) because it was calibrated over the range of 10 to 230 %. As such, the correction factor *FC*, calculated following Eq. (3), was applied if half-range uncertainty estimated from the error propagation method was > 100 and < 230 % following Eq. (4):

$$FC_{EDGARj} = \left[\frac{-0.7200+1.0921\cdot UC_{EDGARj}-1.63\cdot 10^{-3}\cdot UC_{EDGARj}^{2}+1.11\cdot 10^{-5}\cdot UC_{EDGARj}^{3}}{UC_{EDGARj}}\right]^{2}, \quad (3)$$

$$\left(UC_{EDGARj}\right)_{corr} = UC_{EDGARj} \cdot FC_{EDGARj}, \quad (4)$$

where *corr* corresponds to the corrected uncertainty; UC_{EDGARj} is given in %. In cases where UC_{EDGARj} was ≤ 100 and ≥ 230 %, FC_{EDGARj} was assumed to be equal to one. Only four sectors with non energy use of fuels, chemical processes, fuel exploitation and coal production emissions were corrected, Table 3 shows how these uncertainties were corrected. It should be noted that 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 input from expert knowledge or available in-situ data, which lead to the same pattern in final prior uncertainty bounds.

Table 3: Sectors with corrected uncertainties (lower and upper bounds) for countries with well- (WDS) and less well-developed (LDS) statistical infrastructures

№	ECMWF group	IPCC (2006) activities per EDGAR sector	Note	Country type	Prior uncertainty bounds, %			
					Before correction		After correction	
					Low	Up	Low	Up
3	MANUFACTURING (part)	2.D.1, 2.D.2, 2.D.4	Non energy use of fuels	WDS	112.0	112.0	121.7	121.7
				LDS	113.8	113.8	124.0	124.0
		2.B.1, 2.B.2, 2.B.3, 2.B.4, 2.B.5, 2.B.6, 2.B.8	Chemical processes	WDS	100.9	89.9	107.8	89.9
				LDS	100.9	89.9	107.8	89.9
7	OTHER (part)	1.B.2.a.ii, 1.B.2.a.iii.2, 1.B.2.a.iii.3, 1.B.2.b.ii, 1.B.2.b.iii.2, 1.B.2.b.iii.4, 1.B.2.b.iii.5, 1.C	Fuel exploitation	WDS	156.6	215.7	191.1	339.1
				LDS	166.8	223.2	210.9	364.5
		1.B.1.a	Coal production	WDS	107.4	300.5	115.8	300.5
				LDS	107.4	300.5	115.8	300.5

3.2.4 Forcing lognormal distribution on corrected EDGAR sector uncertainty

For models that are purely additive, and for which the half range of uncertainty is less than approximately 50 %, a normal distribution is often an accurate assumption for the model output form. In this case, a symmetric probability distribution with respect to the mean can be assumed. But this is not the case for multiplicative (or mixed) models, or when the uncertainty is large for a non-negative variable such as anthropogenic CO_2 emissions. A log-normal distribution is typically an accurate assumption for the model output form, where the uncertainty range is not symmetric with respect to the mean, even though the variance for the total inventory may be correctly estimated from Approach 1. IPCC (2006) guidelines provide a practical methodology based on Frey (2003) for approximate asymmetric uncertainty range calculations based on the error propagation method. According to this methodology key characteristics of the 95 % confidence intervals are: (i) approximately symmetric for small ranges of uncertainty, and (ii) positively skewed for large ranges of uncertainty. This methodology was applied if the corrected lower half-range uncertainty estimated from error propagation method was ≥ 50 %. More details on the IPCC (2006) parametrisationsuggests to define parameters of the log-normallognormal distribution is given in terms of the geometric mean μg (which can be estimated based upon the Supplementary Information, section S.3. arithmetic mean and the arithmetic standard deviation) following Eq. (5) and geometric standard deviation σg following Eq. (6):

$$\mu g_{EDGARj} = exp\left\{ln\left(E_{EDGARj}\right) - \frac{1}{2}\cdot ln\left(1 + \left[\frac{\left(UC_{EDGARj}\right)_{corr}}{200}\right]^2\right)\right\}, \quad (5)$$

$$\sigma g_{EDGARj} = exp\left\{\sqrt{ln\left(1 + \left[\frac{\left(UC_{EDGARj}\right)_{corr}}{200}\right]^2\right)}\right\}, \quad (6)$$

where E_{EDGARj} is the anthropogenic CO_2 emissions per sector *j*; *corr* corresponds to the corrected uncertainty (i.e. corrected for the systematic underestimation of uncertainty calculated by the error propagation approach used in this study comparing to uncertainties calculated by using the Monte Carlo approach); UC_{EDGARj} is in %.

Because calculations were performed for upper and lower uncertainty limits separately, there are two values of $\left(UC_{EDGARj}\right)_{corr}$: $\left[\left(UC_{EDGARj}\right)_{corr}\right]_{low}$ – the absolute value of the lower uncertainty limit of sector *j*, and $\left[\left(UC_{EDGARj}\right)_{corr}\right]_{high}$ – the absolute value of the upper uncertainty limit of sector *j*. As it is preferred to preserve as much accuracy (extra knowledge) as possible in our calculations and not to inflate uncertainty upper or lower bounds artificially, lower $\left\{\left[\left(UC_{EDGARj}\right)_{corr}\right]_{low}\right\}_{ln}$ and upper $\left\{\left[\left(UC_{EDGARj}\right)_{corr}\right]_{high}\right\}_{ln}$ uncertainty half-range from the error propagation method were calculated with a logarithmic transformation using $\left[\mu g_{EDGARj}\right]_{low}$, $\left[\mu g_{EDGARj}\right]_{high}$ and $\left[\sigma g_{EDGARj}\right]_{low}$, $\left[\sigma g_{EDGARj}\right]_{high}$ respectively according to the following Eq. (7) and Eq. (8) (see Figure 2 for visual representation of these equations):

$$\left\{\left[\left(UC_{EDGARj}\right)_{corr}\right]_{low}\right\}_{ln} = \left(\frac{exp\left\{ln\left(\left[\mu g_{EDGARj}\right]_{low}\right)-1.96\cdot ln\left(\left[\sigma g_{EDGARj}\right]_{low}\right)\right\}-E_{EDGARj}}{E_{EDGARj}}\right)\times 100, \qquad (7)$$

$$\left\{\left[\left(UC_{EDGARj}\right)_{corr}\right]_{high}\right\}_{ln} = \left(\frac{exp\left\{ln\left(\left[\mu g_{EDGARj}\right]_{high}\right)+1.96\cdot ln\left(\left[\sigma g_{EDGARj}\right]_{high}\right)\right\}-E_{EDGARj}}{E_{EDGARj}}\right)\times 100, \qquad (8)$$

where *ln* corresponds to logarithmic transformation of the distribution; resulting values are not absolute.

It should be noted that according to this methodology (with constants for 2.5th and 97.5th percentiles, +1.96 and -1.96 respectively, from the Z-table[2]) the lower uncertainty half-range $\left\{\left[\left(UC_{EDGARj}\right)_{corr}\right]_{low}\right\}_{ln}$ will always be less than 100.0 %. Upper uncertainty half-range $\left\{\left[\left(UC_{EDGARj}\right)_{corr}\right]_{high}\right\}_{ln}$ is approximately symmetric relative to the 0 (Gaussian distribution) up to ~20.0 %, then has rather rapid growth till ~500.0 % (which with logarithmic transformation results in ~486.0 %), maxima at ~1350.0 % (which with logarithmic transformation results in ~582.6 %) and further gradual decrease.

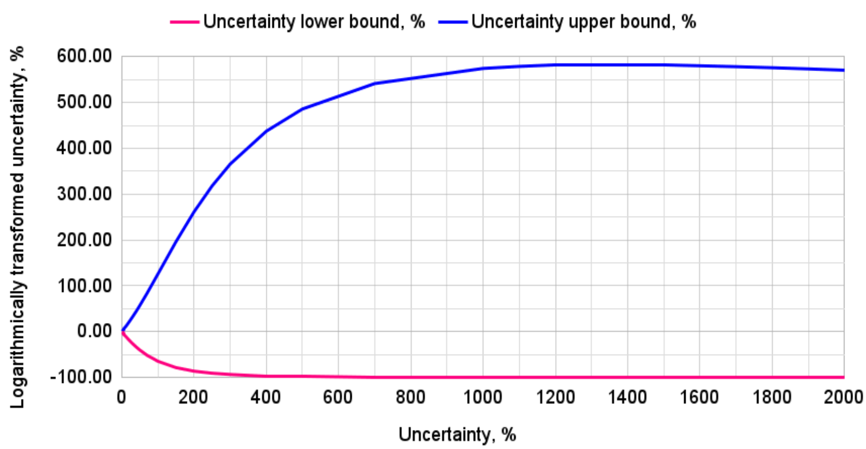


Figure 2: Visual representation of an empirical logarithmic transformation formula for upper and lower uncertainty bounds according IPCC (2006)

Table ~~3~~4 shows the prior uncertainty values for each EDGAR sector and for two geographical entity types (i.e. with well ~~(WDS)~~ and less well ~~(LDS) statistically~~ -developed statistical infrastructure). These values are a combined IPCC activity uncertainty aggregated to EDGAR sectors with the error propagation method and corrected for this method's underestimation. Also, as an example, Table ~~3~~4 shows aggregated to ECMWF groups uncertainties with ensured log-normal distribution for China (CHN), Europe (28 members till end 2019) and all world countries ~~(GLB)~~.

[2] The Z-table is a mathematical table for the values of the cumulative distribution function of the normal distribution.

Table ~~3~~4: Prior uncertainties (lower ~~L~~ and upper ~~U~~ bounds) per each EDGAR emission sector and two geographical entity types (with well- (WDS) and less well-developed (LDS) statistical infrastructure) based on IPCC (2006) and IPCC-TFI (2019), and aggregated to the ECMWF group uncertainties for China (CHN), Europe (E28) and globe (GLB)

№	ECMWF group	IPCC (2006) activities per EDGAR sector	Prior uncertainty bounds, %: WDS countries		Prior uncertainty bounds, %: LDS countries		Uncertainty bounds, %: CHN, WDS		Uncertainty bounds, %: E28, WDS		Uncertainty bounds, %: GLB, mix	
			~~L~~Low	~~U~~Up	~~L~~Low	~~U~~Up	~~L~~Low	~~U~~Up	~~L~~Low	~~U~~Up	~~L~~Low	~~U~~Up
1	ENERGY_S	1.A.1.a (subset)	8.6	3.0	12.2	3.0	8.6	3.0	5.4	1.9	3.6	1.0
2	ENERGY_A	1.A.1.a (rest)	8.6	8.6	12.2	12.2	8.6	8.6	2.8	2.8	3.5	3.5
		4.C	40.3	40.3	41.2	41.2						
3	MANUFACTURING	1.A.2	8.6	8.6	12.2	12.2	12.8	19.4	3.9	5.8	5.7	8.6
		2.C.1, 2.C.2	37.1	37.1	37.1	37.1						
		2.C.3, 2.C.4, 2.C.5, 2.C.6, 2.C.7	73.2	73.2	73.2	73.2						
		2.D.1, 2.D.2, 2.D.4	121.7	121.7	124.0	124.0						
		2.A.1, 2.A.2, 2.A.3, 2.A.4	70.9	70.9	93.0	93.0						
		2.B.1, 2.B.2, 2.B.3, 2.B.4, 2.B.5, 2.B.6, 2.B.8	107.8	89.9	107.8	89.9						
4	SETTLEMENTS	1.A.4, 1.A.5.a, 1.A.5.b.i, 1.A.5.b.ii	12.2	12.2	26.0	26.0	12.2	12.2	4.2	4.2	3.9	3.9
5	AVIATION	1.A.3.a_CRS	5.5	6.4	50.1	106.8	3.5	4.1	1.4	1.6	17.3	58.1
		1.A.3.a_CDS	5.5	6.4	50.1	106.8						
		1.A.3.a_LTO	5.5	6.4	50.1	106.8						
6	TRANSPORT	1.A.3.b	5.4	5.4	7.1	7.1	5.1	8.2	1.6	1.8	4.3	6.4
		1.A.3.d	5.4	5.1	50.0	50.0						
		1.A.3.c, 1.A.3.e	50.3	106.9	50.5	107.0						
7	OTHER	1.A.1.b, 1.A.1.c, 1.A.5.b.iii, 1.B.1.c, 1.B.2.a.iii.4, 1.B.2.a.iii.6, 1.B.2.b.iii.3	54.4	149.3	57.7	151.4	39.7	180.9	10.1	45.3	11.5	52.4
		1.B.2.a.ii, 1.B.2.a.iii.2, 1.B.2.a.iii.3, 1.B.2.b.ii, 1.B.2.b.iii.2, 1.B.2.b.iii.4, 1.B.2.b.iii.5, 1.C	191.1	339.1	210.9	364.5						
		1.B.1.a	115.8	300.5	115.8	300.5						
		3.C.2, 3.C.3, 3.C.4, 3.C.7	70.7	0.0	70.7	0.0						
		2.D.3, 2.B.9, 2.E, 2.F, 2.G	25.0	25.0	50.0	50.0						

3.2.5 Calculating uncertainty for each ECMWF group

The next step is to combine these prior uncertainties for each EDGAR sector into ECMWF group uncertainties (see Table ~~3~~4). Sector uncertainties are combined into group uncertainties by addition following Eq. (~~5~~9) and Eq. (~~6~~10):

$$UC_{ECMWFk} = \frac{\sqrt{(\{(UC_{EDGAR1})_{corr}\}_{ln} \cdot E_{EDGAR1})^2 + (\{(UC_{EDGAR2})_{corr}\}_{ln} \cdot E_{EDGAR2})^2 + \cdots + (\{(UC_{EDGARn})_{corr}\}_{ln} \cdot E_{EDGARn})^2}}{|E_{EDGAR1} + E_{EDGAR2} + \cdots + E_{EDGARn}|}, \qquad (\sout{5}9)$$

$$E_{ECMWFk} = E_{EDGAR1} + E_{EDGAR2} + \cdots + E_{EDGARn}, \qquad (\sout{6}10)$$

where UC_{ECMWFk} and E_{ECMWFk} – combined uncertainty and total emissions per group k; *1,2,…,n* – EDGAR emission sectors that are combined in a particular ECMWF group k; $\{(UC_{EDGAR1})_{corr}\}_{ln}, \{(UC_{EDGAR2})_{corr}\}_{ln}, \ldots, \{(UC_{EDGARn})_{corr}\}_{ln}$ are in %. Combined group uncertainties are country-specific, because they take into account sector budget and adjust uncertainty values accordingly.

3.2.6 Calculating mean and standard deviation of lognormally distributed ECMWF group uncertainty

Finally, we needed to ensure a log-normal distribution of CO_2 emissions. Upper and lower uncertainty half-range values per ECMWF group k $ECMWFk$ are descriptive, but not straight forward to use for emission perturbations in ensemble runs or flux inversions, where mean and standard deviation of the distribution are usually used. The lower and upper bounds of the 95 % probability range, which are the 2.5^{th} and 97.5^{th} percentiles respectively, calculated assuming a log-normal distribution based on a corrected estimated uncertainty half-range from an error propagation approach, are lower and upper uncertainty values. Taking this into account and using the Z-table[3] for 2.5^{th} and 97.5^{th} percentiles p, mean μ^{ln} and standard deviation σ^{ln} of log-normal distribution can be calculated following Eq. (~~7~~11):

$$Z_p = \frac{ln([E_{ECMWFk}]_p) - \mu^{ln}_{ECMWFk}}{\sigma^{ln}_{ECMWFk}}, \qquad (\sout{7}11)$$

where the following variables are known:

$$p = 2.5 \Rightarrow Z_{2.5} = -1.96,\ [E_{ECMWFk}]_{2.5} = E_{ECMWFk} \cdot \left(1 + \frac{[UC_{ECMWFk}]_{low}}{100}\right), \qquad (\sout{8}12)$$

$$p = 97.5 \Rightarrow Z_{97.5} = 1.96,\ [E_{ECMWFk}]_{97.5} = E_{ECMWFk} \cdot \left(1 + \frac{[UC_{ECMWFk}]_{high}}{100}\right), \qquad (\sout{9}13)$$

then simple system could be composed and solved accordingly following Eq. (~~10~~14) and Eq. (~~11~~15):

$$\mu^{ln}_{ECMWFk} = ln(E_{ECMWFk}) + \frac{1}{2} ln\left(1 + \frac{[UC_{ECMWFk}]_{low}}{100}\right) + \frac{1}{2} \sout{ln\left(1 + \frac{[UC_{ECMWFk}]_{high}}{100}\right),}$$

$$\sout{(10} ln\left(1 + \frac{[UC_{ECMWFk}]_{high}}{100}\right), \qquad (14)$$

$$\sigma^{ln}_{ECMWFk} = \frac{ln\left(1 + \frac{[UC_{ECMWFk}]_{low}}{100}\right) - ln\left(1 + \frac{[UC_{ECMWFk}]_{high}}{100}\right)}{-3.92}, \qquad (\sout{11}15)$$

where $[UC_{ECMWFk}]_{low}$ and $[UC_{ECMWFk}]_{high}$ are in %.

3.2.7 Example of uncertainty calculation

Table 5 shows a step-by-step example of how yearly uncertainties are calculated. Example shows calculations for TRANSPORT group, that consists of several EDGAR emission sectors (one EDGAR sector consists even of several IPCC activities). Example shows two countries with different statistical infrastructure development levels (country with well-developed statistical infrastructures is Germany, country with less well-developed statistical infrastructures is the Russian Federation) and significant differences in emission budgets.

Table 5a: Preparatory step for yearly uncertainty calculation – data collection, same values are applied for all countries of the same type, namely for countries with well- (WDS) and less well-developed (LDS) statistical infrastructures

Country	ECMWF	IPCC (2006)	IPCC	Note	Typical fuel	Emission	Activity

~~[3] The Z-table is a mathematical table for the values of the cumulative distribution function of the normal distribution.~~

(Type)	group	activities per EDGAR sector	(2006) activity			factor uncertainty		data uncertainty	
						Low	Up	Low	Up
Germany (WDS)	TRANSPORT	1.A.3.b	1.A.3.b	Road transportation	most typical emission factor	2.0	2.0	5.0	5.0
		1.A.3.d	1.A.3.d	Water-borne navigation	composition of 80 % diesel and 20 % residual fuel oil	2.1	1.1	5.0	5.0
		1.A.3.c, 1.A.3.e	1.A.3.c	Railways	diesel	2.0	0.9	5.0	5.0
			1.A.3.e	Other transportation – Pipeline	none (suggested to neglect)	0.0	0.0	0.0	0.0
				Other transportation – Off-road	most typical emission factor	2.0	2.0	50.0	100.0
Russian Federation (LDS)	TRANSPORT	1.A.3.b	1.A.3.b	Road transportation	most typical emission factor	5.0	5.0	5.0	5.0
		1.A.3.d	1.A.3.d	Water-borne navigation	composition of 80 % diesel and 20 % residual fuel oil	2.1	1.1	50.0	50.0
		1.A.3.c, 1.A.3.e	1.A.3.c	Railways	diesel	2.0	0.9	5.0	5.0
			1.A.3.e	Other transportation – Pipeline	none (suggested to neglect)	0.0	0.0	0.0	0.0
				Other transportation – Off-road	most typical emission factor	5.0	5.0	50.0	100.0

Table 5b: First part of yearly uncertainty calculation – same values are applied for all countries of the same type, namely for countries with well- (WDS) and less well-developed (LDS) statistical infrastructures

Country (Type)	IPCC (2006) activities per EDGAR sector	IPCC (2006) activity	Combined uncertainty per IPCC (2006) activity, see Eq. (1)		Combined uncertainty per EDGAR sector, see Eq. (2)		Corrected combined uncertainty per EDGAR sector, see Eq. (3)-(4)	
			Low	Up	Low	Up	Low	Up
Germany (WDS)	1.A.3.b	1.A.3.b	5.4	5.4	5.4	5.4	5.4	5.4
	1.A.3.d	1.A.3.d	5.4	5.1	5.4	5.1	5.4	5.1
	1.A.3.c, 1.A.3.e	1.A.3.c	5.4	5.1	50.3	100.1	50.3	106.9
		1.A.3.e	0.0	0.0				
			50.0	100.0				
Russian Federation (LDS)	1.A.3.b	1.A.3.b	7.1	7.1	7.1	7.1	7.1	7.1
	1.A.3.d	1.A.3.d	50.0	50.0	50.0	50.0	50.0	50.0
	1.A.3.c, 1.A.3.e	1.A.3.c	5.4	5.1	50.5	100.3	50.5	107.0
		1.A.3.e	0.0	0.0				
			50.2	100.1				

Table 5c: Second part of yearly uncertainty calculation – values are specific per each geographical entity, take into account country type, namely if country has well- (WDS) or less well-developed (LDS) statistical infrastructure, and countries emission budget (values are from CHE_EDGAR-ECMWF_2015)

Country (Type)	IPCC (2006) activities per EDGAR sector	Emission budget 2015 per EDGAR sector, Mton	Uncertainty with assumed lognormal distribution per EDGAR sector, see Eq. (5)-(8)		Emission budget 2015 per ECMWF group, Mton	Grouped uncertainty with assumed lognormal distribution per ECMWF group, see Eq. (9)-(10)		Lognormal parameters of grouped uncertainty with assumed lognormal distribution per ECMWF group, see Eq. (14)-(15)	
			Low	Up		Low	Up	mean	standard deviation
Germany (WDS)	1.A.3.b	139.6	5.4	5.4	143.0	5.3	5.7	11.9	0.0
	1.A.3.d	1.0	5.4	5.1					
	1.A.3.c, 1.A.3.e	2.3	40.3	135.5					
Russian	1.A.3.b	131.7	7.1	7.1	206.9	14.1	44.8	12.3	0.1

Federation (LDS)	1.A.3.d	7.4	40.1	57.2				
	1.A.3.c, 1.A.3.e	67.9	40.5	135.7				

3.3 Monthly uncertainties

For Earth system modelling and data assimilation purposes a sub-yearly time scale is more appropriate. Monthly profiles of anthropogenic emissions are available and used in air quality models and are more certain than the sub-monthly profiles. The monthly profiles used in EDGARv4.3.2 are standardised to 12 monthly shares per EDGAR sector and per region (i.e. Northern temperate zone, Equator, Southern temperate zone). They do not take into account the specificity of a single year and are not varying within a geographical entity (country). We used these global yearly and monthly emission maps for 2010 to calculate for each month a multiplication factor per 0.1°×0.1° grid-cell of the sector-specific maps. Then multiplication factors were combined with CHE_EDGAR-ECMWF_2015 maps and monthly country- and sector-specific CO_2 emission budgets are calculated.

Uncertainties for monthly budgets are obviously larger than yearly ones and instead of one standard deviation σ (Quilcaille et al, 2018) two or three standard deviations, 2σ or 3σ respectively are commonly used (Oda et al., 2018; Andres et al., 2014; Andres et al., 2011). We decided to be more analytical:

1) to use the same procedure as for annual uncertainty calculation but base it on monthly emission budgets (i.e. uncertainties for IPCC activities are combined to EDGAR sectors with error propagation method, corrected for systematic underestimation by error propagation method, and adapted to have log-normal distribution~~).~~; see Eq. (1)-(8)). Obtained monthly uncertainties are the same or even smaller than the yearly ones, because empirical equations applied use emission budgets, which are smaller for individual months compared to the yearly values;

2) to calculate the correlation α (an uncertainty boosting parameter) between yearly and monthly uncertainties based on an analysis of the variations over the different months following Eq. (~~12~~16):

$$(E_{YEAR} \cdot UC_{YEAR})^2 = \alpha^2 \cdot \cancel{((E_{MONTH1} \cdot UC_{MONTH1})^2 + (E_{MONTH2} \cdot UC_{MONTH2})^2 + \cdots + (E_{MONTH12} \cdot UC_{MONTH12})^2)} \quad \cancel{,}$$

$$\cancel{(12}((E_{MONTH1} \cdot UC_{MONTH1})^2 + (E_{MONTH2} \cdot UC_{MONTH2})^2 + \cdots + (E_{MONTH12} \cdot UC_{MONTH12})^2), \qquad (16)$$

where E and UC correspond to sectoral emission budget and uncertainty in kton and % respectively, $YEAR, MONTH1, MONTH2, \ldots, MONTH12$ – yearly and monthly (January, February, …, December) values. Eq. (~~12~~16) is based on the rule for combining uncorrelated uncertainties under addition of the error propagation equation (see Eq. (~~5~~9)) and assumption that each month's uncertainty should be enhanced (boosted) by the same value;

3) to multiply the prior yearly uncertainties from Table ~~3~~4 by the boosting parameter (specific per country and emission sector) and use the result as a first guess of monthly prior uncertainties;

4) to iterate calculation steps 1) to 3) in order to find the best boosting parameter (to have the best fit between yearly and combined 12-month uncertainties) from Eq. (16) for each country and emission sector. Once the best boosting parameter was found (i.e. maximum difference between α from previous iteration and the current one over all countries and emission

sectors became less than acceptable threshold) calculated monthly uncertainties per each EDGAR sector were grouped into 7 ECMWF groups and log-normal distribution of CO_2 emissions was ensured~~.~~, see Eq. (9)-(15).

Figure ~~2~~3 has simplified roadmaps for yearly and monthly uncertainty calculations.

~~3.4 Covariance matrices~~

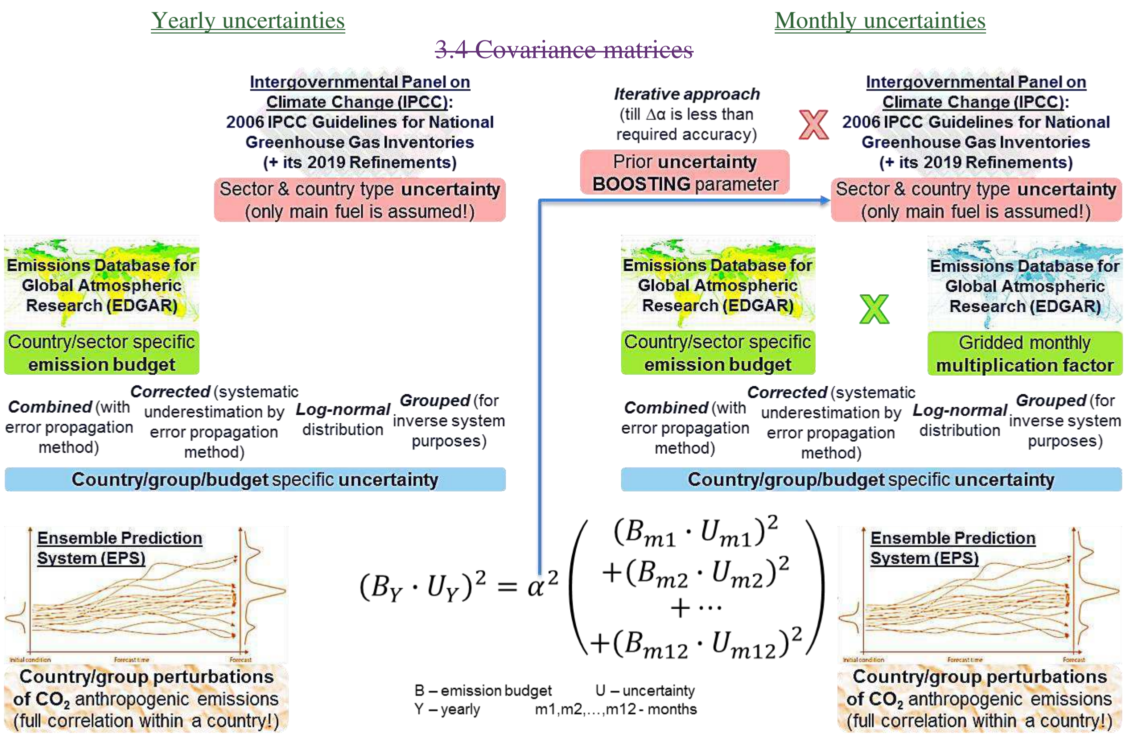


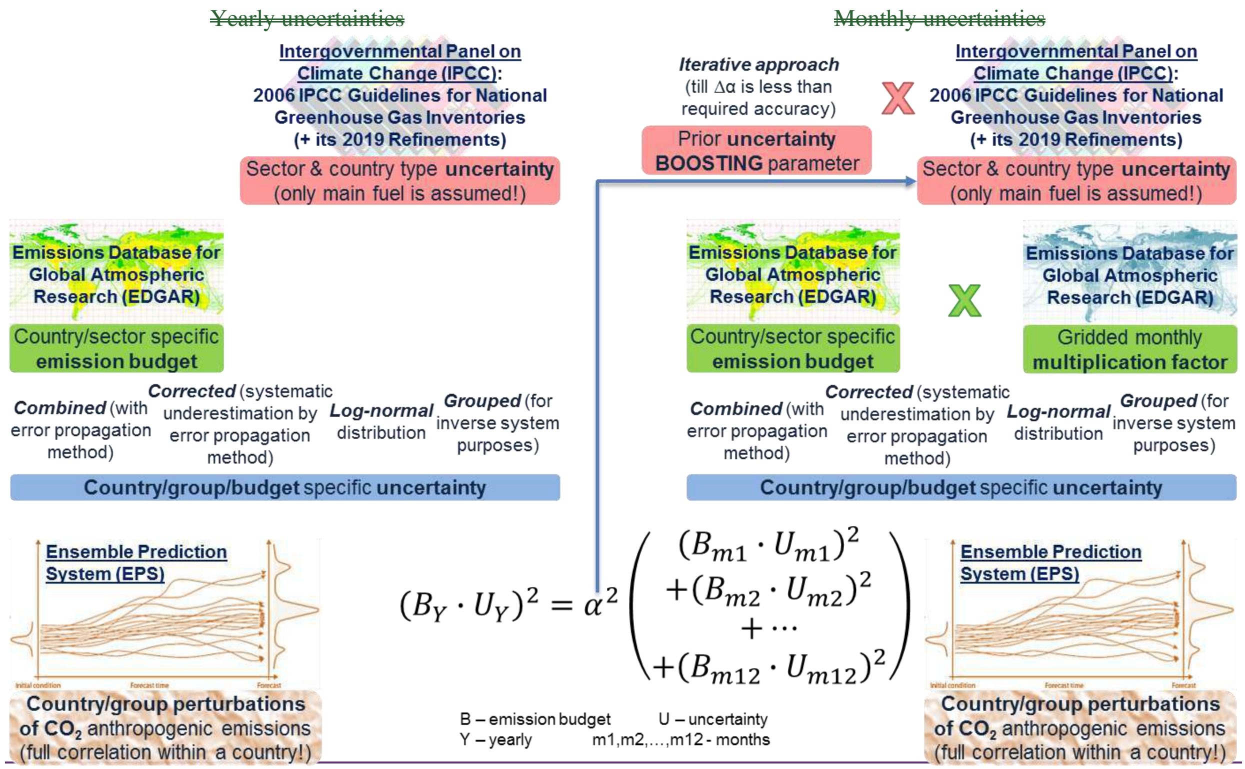
Figure 3: Simplified roadmaps for yearly (left) and monthly (right) uncertainty calculation and their relation (bottom)

The prior error covariance matrix of the emission inventory is required as an input to the inversion system. According to the IPCC (2006) all anthropogenic CO_2 emissions are assumed to be fully uncorrelated, hence the prior error correlations between grid-cell emissions from the same sector should be assumed negligible if country- and/or sector-specific information is lacking. ~~Only by assuming full absence of correlation it is possible to calculate emission uncertainties for each geographical entity and group of sectors with rather limited globally available information.~~ For the first implementation, ECMWF group covariance matrices per each geographical entity have the same representation – emission group is fully correlated with itself and fully uncorrelated with any other group. ~~Table 4 shows~~For an example ~~for Europe (28 members till end 2019).~~

~~Table 4: Representation~~ of ECMWF group covariance matrices~~, example for Europe (E28) with diagonal values being log-normal variances ·10^-5^~~

~~E28 Group of sectors~~	~~ENERGY_S~~	~~ENERGY_A~~	~~MANUFACTURING~~	~~SETTLEMENTS~~	~~AVIATION~~	~~TRANSPORT~~	~~OTHER~~
~~ENERGY_S~~	~~26.2~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~
~~ENERGY_A~~	~~0.0~~	~~2286.8~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~
~~MANUFACTURING~~	~~0.0~~	~~0.0~~	~~3435.5~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~
~~SETTLEMENTS~~	~~0.0~~	~~0.0~~	~~0.0~~	~~1518.3~~	~~0.0~~	~~0.0~~	~~0.0~~
~~AVIATION~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.1~~	~~0.0~~	~~0.0~~
~~TRANSPORT~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~473.8~~	~~0.0~~
~~OTHER~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~0.0~~	~~9472.5~~

see Table S4 from the Supplementary Information, section S.3.



~~**Figure 2: Simplified roadmaps for yearly (left) and monthly (right) uncertainty calculation and their relation (bottom)**~~

Due to the lack of information available to properly characterize the error correlations and error variances in the inventory, a refinement of those prior statistics will be carried out in a follow-on paper (~~Busserez et al. in preparation~~Bousserez, 2019) using atmospheric CO_2 observations. For this, the maximum likelihood of the prior error standard deviations and error correlation lengths will be estimated following approaches described in Wu et al. (2013).

3.4 Gridding uncertainties

Calculated yearly and monthly uncertainties per country and sector were assigned to each grid-box on the global map. National uncertainties were applied uniformly across each country. Figure 4 shows an example of the upper and lower uncertainty limits of anthropogenic CO_2 emission flux for TRANSPORT group. It should be noted that uncertainties related to the spatial distribution (representativeness of the proxy data and their uncertainty) should be much higher than the ones presented in this study. This research does not address uncertainties related to the spatial distribution. In the future we plan to address these uncertainties too. For example, following Oda et al. (2019) to characterize spatial patterns of the disaggregation errors in our emission maps.

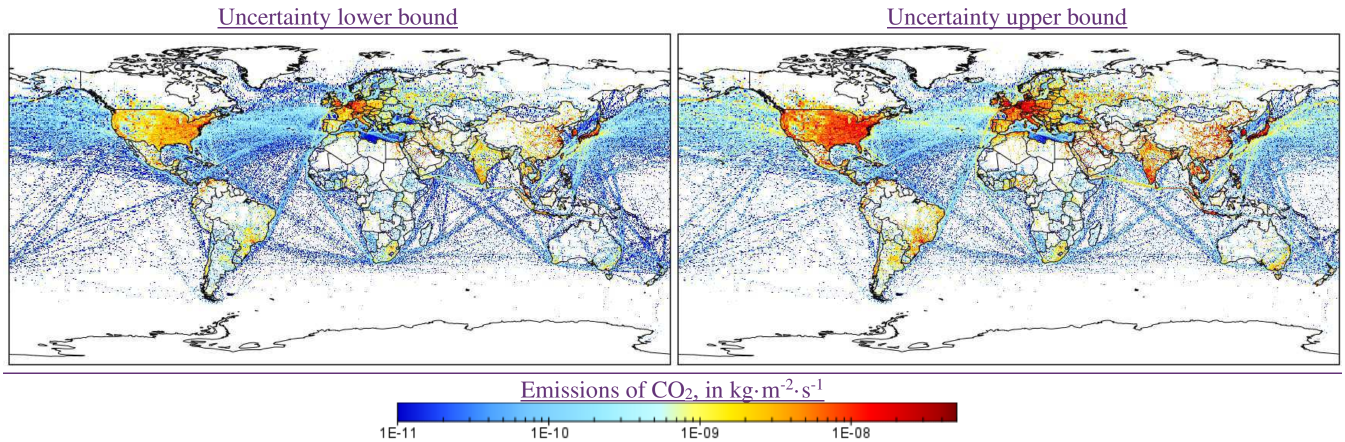


Figure 4: CO_2 emission flux uncertainties (lower (left) and upper (right) bounds) for TRANSPORT group in $kg \cdot m^{-2} \cdot s^{-1}$

4 Generated dataset

CHE_EDGAR-ECMWF_2015 data (Choulga et al., 2020) are freely available https://doi.org/10.5281/zenodo.3967439, and consist of 11 grid-maps in NetCDF format and one Excel file with information on anthropogenic CO_2 emissions and their uncertainties. For detailed information on each file see Table 6.

Table 6: Detailed information on CHE_EDGAR-ECMWF_2015 data

File	General note	Field / Spreadsheet
Lower_Uncertainties_Percentage_0.1	file has 2×8 fields with annual upper and lower uncertainty bounds in % per each emission group and for all groups summed together on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where	"Lower" – lower uncertainty bound (2.5th percentile of log-normal distribution) for yearly emissions, in %
		"Upper" – upper uncertainty bound (97.5th percentile of log-normal distribution) for yearly emissions, in %
		"Sector" – emission sector numerical name. "0" represents emission group ENERGY_S (with IPCC (2006) activity 1.A.1.a (subset)) standing for power industry emissions from

	values represent centre of the grid-cell	super emitting power plants; “1” group ENERGY_A (1.A.1.a (rest), 4.C) – power industry emissions from standard emitting power plants, & solid waste incineration; “2” group MANUFACTURING (1.A.2, 2.C.1, 2.C.2, 2.C.3, 2.C.4, 2.C.5, 2.C.6, 2.C.7, 2.D.1, 2.D.2, 2.D.4, 2.A.1, 2.A.2, 2.A.3, 2.A.4, 2.B.1, 2.B.2, 2.B.3, 2.B.4, 2.B.5, 2.B.6, 2.B.8) – combustion for manufacturing (including autoproducers), & iron and steel production, & non-ferrous metals production, & non energy use of fuels, & non-metallic minerals production, & chemical processes; “3” group SETTLEMENTS (1.A.4, 1.A.5.a, 1.A.5.b.i, 1.A.5.b.ii) – energy for buildings, residential heating; “4” group AVIATION (1.A.3.a_CRS, 1.A.3.a_CDS, 1.A.3.a_LTO) – aviation cruise, & climbing and descent, & landing and take off; “5” group TRANSPORT (1.A.3.b, 1.A.3.d, 1.A.3.c, 1.A.3.e) – road transportation, & shipping, & railways, pipelines, off-road transport; “6” group OTHER (1.A.1.b, 1.A.1.c, 1.A.5.b.iii, 1.B.1.c, 1.B.2.a.iii.4, 1.B.2.a.iii.6, 1.B.2.b.iii.3, 1.B.2.a.ii, 1.B.2.a.iii.2, 1.B.2.a.iii.3, 1.B.2.b.ii, 1.B.2.b.iii.2, 1.B.2.b.iii.4, 1.B.2.b.iii.5, 1.C, 1.B.1.a, 3.C.2, 3.C.3, 3.C.4, 3.C.7, 2.D.3, 2.B.9, 2.E, 2.F, 2.G) – oil refineries and transformation industry, & fuel exploitation, & coal production, & agricultural soils, & solvents and products use; “7” represents all groups summed together
Monthly_Upper_Lower_Uncertainties_Percentage_0.1_0.1.nc	file has 2×8×12 fields with monthly upper and lower uncertainty bounds in % per each emission group and for all groups summed together on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	file structure is identical to the file Annual_Upper_Lower_Uncertainties_Percentage_0.1_0.1.nc, but per month (1, 2, …, 12 correspond to January, February, …, December)
Annual_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×8 fields with annual emissions, and upper and lower uncertainty bounds in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ per each emission group and for all groups summed together on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	“Sup_lower” – lower uncertainty bound (2.5th percentile of log-normal distribution) for yearly emissions of ENERGY_S group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, “Sup_upper” – upper uncertainty bound (97.5th percentile of log-normal distribution) for yearly emissions of ENERGY_S group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, “Sup_flux” – yearly emissions of ENERGY_S group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Ene_lower”, “ene_upper”, “ene_flux” – same, but for ENERGY_A group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Man_lower”, “man_upper”, “man_flux” – same, but for MANUFACTURING group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Set_lower”, “set_upper”, “set_flux” – same, but for SETTLEMENTS group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Avi_lower”, “avi_upper”, “avi_flux” – same, but for AVIATION group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Tra_lower”, “tra_upper”, “tra_flux” – same, but for TRANSPORT group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Oth_lower”, “oth_upper”, “oth_flux” – same, but for OTHER group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “All_lower”, “all_upper”, “all_flux” – same, but for all groups summed together, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$
Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ per ENERGY_S emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	“Sup_lower” – lower uncertainty bound (2.5th percentile of log-normal distribution) for monthly emissions of ENERGY_S group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Sup_upper” – upper uncertainty bound (97.5th percentile of log-normal distribution) for monthly emissions of ENERGY_S group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Sup_flux” – monthly emissions of ENERGY_S group, in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ “Month” – month numerical name, where 1, 2, …, 12 correspond to January, February, …, December
Monthly_Ene_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ per ENERGY_A emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	file structure is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with “ene_lower”, “ene_upper”, “ene_flux” fields

Monthly_Man_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per MANUFACTURING emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	file structure is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with "man_lower", "man_upper", "man_flux" fields
Monthly_Set_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per SETTLEMENTS emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	file structure is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with "set_lower", "set_upper", "set_flux" fields
Monthly_Avi_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per AVIATION emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	file structure is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with "avi_lower", "avi_upper", "avi_flux" fields
Monthly_Tra_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per TRANSPORT emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	file structure is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with "tra_lower", "tra_upper", "tra_flux" fields
Monthly_Oth_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per OTHER emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	file structure is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with "oth_lower", "oth_upper", "oth_flux" fields
Monthly_All_Upper_Lower_Uncertainties_0.1_0.1.nc	file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} for all groups summed together on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell	file structure is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with "all_lower", "all_upper", "all_flux" fields
CHE_EDGAR_2015.xlsx	file has 16 spreadsheets with listed information per country (metadata, emissions, uncertainties, statistical parameters)	"COUNTRY" – ISO Code (3-letter abbreviation of a geographical entity), Geographical name (name of a geographical entity), Type (development level of countries statistical infrastructure, meaning with well-/less well-developed statistical infrastructure), Main country (dependency, which country geographical entity in question belongs to), Full information (full name of a geographical entity, and what territory it occupies on the map of this study)

		"GROUP" – № (number of anthropogenic CO_2 emission group), ECMWF group (group name), IPCC (2006) activity (IPCC activities that are included in each group), Note (short explanation of the group), Global emission budget 2015, Mton (total global emissions per group), Prior uncertainty bounds, % (initial, calculated purely based on assumptions from IPCC, lower and upper uncertainty bounds for countries with well-/less well-developed statistical infrastructures)
		"YEARLY" – ISO Code (3-letter abbreviation of a geographical entity), ECMWF group (group name), Budget, kton (yearly anthropogenic CO_2 emission budget per group and total per geographical entity), Uncertainty bounds, % (calculated based on Prior uncertainty bounds and Budgets yearly uncertainties per group and total per geographical entity, uncertainties lower/upper/symmetrical bounds), Contribution to total countries uncertainty, % (share of each group in geographical entities total yearly uncertainty, total contribution is always 100 %), Parameters of log-normal distribution (anthropogenic CO_2 emission distribution is assumed to be log-normal, so additionally for modelling purposes log-normal mean, log-normal standard deviation and log-normal variance were calculated)
		"MONTHLY_01", "MONTHLY_02", …, "MONTHLY_12" – same explanation as for spreadsheet "YEARLY", but for a month (01, 02, …, 12 correspond to January, February, …, December)

5 Comparison and ~~discussion~~sensitivity

Calculated emissions and uncertainties of fossil CO_2 have been compared to other data sets based on the country-specific data reported to UNFCCC and on fuel-specific data reported in the energy statistics of IEA. The global values and their uncertainty at a 2σ range for the CHE_EDGAR-ECMWF_2015 dataset show the lowest value of -4.7/+9.6 % or ±7.1 % range, see Table 7. This result might be attributed to the methodology, in particular considering that (i) all calculations were done at the country level and then aggregated to global level assuming no correlation following IPCC (2006), (ii) all calculations were done separately for upper and lower uncertainty bounds to preserve original information with asymmetric confidence intervals for large uncertainties (not required for the Approach 1 described in IPCC (2006), according to Approach 1 from IPCC (2006) only higher uncertainty value of asymmetric interval should be used – leads to artificial inflation of uncertainty upper or lower limit), and (iii) might be also because in this study we were not taking into account proxy grid-map uncertainties.

Table 7: Comparison of global anthropogenic CO_2 emission uncertainty at 2σ associated with certain emission datasets

Name	Global uncertainty at 2σ, %	References
BP	no quantitative assessment of uncertainty associated with its emissions dataset	Andrew (2020)
CDIAC	±8.4 %	Andres et al. (2016)
CEDS	no quantitative assessment of uncertainty associated with its emissions dataset, limited information in	Hoesly et al. (2018)
CHE_EDGAR-ECMWF_2015	±7.1 % (-4.7/+9.6 %)	Andrew (2020)
EDGAR	±9.0 %	Janssens-Maenhout et al. (2019)
EIA	no quantitative assessment of uncertainty associated with its emissions dataset	Andrew (2020)

Global Carbon Project (GCP)	±10.0 %	Friedlingstein et al. (2019)
IEA	no quantitative assessment of uncertainty associated with its emissions dataset	Andrew (2020)
ODIAC	±8.4 %[4]	Oda et al. (2018)

In this paper we decided to focus on some ~~of the~~specific geographical areas – chosen to be among most emitting in total or per emission group, most typical or most influential for a certain region. A list of these geographical entities and development levels of their statistical infrastructures are presented in Table ~~5~~8.

Table ~~5~~8: List of selected geographical entities with their statistical infrastructure's development levels

ISO Code	Geographical name	Type
GLB	All World Countries	~~ALL~~mixed-developed statistical infrastructure
E28	Europe (28 members till end 2019)	~~WDS~~well-developed statistical infrastructure
DEU	Germany	~~WDS~~well-developed statistical infrastructure
ESP	Spain	~~WDS~~well-developed statistical infrastructure
FRA	France	~~WDS~~well-developed statistical infrastructure
GBR	United Kingdom	~~WDS~~well-developed statistical infrastructure
POL	Poland	~~WDS~~well-developed statistical infrastructure
BRA	Brazil	~~LDS~~less well-developed statistical infrastructure
CHN	China	~~WDS~~well-developed statistical infrastructure
IDN	Indonesia	~~LDS~~less well-developed statistical infrastructure
IND	India	~~WDS~~well-developed statistical infrastructure
JPN	Japan	~~WDS~~well-developed statistical infrastructure
RUS	Russian Federation	~~LDS~~less well-developed statistical infrastructure
USA	United States of America	~~WDS~~well-developed statistical infrastructure

~~4~~5.1 Global versus country-specific results

In order to see how development level of country's or geographical entity's statistical infrastructure is influencing emission uncertainty of that country or geographical entity itself and (possibly) ~~global one~~the globe, uncertainty calculations for

[4] The difference between ODIAC and CDIAC gridded data is 3.3-5.7 % (Oda et al., 2018).

selected entities were performed twice – with their original and inverse types (i.e. ~~WDS becomes LDS~~country with well-developed statistical infrastructure becomes country with less well-developed statistical infrastructure and vice versa). More details on geographical entity's statistical infrastructure development level (e.g. how it was determined) are given in the Supplementary Information, section S.4. Figure ~~3~~5 shows sectoral emission budgets, uncertainties and contributions in percentage to the total uncertainty of country or geographical entity with its original and inverse statistical infrastructure development levels. The biggest impact of development level change can be noticed for countries with larger emission budgets. On average total uncertainties of selected countries (see Table ~~5~~8) changed by 1-2 %; group uncertainties changed in line with prior uncertainties from Table ~~3~~4 and countries emission budgets~~:~~, as reported in Table 9.

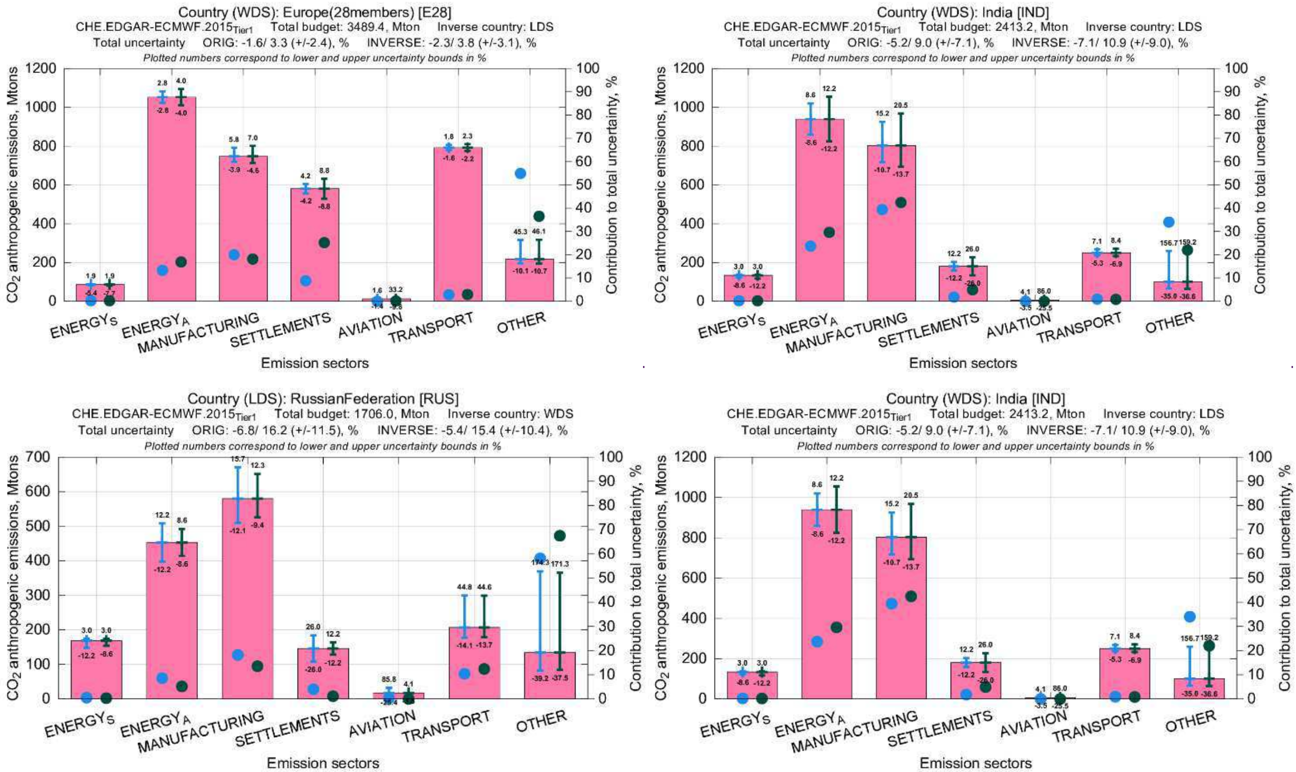
- ~~most substantial uncertainty changes for SETTLEMENTS group (consists only from residential heating emissions) with rather high differences in prior uncertainties for WDS and LDS, ±12.2 % and ±26.0 % respectively;~~
- ~~strongly changes – for MANUFACTURING and ENERGY_A (and ENERGY_S where present) groups as their budgets usually make a significant part of country's total emission budget. The MANUFACTURING group is globally mainly composed from combustion for manufacturing with rather low prior uncertainty (±8.6 % and ±12.2 % for WDS and LDS respectively) and non-metallic minerals production with much higher uncertainties (±70.9 % and ±93.0 % for WDS and LDS respectively). It also contains emissions from very uncertain non-energy use of fuels (±121.7 % and ±124.0 % for WDS and LDS respectively) and chemical processes (-107.8/+89.9 % both for WDS and LDS) emissions, though their global share in this group is ~7.0 %. The ENERGY_A group is composed of emissions from average power plants with rather low uncertainties (±8.6 % and ±12.2 % for WDS and LDS respectively) and solid waste incineration with much higher uncertainties (±40.3 % and ±41.2 % for WDS and LDS respectively). For the Globe the ratio of solid waste incineration to energy emissions is ~1/100, which keeps the total ENERGY_A group prior uncertainty quite low ±3.5 % (NB! geographical entities with higher ratios will have higher uncertainties). The ENERGY_S group has emissions from super power plants only with rather low prior uncertainties (-8.6/+3.0 % and -12.2/+3.0 % for WDS and LDS respectively) for all geographical entities;~~
- ~~mildly changes – for TRANSPORT group which globally is mainly composed of road transportation with rather low uncertainty (±5.4 % and ±7.1 % for WDS and LDS respectively) and shipping emissions (NB! all international shipping is included in All World Countries) with low uncertainties -5.4/+5.1 % for WDS and high uncertainties ±50.0 % for LDS countries. In addition, this group contains rather uncertain railways, pipelines and off-road transport emissions (~ -50.4/+107.0 % for both WDS and LDS), though their global share in this group is ~16.0 % only;~~
- ~~small changes (though huge in % value) – for AVIATION group as its prior uncertainties change dramatically from WDS to LDS (-5.5/+6.4 % and -50.1/+106.8 % respectively), though its share in global emissions is only 2.3 % (NB! all international aviation is included in All World Countries);~~
- ~~barely changes – for OTHER group as all its components are very uncertain and usually have the same prior uncertainties for both statistical infrastructure's development levels. Its main composite globally (~78.0 %) are emissions~~

~~from oil refineries and transformation industry with prior uncertainties -54.4/+149.3 % and -57.7/+151.4 % for WDS and LDS respectively. Also, this group usually has the highest contribution to the geographical entity's total uncertainty.~~

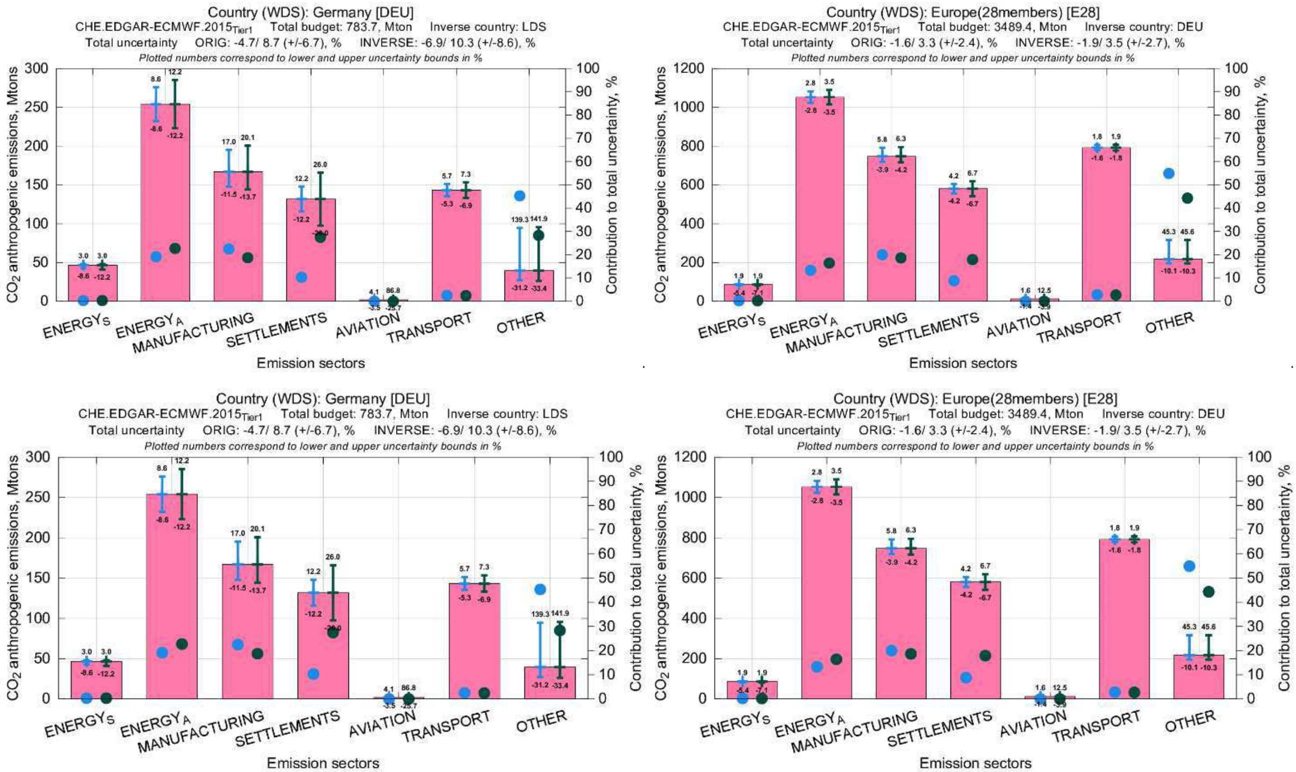
Table 9: Country's statistical infrastructure (countries with well- (WDS) and less well-developed (LDS) statistical infrastructures) influence on emission uncertainty

Impact on the uncertainty	ECMWF group	Cause description
most substantial	SETTLEMENTS	• consists only from residential heating emissions; • high differences in prior uncertainties for WDS and LDS, ±12.2 % and ±26.0 % respectively
strong	MANUFACTURING	• budget usually makes a significant part of country's total emission budget; • globally mainly composed from combustion for manufacturing with rather low prior uncertainty (±8.6 % and ±12.2 % for WDS and LDS respectively) and non-metallic minerals production with much higher uncertainties (±70.9 % and ±93.0 % for WDS and LDS respectively); • also contains emissions from very uncertain non-energy use of fuels (±121.7 % and ±124.0 % for WDS and LDS respectively) and chemical processes (-107.8/+89.9 % both for WDS and LDS) emissions, though their global share in this group is only ~7.0 %
	ENERGY_A	• budget usually makes a significant part of country's total emission budget; • composed of emissions from standard power plants with rather low uncertainties (±8.6 % and ±12.2 % for WDS and LDS respectively) and solid waste incineration with much higher uncertainties (±40.3 % and ±41.2 % for WDS and LDS respectively); • for the Globe the ratio of solid waste incineration to energy emissions is ~1/100, which keeps the total group prior uncertainty quite low ±3.5 %; • NB! geographical entities with higher ratios will have higher uncertainties
	ENERGY_S	• composed of emissions from super power plants only with rather low prior uncertainties (-8.6/+3.0 % and -12.2/+3.0 % for WDS and LDS respectively) for all geographical entities
mild	TRANSPORT	• globally mainly composed of road transportation with rather low uncertainty (±5.4 % and ±7.1 % for WDS and LDS respectively) and shipping emissions with low uncertainties -5.4/+5.1 % for WDS and high uncertainties ±50.0 % for LDS; • also contains rather uncertain railways, pipelines and off-road transport emissions (~ -50.4/+107.0 % for both WDS and LDS), though their global share in this group is ~16.0 % only; • NB! all international shipping is included in All World Countries geographical entity
small	AVIATION	• extremely high differences in prior uncertainties for WDS and LDS (-5.5/+6.4 % and -50.1/+106.8 % respectively), though this groups share in global emissions is only 2.3 %; • NB! all international aviation is included in All World Countries geographical entity
negligible	OTHER	• composed of very uncertain components with usually almost the same prior uncertainties for WDS and LDS; • main composite globally (~78.0 %) are emissions from oil refineries and transformation industry with prior uncertainties -54.4/+149.3 % and -57.7/+151.4 % for WDS and LDS respectively; • also usually has the highest contribution to the country's total uncertainty

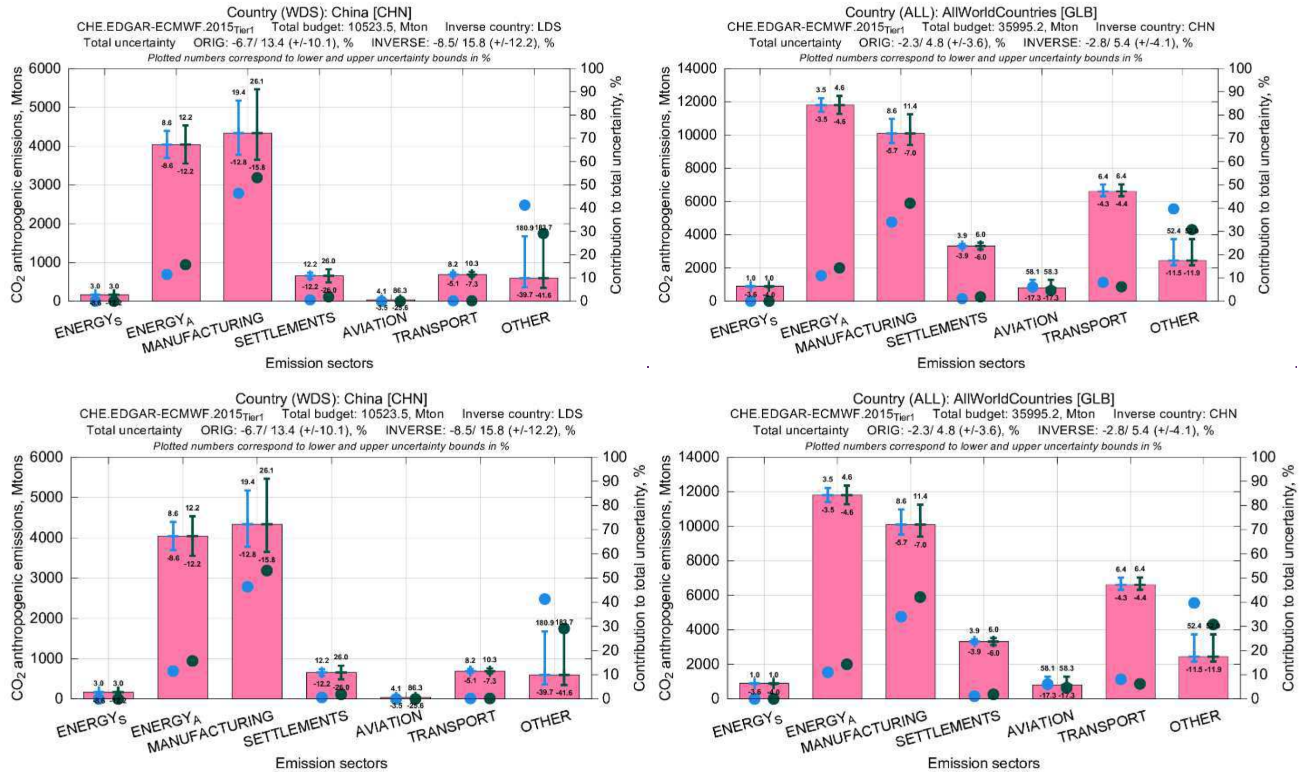
Country type impacts country itself



Country type impacts E28



Country type impacts ~~GLB~~global values



Group emission budget, in Mtons Upper and lower group uncertainty bound for countries original and inverse type, in Mtons Group contribution to countries total uncertainty for countries original and inverse type, in %

Group uncertainty 45.6, in %

Figure ~~3~~5: Emission budgets, uncertainties and contributions in percentage to the total uncertainty of the country with their original and inverse types (countries with well- (WDS) and less well-developed (LDS) statistical ~~infrastructure development types:~~infrastructures): impacting mainly country itself, e.g. ~~Europe (E28~~the Russian Federation (RUS), India (IND), impacting also Europe (E28), e.g. Germany (DEU), impacting even ~~globe (GLB),~~global values, e.g. China (CHN)

Alterations in some countries' (e.g. Germany, France) statistical infrastructure's development levels lead to changes in Europe (28 members till end 2019) uncertainties, with most substantial change for SETTLEMENTS group (e.g. 2.5 and 1.0 % respectively). Huge changes (> 10.0 %) in Europe's (28 members till end 2019) AVIATION group uncertainty % value can be due to the variation of statistical infrastructure development level for Germany, United Kingdom, France or Spain, though this groups contribution to the Europe's (28 members till end 2019) total uncertainty remains negligible. Alterations in statistical infrastructure development levels for China or the United States of America modify even global uncertainties because these countries substantially contribute to the global emission budget~~—~~, e.g. China emits ~1/3 of the global anthropogenic CO_2 budget and can change global total uncertainty up to 0.5 %.

~~4~~5.2 Yearly and monthly uncertainties

In order to increase the emission temporal resolution, monthly emissions and their uncertainties were calculated combining yearly emissions, monthly multiplication factors, and adapted uncertainty calculation methodology (see Section 3.3). Prior yearly uncertainties were multiplied by dimensionless uncertainty boosting parameter α (same value for each month) to compute prior monthly uncertainties, which were further used together with monthly emission budgets for countries monthly uncertainty calculation. Monthly uncertainties (just like yearly uncertainties) are determined by empirical formulas from IPCC (2006), hence their values depend on monthly emission budgets, which relate to number of days in a month (e.g. even with a flat yearly cycle months with more days have higher emission budgets, i.e. month emissions are sum of daily values). To eliminate this dependency, we looked straight away at dimensionless uncertainty boosting parameter α, see Table ~~6~~10 for most common values for ~~WDS and LDS~~ countries with well- and less well-developed statistical infrastructures per EDGAR sectors. Boosting parameters become active ($\alpha \neq 1$) when absolute uncertainty values are ≥ 25.0 %, α increases with the increase of absolute uncertainty following third order polynomial. For lower bound uncertainties α has bigger values and steeper growth than for upper bound uncertainties (e.g. -25.0 % ≙ α = 1.5 and -124.0 % ≙ α = 2.6; +25.0 % ≙ α = 0.8 and +124.0 % ≙ α = 1.2), α behaves in the same way for ~~WDS and LDS~~ countries with well- and less well-developed statistical infrastructures. Discrepancies in different geographical entity's (country's) boosting parameters might be for several reasons, main ones are: (i) sector emissions were zero (e.g. super power plant emissions of the energy ~~(ENE)~~ sector had no emissions); (ii) sector uncertainties were ≥ 50.0 % and needed to be adapted accordingly by log-normal distribution technique (e.g. agriculture soils ~~(AGS)~~ sector with prior uncertainties -70.7/+0.0 % both for ~~WDS~~countries with well- and ~~LDS).~~ less well-developed statistical infrastructures). Most significant discrepancies in α are for ~~AGS~~agriculture soils sector (e.g. instead of lower/upper values from Table ~~6~~10 for ~~WDS~~countries with well-developed statistical infrastructures France has α = 1.8/3.1, United Kingdom – 1.8/7.2, China – 1.8/8.4, Japan – 1.8/10.8; instead of lower/upper values from Table ~~6~~10 for ~~LDS~~countries with less well-developed statistical infrastructures Brazil has α = 1.8/0.0, the Russian Federation – 1.8/5.6). ~~In general, Brazil, Indonesia and India have a very weak yearly cycle with quite high monthly uncertainties throughout the year. Globe, Europe (28 members till end 2019), Germany, Spain, France, United Kingdom, Poland, China, Japan, Russian Federation, and United States of America have more pronounced yearly cycle, most significant for SETTLEMENTS and ENERGY_A (and ENERGY_S where present) groups, and less significant for AVIATION, TRANSPORT and MANUFACTURING groups. This is in line with the monthly profiles applied in EDGARv4.3.2 for Northern and Southern temperate zones, and Equator (see Janssens-Maenhout et al. (2019)). In summer months for Northern temperate zone, a strong decrease in SETTLEMENT and ENERGY_A (and ENERGY_S where present) groups emissions was observed, a light decrease in MANUFACTURING group emissions, and a light increase in AVIATION and TRANSPORT groups emissions. This corresponds rather well with the assumption that most of the population in the Northern hemisphere must heat their houses during winter, and that they take holidays and travel more during summer.~~

Table ~~6~~10: Dimensionless (DN) boosting parameter uncertainties (lower ~~L~~ and upper ~~U~~ bounds) for ~~statistically~~countries with well- (WDS) and less well-developed (LDS) ~~countries~~statistical infrastructures

№	ECMWF group	IPCC (2006) activities per EDGAR sector	Uncertainty boosting parameter, DN			
			WDS countries		LDS countries	
			~~L~~Low	~~U~~Up	~~L~~Low	~~U~~Up
1	ENERGY_S	1.A.1.a (subset)	1.0	1.0	1.0	1.0
2	ENERGY_A	1.A.1.a (rest)	1.0	1.0	1.0	1.0
		4.C	1.8	0.8	1.9	0.8
3	MANUFACTURING	1.A.2	1.0	1.0	1.0	1.0
		2.C.1, 2.C.2	1.7	0.8	1.7	0.8
		2.C.3, 2.C.4, 2.C.5, 2.C.6, 2.C.7	2.0	0.9	2.0	0.9
		2.D.1, 2.D.2, 2.D.4	2.6	1.2	2.6	1.2
		2.A.1, 2.A.2, 2.A.3, 2.A.4	2.0	0.9	2.3	1.0
		2.B.1, 2.B.2, 2.B.3, 2.B.4, 2.B.5, 2.B.6, 2.B.8	2.4	1.0	2.4	1.0
4	SETTLEMENTS	1.A.4, 1.A.5.a, 1.A.5.b.i, 1.A.5.b.ii	1.0	1.0	1.5	0.9
5	AVIATION	1.A.3.a_CRS	1.0	1.0	1.7	1.1
		1.A.3.a_CDS	1.0	1.0	1.7	1.1
		1.A.3.a_LTO	1.0	1.0	1.7	1.1
6	TRANSPORT	1.A.3.b	1.0	1.0	1.0	1.0
		1.A.3.d	1.0	1.0	1.7	0.9
		1.A.3.c, 1.A.3.e	1.7	1.1	1.7	1.1
7	OTHER	1.A.1.b, 1.A.1.c, 1.A.5.b.iii, 1.B.1.c, 1.B.2.a.iii.4, 1.B.2.a.iii.6, 1.B.2.b.iii.3	1.7	1.4	1.8	1.4
		1.B.2.a.ii, 1.B.2.a.iii.2, 1.B.2.a.iii.3, 1.B.2.b.ii, 1.B.2.b.iii.2, 1.B.2.b.iii.4, 1.B.2.b.iii.5, 1.C	3.0	2.4	3.1	2.5
		1.B.1.a	2.5	2.2	2.5	2.2
		3.C.2, 3.C.3, 3.C.4, 3.C.7	1.8	0.0	2.0	0.0
		2.D.3, 2.B.9, 2.E, 2.F, 2.G	1.5	0.8	1.7	0.9

4In general, Brazil, Indonesia and India have a very weak yearly cycle with quite high monthly uncertainties throughout the year. Globe, Europe (28 members till end 2019), Germany, Spain, France, United Kingdom, Poland, China, Japan, the Russian Federation, and the United States of America have more pronounced yearly cycles, most significant for SETTLEMENTS and ENERGY_A (and ENERGY_S where present) groups, and less significant for AVIATION, TRANSPORT and MANUFACTURING groups. This is in line with the monthly profiles applied in EDGARv4.3.2 for Northern and Southern temperate zones, and Equator, see Janssens-Maenhout et al. (2019). In summer months for Northern temperate zone, a strong decrease in SETTLEMENT and ENERGY_A (and ENERGY_S where present) groups emissions was observed, a light decrease in MANUFACTURING group emissions, and a light increase in AVIATION and TRANSPORT groups emissions. This corresponds rather well with the assumption that most of the population in the Northern hemisphere must heat their houses during winter, and that they take holidays and travel more during summer.

5.3 Comparison with UNFCCC, TNO and other data

The CHE_EDGAR-ECMWF_2015 dataset containing 7 global gridded fossil CO_2 emission flux maps, and country- and ECMWF-group-specific emission budgets and uncertainties have been assessed with independent data. Global emission budget values from different datasets are almost never the same, therefore it is important to first identify why estimates differ between datasets – datasets might use same country-level information as primary input, nevertheless differences in inclusion, interpretation, and treatment of that data lead to diverse results in emissions; second – try to harmonise e.g. data inclusion or omission across datasets to have more clarity in the discrepancies.

For Europe (28 members till end 2019), Germany, Spain, France, United Kingdom, Poland, Japan, the Russian Federation and the United States of America emission and uncertainty data was collected from UNFCCC NIR. The aggregation of the IPCC (2006) activity-specific emissions and uncertainties into 7 ECMWF groups was done assuming no correlation, following IPCC (2006). Although IPCC (2006) has a standard table to report GHG emissions, uncertainties can be reported in less detail by a more general category (e.g. 2.D only instead of 2.D.1, 2.D.2, 2.D.3, 2.D.4), meaning information harmonization required lots of careful time-consuming country-specific technical work.

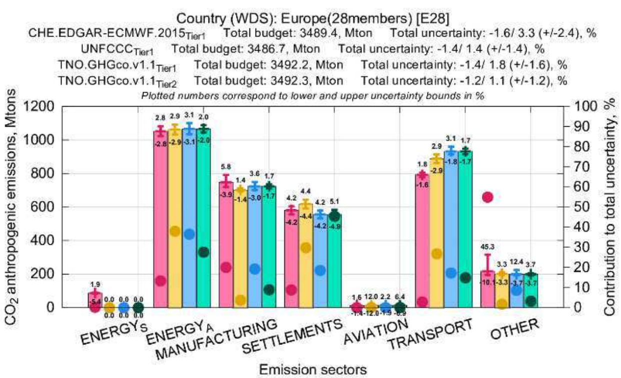
The Netherlands Organisation for Applied Scientific Research (TNO) has recently prepared the first version of their GHG and co-emitted species emission database (TNO_GHGco_v1.1) that covers the entire European domain (at 0.1°×0.05° resolution) also for CO_2 (distinguishing between fossil fuel and biofuel). Initial emission data is from the UNFCCC (Common reporting format (CRF) tables) and the European Monitoring and Evaluation Programme/Centre on Emission Inventories and Projections for air pollutants (EMEP/CEIP). These data were harmonized, checked for gaps, errors and inconsistencies, and (where needed) replaced or completed using emission data from the Greenhouse gas-Air pollution Interactions and Synergies (GAINS) model (Amann et al., 2011). Moreover, inland shipping emissions were replaced with TNO's own estimates and sea shipping is based on automatic identification system (AIS) based tracks. Expert judgement is used to assess the quality of each data source and to make choices on which source to use. The resulting emissions were checked in detail with regard to their absolute value and trends (Kuenen et al., 2014). In this study we used emission budgets from 30 TNO sectors provided by TNO (Super et al., February 2020, personal communication), and prior uncertainties calculated from IPCC (2006) and IPCC-TFI (2019) see Table ~~7~~11 (NB! all uncertainty calculations were done per country and only then for comparison purposes aggregated to Europe (28 members till end 2019) values assuming no correlation following IPCC (2006)). In addition, TNO has provided Tier 2 (Monte Carlo approach) uncertainties based on the same budgets and uncertainties from submitted NIR reports based on Tier 1 approach. The Monte Carlo simulations were done at the highest detail level (nomenclature for reporting (NFR) sector/fuel type) assuming correlations between certain sectors (for more information see Super et al. (2020)), and then emissions were aggregated to ECMWF groups assuming no correlation.

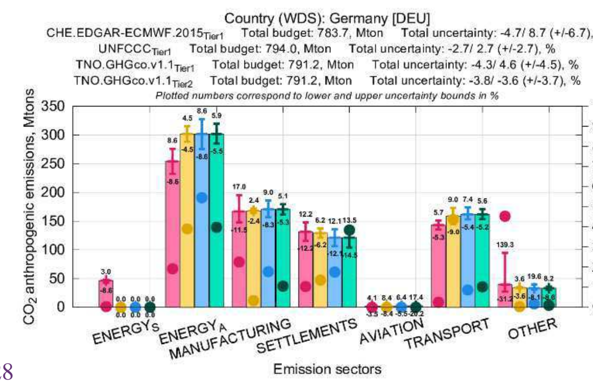
Table ~~7~~11: Prior uncertainties (lower ~~L~~ and upper ~~U~~ bounds) per each TNO emission sector based on IPCC (2006) and IPCC-TFI (2019), and aggregated to the ECMWF group uncertainties for Germany (DEU) and Europe (E28)

№	ECMWF group	IPCC (2006) activities per TNO sector	Prior uncertainty bounds, % WDS countries		Uncertainty bounds, % DEU		Uncertainty bounds, % E28	
			~~L~~Low	~~U~~Up	~~L~~Low	~~U~~Up	~~L~~Low	~~U~~Up
1	ENERGY_S	1.A.1.a (subset)	8.6	3.0	0.0	0.0	0.0	0.0
2	ENERGY_A	1.A.1.a (rest)	8.6	8.6	8.6	8.6	3.1	3.1
		4.C	40.3	40.3				
3	MANUFACTURING	1.A.2	8.6	8.6	8.3	9.0	3.0	3.6
		2.C.1, 2.C.2	37.1	37.1				
		2.C.3	10.2	10.2				
		2.C.4, 2.C.5, 2.C.6, 2.C.7	72.5	72.5				
		2.D.2	106.8	106.8				
		2.D.1, 2.D.4	50.3	50.3				
		2.A.1	36.7	36.7				
		2.A.2, 2.A.3, 2.A.4	60.7	60.7				
		2.B.1, 2.B.2, 2.B.3, 2.B.4, 2.B.5, 2.B.6, 2.B.8	107.8	89.9				
4	SETTLEMENTS	1.A.4	12.2	12.2	12.1	12.1	4.2	4.2
		1.A.5.a, 1.A.5.b.i, 1.A.5.b.ii	0.0	0.0				
5	AVIATION	1.A.3.a_CRS	5.5	6.4	5.5	6.4	1.9	2.2
		1.A.3.a_CDS	5.5	6.4				
		1.A.3.a_LTO	5.5	6.4				
6	TRANSPORT	1.A.3.b	5.4	5.4	5.4	7.4	1.8	3.1
		1.A.3.d	5.4	5.1				
		1.A.3.c	5.4	5.1				
		1.A.3.e	50.0	106.7				
7	OTHER	1.A.1.b	8.6	8.6	8.1	19.6	3.7	12.4
		1.A.1.c	12.2	12.2				
		1.A.5.b.iii, 1.B.1.c, 1.B.2.a.iii.4, 1.B.2.a.iii.6, 1.B.2.b.iii.3	0.0	0.0				
		1.B.2.a.ii, 1.B.2.a.iii.2, 1.B.2.a.iii.3, 1.B.2.b.ii, 1.B.2.b.iii.2, 1.B.2.b.iii.4, 1.B.2.b.iii.5	176.3	267.2				
		1.C	50.0	100.0				
		1.B.1.a	115.8	300.5				
		3.C.2	50.0	0.0				
		3.C.3, 3.C.4, 3.C.7	50.0	0.0				
		2.D.3, 2.B.9, 2.E, 2.F, 2.G	25.0	25.0				

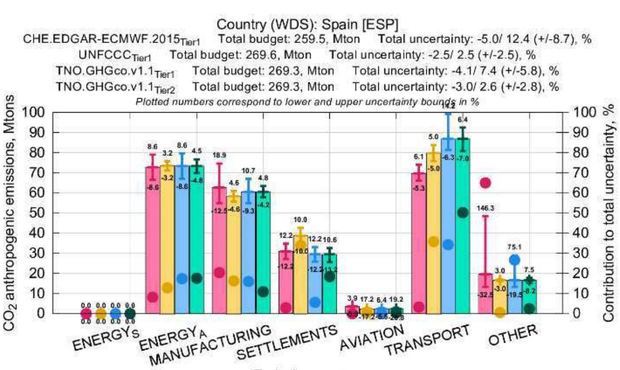
Figure ~~4~~6 shows emission budgets and uncertainties in Mtons, and contributions in % to the total geographical entity's uncertainty for Europe (28 members till end 2019), Germany, ~~Spain,~~ France~~,~~ and United Kingdom~~ and Poland~~ with their original statistical infrastructure development types based on data from CHE_EDGAR-ECMWF_2015 (in pink), UNFCCC (in yellow), and TNO_GHGco_v1.1 Tier 1 (in blue) and Tier 2 (in green~~).~~); plots for Spain and Poland are not shown here. 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. It should be noted that: (i) UNFCCC uncertainties were aggregated to ECMWF groups individually per each country as uncertainties are reported in a rather free form thus could be aggregated from different levels of precision, (ii) uncertainties for Europe (28 members till end 2019) from CHE_EDGAR-

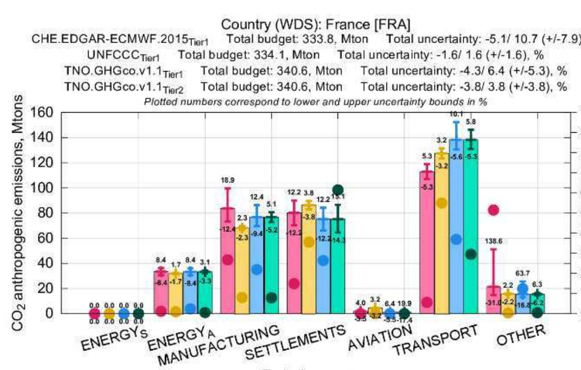
ECMWF_2015 are rather low as they were calculated by aggregating information of 28 countries, rather than assuming it to be a one geographical entity from the beginning as it is done in UNFCCC, and (iii) differences in uncertainties of CHE_EDGAR-ECMWF_2015 with other sources, especially in fuel dependent emission groups, might be due to biofuels, as CHE_EDGAR-ECMWF_2015 is not taking them into account, (NB! other datasets do take biofuels into account), and other sources (e.g. according to UNFCCC SETTLEMENT group uncertainties for United Kingdom are ±24.5 % (contributes 95 % of United Kingdom's total uncertainty), which is twice higher according to other sources – it might be explained by use of other fuels, e.g. wood and/or coal for residential heating). Differences in uncertainties between CHE_EDGAR-ECMWF_2015 and TNO_GHGco_v1.1 Tier 1 show additional value in more detailed emission budget knowledge, i.e. if we know for certain that country has no glass production then this rather uncertain activity can be excluded from non-metallic minerals production sector overall uncertainty calculation. Differences in uncertainties between TNO_GHGco_v1.1 Tier 1 and TNO_GHGco_v1.1 Tier 2 show additional value in advanced calculation technique, using a more sophisticated, data demanding Monte Carlo approach instead of simple error propagation. Overall there is quite good agreement in emission budgets and uncertainties from different sources of emission data.



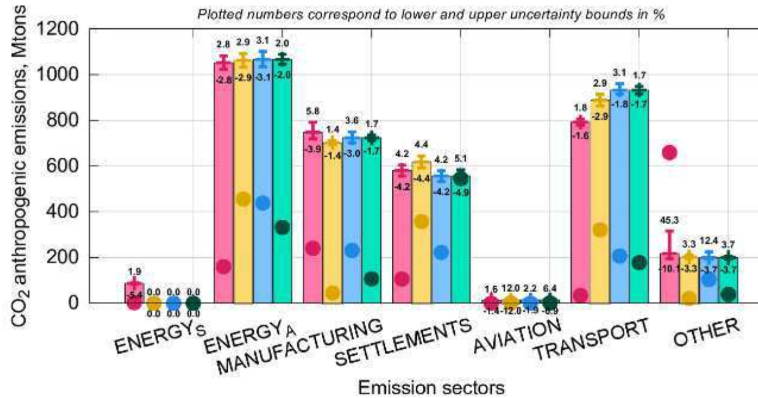


Europe (28 members till end 2019), country with well-developed statistical infrastructure

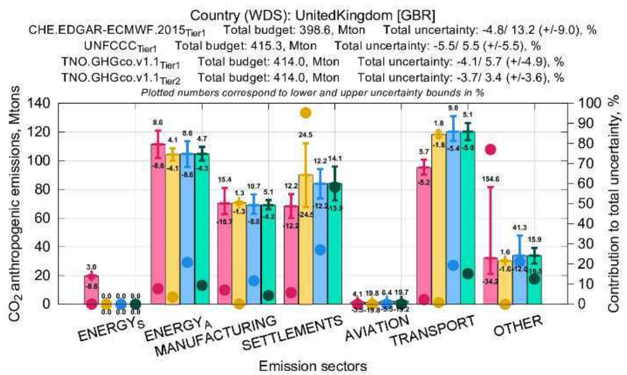


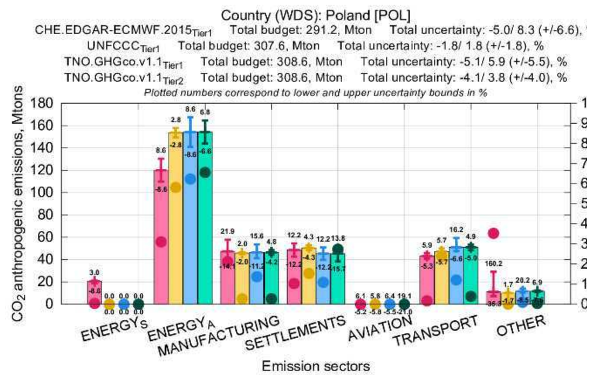


Data	Emission budget 2015, Mton	Uncertainty bound, %	
		Low/Up	Range
CHE_EDGAR-	3489.4	-1.6/+3.3	±2.4

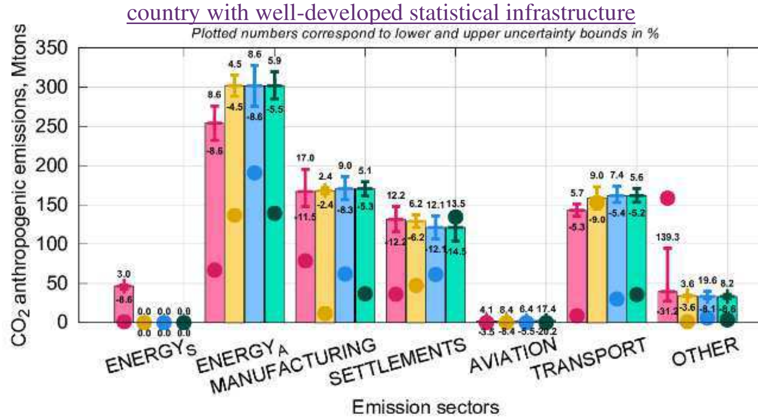


ECMWF_2015			
UNFCCC$_{Tier1}$	3486.7	-1.4/+1.4	±1.4
TNO_GHGco _v1.1$_{Tier1}$	3492.2	-1.4/+1.8	±1.6
TNO_GHGco _v1.1$_{Tier2}$	3492.3	-1.2/+1.2	±1.2



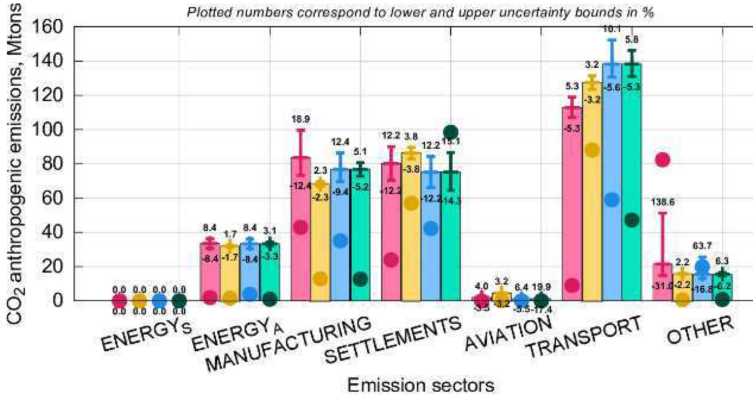


Germany, country with well-developed statistical infrastructure



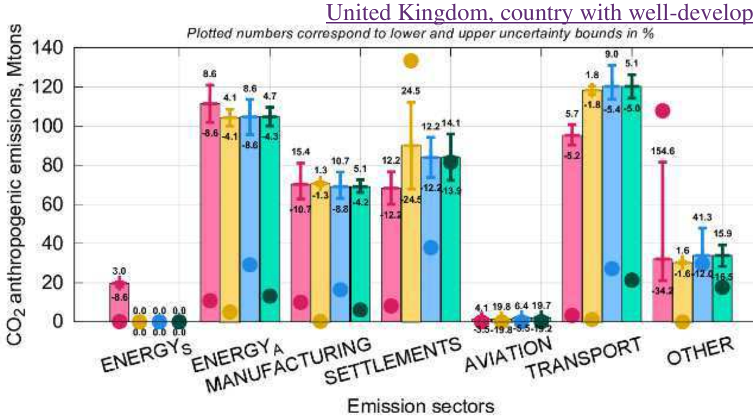
Data	Emission budget 2015, Mton	Uncertainty bound, %	
		Low/Up	Range
CHE_EDGAR-ECMWF_2015	783.7	-4.7/+8.7	±6.7
UNFCCC$_{Tier1}$	794.0	-2.7/+2.7	±2.7
TNO_GHGco _v1.1$_{Tier1}$	791.2	-4.3/+4.6	±4.5
TNO_GHGco _v1.1$_{Tier2}$	791.2	-3.8/+3.6	±3.7

France, country with well-developed statistical infrastructure



Data	Emission budget 2015, Mton	Uncertainty bound, %	
		Low/Up	Range
CHE_EDGAR-ECMWF_2015	333.8	-5.1/+10.7	±7.9
UNFCCC$_{Tier1}$	334.1	-1.6/+1.6	±1.6
TNO_GHGco_v1.1$_{Tier1}$	340.6	-4.3/+6.4	±5.3
TNO_GHGco_v1.1$_{Tier2}$	340.6	-3.8/+3.8	±3.8

United Kingdom, country with well-developed statistical infrastructure



Data	Emission budget 2015, Mton	Uncertainty bound, %	
		Low/Up	Range
CHE_EDGAR-ECMWF_2015	398.6	-4.8/+13.2	±9.0
UNFCCC$_{Tier1}$	415.3	-5.5/+5.5	±5.5
TNO_GHGco_v1.1$_{Tier1}$	414.0	-4.1/+5.7	±4.9
TNO_GHGco_v1.1$_{Tier2}$	414.0	-3.7/+3.4	±3.6

Group emission budget, in Mtons for UNFCCC, CHE_EDGAR-ECMWF_2015, TNO_GHGco_v1.1 Tier 1 & Tier 2. Group uncertainty 45.6, in %

Upper and lower group uncertainty bound, in Mtons for UNFCCC, CHE_EDGAR-ECMWF_2015, TNO_GHGco_v1.1 Tier 1 & Tier 2

Group contribution to countries total uncertainty, in % for UNFCCC, CHE_EDGAR-ECMWF_2015, TNO_GHGco_v1.1 Tier 1 & Tier 2

Figure 46: Emission budgets, uncertainties and contributions in percentage to the total uncertainty for Europe (E28), Germany (DEU), ~~Spain (ESP),~~ France (FRA~~),~~) and United Kingdom (GBR~~) and Poland (POL~~) with their original statistical infrastructure development types

Emission budgets, Tier 1 uncertainties, and contributions in percentage to the total geographical entity's uncertainty for Japan, the Russian Federation and the United States of America from CHE_EDGAR-ECMWF_2015 could be compared only with UNFCCC data (plots not shown here). UNFCCC uncertainties are usually lower than the ones calculated in this study. Main reason for that is use of country-specific emission data and ~~AD~~activity data uncertainties, which are lower than default values suggested by IPCC (2006) and IPCC-TFI (2019). Only for fuel dependent groups (e.g. AVIATION) UNFCCC uncertainties might be higher than in this study as rather uncertain biofuels might be taken into account~~.~~ (NB! CHE_EDGAR-ECMWF_2015 does not take biofuels into account). Also, emission budgets reported to UNFCCC show some differences from the ones from CHE_EDGAR-ECMWF_2015. For Japan group budgets agree rather well, and total budget difference is ~1.0 %. For the Russian Federation major differences are in ENERGY_A (and ENERGY_S) and

MANUFACTURING groups, which results in ~6.0 % higher total budget of CHE_EDGAR-ECMWF_2015. For the United States of America major differences are ~200 Mton and ~100 Mton for SETTLEMENTS and OTHER groups respectively, which results in ~4.0 % higher total budget than based on UNFCCC data.

Recent comparison of different gridded global datasets by Andrew (2020) pointed out that only few of these datasets provide quantitative uncertainty assessment, see summary in Table 8. Comparing to other global emission uncertainty values CHE_EDGAR-ECMWF_2015 shows lowest values – it might be rather deceptive as all calculations were done at the country level and then aggregated to global level assuming no correlation following IPCC (2006), we have also calculated separately upper and lower uncertainty bounds to preserve as much initial information as possible especially of asymmetric confidence intervals for large uncertainties although it is not required by the Approach 1 methodology (according to Approach 1 from IPCC (2006) only higher uncertainty value of asymmetric interval should be used – leads to artificial inflation of uncertainty upper or lower limit); on the other hand it might be also because in this study we were not taking into account proxy grid-map uncertainties. Proxy grid-map uncertainties can be rather easily added on top of calculated uncertainties by the end user.7. Comparing to other global emission uncertainty values CHE_EDGAR-ECMWF_2015 shows lowest values mainly due to the aggregation technique.

Table 8: Comparison of global anthropogenic CO_2 emission uncertainty at 2σ associated with certain emission dataset

Name	Global uncertainty at 2σ, %
BP	no quantitative assessment of uncertainty associated with its emissions dataset
CDIAC	±8.4 %
CEDS	no quantitative assessment of uncertainty associated with its emissions dataset, limited information in Hoesly et al. (2018)
CHE_EDGAR-ECMWF_2015	±7.1 % (-4.7/+9.6 %)
EDGAR	±9.0 %
EIA	no quantitative assessment of uncertainty associated with its emissions dataset
Global Carbon Project (GCP)	±10.0 %
IEA	no quantitative assessment of uncertainty associated with its emissions dataset

45.4 Sensitivity to the fuel specificity

As mentioned above, for transport related emission uncertainty calculations only the most typical fuel type (for aviation, railways, shipping) and EFemission factor uncertainty (for road and off-road transport) were used, because detailed fuel consumption information per IPCC activity was not available for this study. EDGAR dataset development team do have specific fuel information globally, which could be used for uncertainty calculation. EDGAR dataset with incorporated fuel-specific ADactivity data and EFemission factor uncertainties and Tier 1 approach for uncertainty calculation (see Supplementary Information, section S.5) hereinafter referred to as EDGAR-JRC. Country budget uncertainties were calculated by considering “full fuel” splitting and by taking into consideration the assumption that EFemission factor from sectors sharing the same fuel are fully correlated. This latter assumption transformed the sum in quadrature of Eq. (2) into a linear summation (Bond et al., 2004; Bergamaschi et al., 2015). The uncertainty of ~~AD~~activity data were set in accordance with IPCC (2006) guidelines, in the range 5.0 to 10.0 % for combustion activities, 10.0 to 20.0 % for combustion in the residential sector, 25.0 % for bunker fuels in the marine transport, 35.0 % for industrial processes of cement, lime, glass, ammonia (the range of uncertainty values refers to the 95 % confidence interval of the mean, assigned separately to ~~WDS and LDS~~ countries with well- and less well-developed statistical infrastructures). Uncertainties from EDGAR-JRC dataset aggregated to the ECMWF group level were compared with the ones from CHE_EDGAR-ECMWF_2015, see Table ~~9~~12 for Europe (28 members till end 2019) and all world countries~~ (GLB),~~, and Table ~~S6~~S8 from the Supplementary Information, section S.5, for all the rest geographical entities from Table ~~5~~8. NB! Group contribution to the geographical entity's (country's) total uncertainty is zero when group has no emissions. Emission uncertainties from EDGAR-JRC reflect the share of fuel composing the emission of each country and are in line with the estimates by CHE_EDGAR-ECMWF_2015 for those countries where the fuel-composite uncertainty is closer to the average value assigned (see Table ~~3~~4). Uncertainties calculated with fuel-specific data are usually smaller; when prevailing fuel coincides with typical fuel type from CHE_EDGAR-ECMWF_2015 emission group uncertainties from both sources are quite similar. It should be noted here that: (i) countries total uncertainty is higher in EDGAR-JRC due to aggregation technique (full correlation is assumed), (ii) AVIATION group uncertainties are higher in EDGAR-JRC due to prior aggregation of all three aviation connected sectors (cruise, climbing & descent, and landing & take off).

Table ~~9~~12: Aggregated to the ECMWF group level uncertainties (lower ~~L~~ and upper ~~U~~ bounds) in % and contributions in % to the total uncertainty (CV) for Europe (E28) and globe (GLB) from EDGAR-JRC (with extra fuel type knowledge) and CHE_EDGAR-ECMWF_2015 (with typical fuel only)

Country	ECMWF group	EDGAR-JRC			CHE_EDGAR-ECMWF_2015		
		~~L~~Low, %	~~U~~Up, %	CV, %	~~L~~Low, %	~~U~~Up, %	CV, %
GLB	ENERGY_S	0.0	0.0	0.0	-3.6	1.0	0.0
	ENERGY_A	-2.9	2.7	42.4	-3.5	3.5	11.0
	MANUFACTURING	-4.3	4.3	41.3	-5.7	8.6	34.0
	SETTLEMENTS	-2.5	2.5	1.9	-3.9	3.9	1.1
	AVIATION	-4.2	5.8	0.5	-17.3	58.1	6.1
	TRANSPORT	-2.5	2.6	7.7	-4.3	6.4	8.1
	OTHER	-5.9	6.2	6.2	-11.5	52.4	39.7
	TOTAL	*-4.8*	*4.8*	*100.0*	*-2.3*	*4.8*	*100.0*
E28	ENERGY_S	0.0	0.0	0.0	-5.4	1.9	0.2
	ENERGY_A	-2.0	2.4	56.4	-2.8	2.8	13.3
	MANUFACTURING	-2.2	2.2	12.6	-3.9	5.8	20.0
	SETTLEMENTS	-2.5	2.5	15.1	-4.2	4.2	8.8
	AVIATION	-2.4	2.8	0.0	-1.4	1.6	0.0
	TRANSPORT	-1.3	1.3	7.2	-1.6	1.8	2.8
	OTHER	-5.0	5.0	8.7	-10.1	45.3	54.9
	TOTAL	*-3.3*	*3.6*	*100.0*	*-1.6*	*3.3*	*100.0*

The uncertainties derived in this study are an upper bound of the uncertainty estimation compared to the uncertainties calculated with more detailed information, as done by the countries and reported to UNFCCC or to the uncertainties calculated with fuel-specific data. Even though sometimes differences might be quite high in %, they are usually quite small in Mtons. Taking into account that fuel data is not publicly available, requires a lot of time to collect and implement, and is not available globally – it was decided not to use it in this study for Tier 1 uncertainty calculations.

45.5 Atmospheric sensitivity to nationally disaggregated emissions

The gridded emissions are a required input to the ECMWF model used to simulate atmospheric CO_2 globally (Agusti-Panareda et al., 2014; Agusti-Panareda et al., 2019). Ideally, uncertainties at a grid-cell level would be preferred by the models, which is a difficult time-consuming task. In order to check if these calculations are necessary it was decided to run some experiments. High-resolution (~25 km horizontal resolution, 137 vertical levels) simulations with ECMWF Integrated Forecasting System (IFS) model have been performed to assess the atmospheric sensitivity to fully resolved emissions compared to nationally smoothed (global emission budget is conserved), see Figure 57.

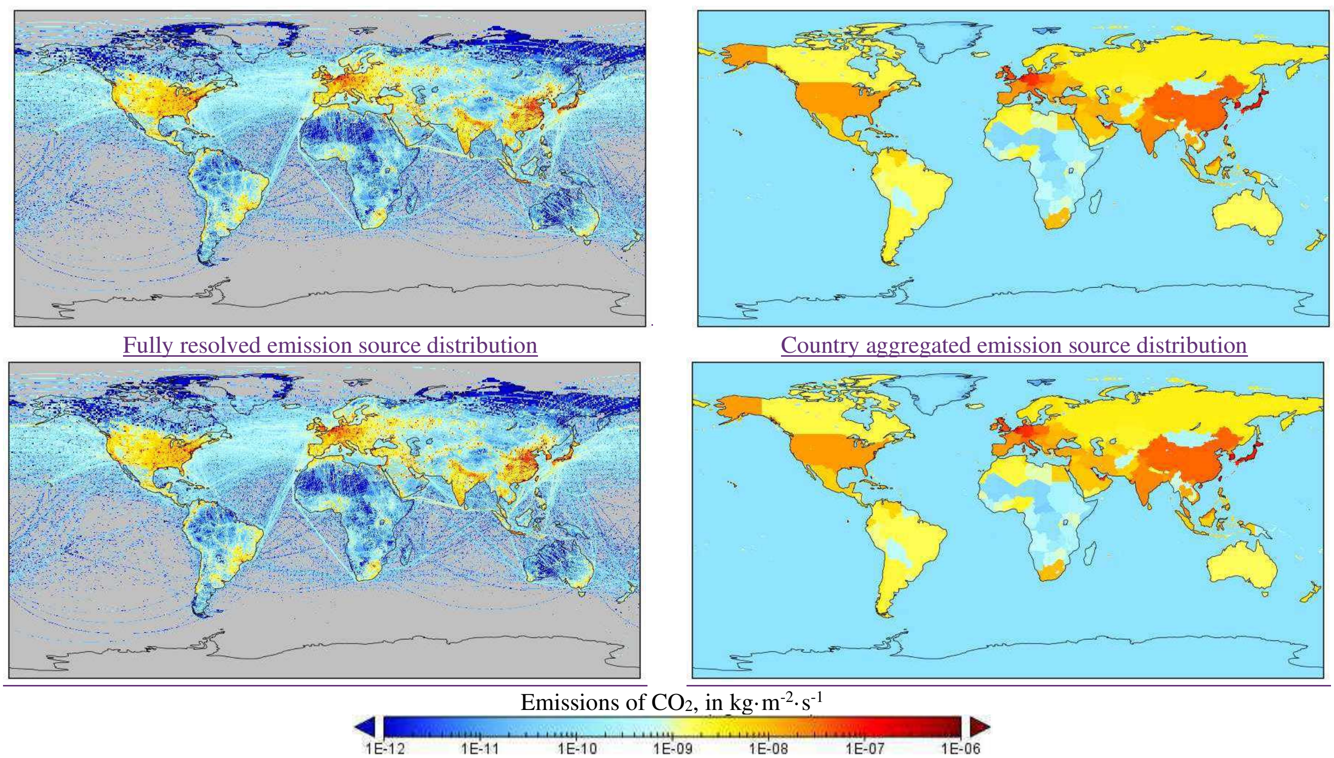


Figure 57: Anthropogenic CO_2 flux source distribution – fully resolved (left), country aggregated (right)

Model simulations were performed for January 2015 with 3 hourly output. Anthropogenic, fire, ocean and biogenic fluxes (large-scale model BIAS mitigated by biogenic CO_2 flux adjustment scheme BFAS) were considered. For the full model configuration description see McNorton et al. (2020). The atmospheric response to using either fully or partially resolved emissions compared with nationally smoothed emissions after a 10-day period are shown in Figure 68. It was noted that point sources (e.g. power plants, factories) can be easily detected if they comprise substantial part of countries total emission budget (e.g. in South Africa). If point sources are distributed homogeneously over the country and other areal sources are rather high as well it becomes really difficult to detect one extra/missing emitting hotspot (e.g. in Germany). China is a very good example for both cases as its western part has very little hotspots and they are easy to detect over the low emitting background, and its eastern part has lots of hotspots and high emitting areal sources which make it almost impossible to disentangle emissions from single power plant or factory from high emitting background. In general, even by resolving a single sector, in this case the energy sector (see Figure 68), a difference in the atmospheric response is evident. Differences of several ppm are detected over multiple regions, highlighting the importance of using high resolution spatially resolved emissions. With increase of both flux and transport model resolutions these differences are expected to increase further with steeper atmospheric CO_2 gradients.

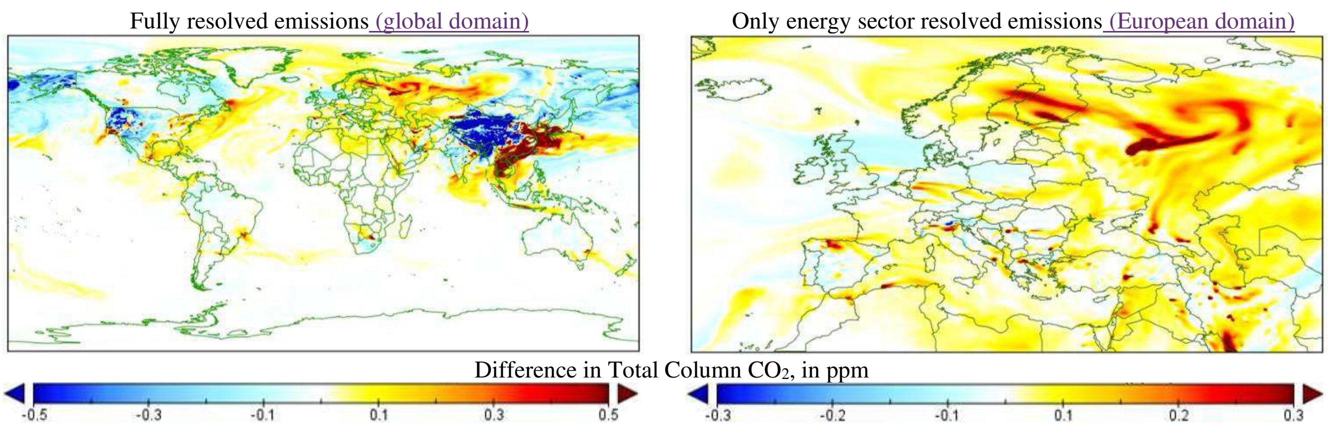
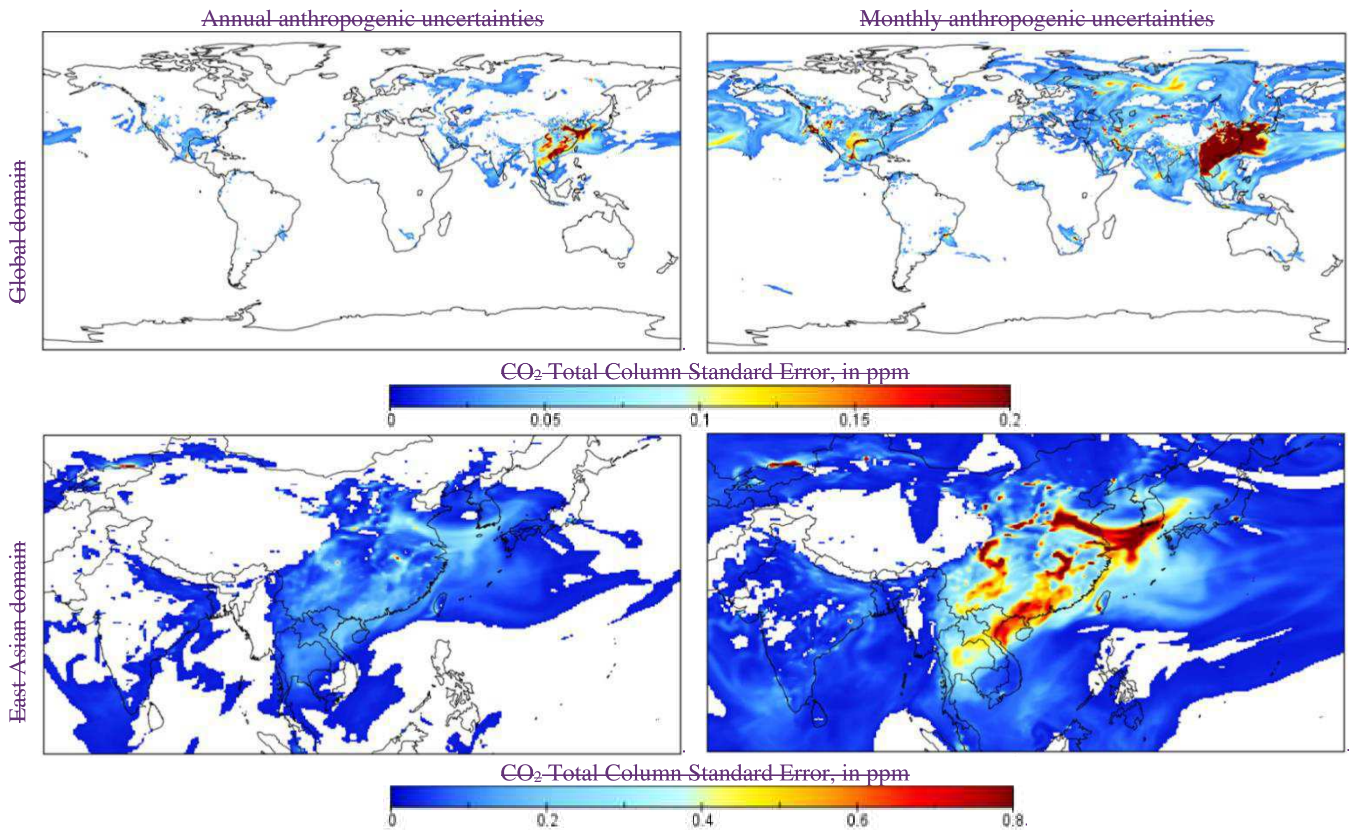


Figure 68: Difference in atmospheric response to using resolved and country aggregated emissions for January 2015 with IFS model at ~25 km resolution after 10-day simulation; the difference is calculated using both fully resolved emissions (left) or by only resolving the energy sector emissions (right)

In McNorton et al. (2020) an ensemble of the ECMWF IFS model has been used to represent the atmospheric CO_2 response to flux uncertainties derived in this study. The 50-member ensemble used spatially resolved anthropogenic, ocean, fire and biogenic fluxes. Prior anthropogenic emissions were taken from the CHE_EDGAR-ECMWF_2015 dataset and were perturbed using random noise and the log-normal yearly and monthly uncertainties reported here. They assumed that uncertainties have perfect spatial correlation within national domains and within a one-month period. No correlation was assumed between months and across different emission groups. McNorton et al. (2020) concludes that the atmospheric response to the combined anthropogenic uncertainty is between 0.1-1.4 ppm for column-averaged CO_2 over emission hot spots (see Tables 2 and 3 from McNorton et al. (2020)), these values are expected to increase further using weekly or daily

~~uncertainty estimates. Figure 7 shows error growth in column CO_2 (XCO_2) from the ensemble simulations after 10 days using yearly and monthly uncertainties (from McNorton et al. (2020)). Results show a strong atmospheric signal from monthly uncertainties over the East Asia region, which is expected to increase further globally with hourly, daily or weekly uncertainties.~~



~~**Figure 7: Error growth in column CO_2 (XCO_2) derived using IFS ensemble simulations after a 10-day period; using yearly (left) and monthly (right) uncertainties for both the global (top) and East Asian (bottom) domain**~~

~~5~~6 Conclusions and discussion

The new CHE_EDGAR-ECMWF_2015 dataset with anthropogenic fossil CO_2 emissions and their uncertainties and with a new 7×7 covariance matrix for the atmospheric transport model was compiled and tested. The fossil CO_2 emissions include all long-cycle carbon emissions from human activities, such as fossil fuel combustion, industrial processes (e.g. cement) and products use, but excludes emissions from land-use change and forestry. Human CO_2 emission inventories were processed into gridded maps to provide an estimate of prior CO_2 emissions, aggregated in 7 main emissions groups: 1) ~~power~~

~~generation~~ energy production super-emitters~~ and~~, 2) energy production ~~average~~standard-emitters, 3) manufacturing, 4) settlements, 5) aviation, 6) other transport at ground level and 7) others, with estimation of their uncertainty and covariance. For the first implementation it is assumed that each emission group is fully correlated with itself and fully uncorrelated with any other group (only diagonal values are non-zero and equal to log-normal variance). ~~A covariance matrix of 7×7 maintains the size for the inversion system to less than 50, which is adequate and computationally affordable.~~

The CHE_EDGAR-ECMWF_2015 represents the 2015 global fossil CO_2 emissions prior at 0.1º×0.1º resolution that has been for the first time to our knowledge ~~completed with full~~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 these datasets as prior input were processed such that the uncertainty information ~~with global coverage.~~could be fully taken up by the ECMWF model IFS. Estimation of emission uncertainties is purely based on IPCC (2006) and IPCC-TFI (2019) ~~EF~~emission factor and ~~AD~~activity data uncertainty values and assumptions – mainly that emissions are fully uncorrelated. Uncertainties related to the spatial distribution (representativeness of the proxy data and their uncertainty) were not assessed in this study, but they can be included by the user on top of the calculated emission uncertainties. All calculations, performed for the year 2015, are documented so that the methodology and algorithms used can be easily adapted for any other year. The dataset can be directly used in inverse modelling, and ensemble data assimilation applications, such as those envisaged within the Copernicus Atmosphere Monitoring Service (CAMS) system.

~~The CHE_EDGAR-ECMWF_2015 dataset consists of: (i) 1 grid-map with yearly anthropogenic CO_2 emission fluxes per each of 7 groups and 1 all groups summed together (total of 8 grid-maps), in $kg \cdot m^{-2} \cdot s^{-1}$; (ii) 2 grid-maps with yearly emissions upper and lower uncertainty bounds per each of 7 groups and 1 all groups summed together (total of 16 grid-maps), in %; (iii) 12 grid-maps with monthly anthropogenic CO_2 emission fluxes per each of 7 groups and 1 all groups summed together (total of 96 grid-maps), in $kg \cdot m^{-2} \cdot s^{-1}$; (iv) 2 grid-maps with monthly emissions upper and lower uncertainty bounds per each of 12 months and per each of 7 groups and 1 all groups summed together (total of 192 grid-maps), in %; (v) Excel file with listed information per country. The Excel file is organized in spreadsheets by: 1) geographical entities and their statistical infrastructure development levels, 2) emission groups with their prior upper and lower uncertainty bounds per each geographical entities level type and IPCC activities included in each group, 3) yearly and monthly emission budgets (per group and per geographical entity – total), uncertainties (per group and total), contribution of each group to total geographical entities uncertainty in %. For modelling purposes the CO_2 emission distribution is assumed to be log-normal with reported mean, standard deviation and variance (for the covariance matrices).~~

CHE_EDGAR-ECMWF_2015 consists of 11 global NetCDF files with gridded yearly and monthly upper and lower bounds of uncertainties in % and $kg \cdot m^{-2} \cdot s^{-1}$ 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).

Calculated emissions and uncertainties of fossil CO_2 have been compared to other data sets based on the country-specific data reported to UNFCCC and on fuel-specific data reported in the energy statistics of IEA. ~~Calculated emissions and~~

uncertainties of fossil CO_2 have been compared to other data sets based on the country-specific data reported to UNFCCC and on fuel-specific data reported in the energy statistics of IEA. The global values and their uncertainty at a 2σ range for the CHE_EDGAR-ECMWF_2015 dataset show the lowest value of -4.7/+9.6 % or ±7.1 % range (compared to CDIAC ±8.4 %, EDGAR ±9.0 %, GCP ±10.0 %), which is attributed to the methodology, in particular considering that (i) all calculations were done at the country level and then aggregated to global level assuming no correlation following IPCC (2006), and (ii) all calculations were done separately for upper and lower uncertainty bounds to preserve original information with asymmetric confidence intervals for large uncertainties (not required for the Approach 1 described in IPCC (2006)), but not specified for other datasets. At country level the CHE_EDGAR-ECMWF_2015 dataset providedue to the methodology used. At country level the CHE_EDGAR-ECMWF_2015 dataset provides generally larger uncertainty ranges, that are reduced when more detailed information is available to reduce the uncertainties; in summary, using the information that is uniformly available for all countries a coherent uncertainty representation is obtained.

The CHE_EDGAR-ECMWF_2015 dataset has been tested to provide the ECMWF Earth system ensemble spread to characterise the CO_2 atmospheric concentrations' uncertainties in the prototype of the Copernicus CO_2 Monitoring and Verification Support Capacity. Annual and monthly uncertainties have been evaluated in the ECMWF's atmospheric transport model IFS ensemble simulations as well as the sensitivity to the spatial distribution of anthropogenic CO_2 emissions. (McNorton et al., 2020). Results show to be rather sensitive to the spatial distribution proxies, and most updated proxies and prior uncertainties are better adapted for data assimilation applications. This needs to be studied in a future research project, the Prototype system for a Copernicus CO_2 service (CoCO_2), that follows the current CHE research project.

Contribution of representativeness errors to uncertainties and time correlation are neglected in CHE_EDGAR-ECMWF_2015 and will need to be assessed in successive future studies. The estimation of global gridded emissions with their spatially and temporally distributed uncertainties constitute the backbone for atmospheric inversions to estimate anthropogenic emissions from atmospheric concentrations (Pinty et al., 2017). Dedicated satellite missions (e.g. Copernicus anthropogenic CO_2 monitoring mission CO_2M described in Janssens-Maenhout et al. (2020)) are being planned to monitor anthropogenic emissions from space and substantially reduce emission uncertainties. The developments in the emission uncertainty based on prior knowledge computation presented in this paper is an important preparatory step for an ensemble-based CO_2 Monitoring and Verification System prototype, such as the one developed within the CHE project.

Data availability. EDGARv4.3.2 data are open access and available at http://edgar.jrc.ec.europa.eu/overview.php?v=432&SECURE=123, last access: 26 February29 June 2020, doi:https://data.europa.eu/doi/10.2904/JRC_DATASET_EDGAR, documented in Janssens-Maenhout et al. (2019). CHE_EDGAR-ECMWF_2015 data (Choulga et al., 2020) are freely available https://doi.org/10.5281/zenodo.3712339, and consist of following files with information on anthropogenic CO_2 emissions and their uncertainties: https://doi.org/10.5281/zenodo.3967439, and documented in this paper.

- Annual_Upper_Lower_Uncertainties_Percentage_0.1_0.1.nc – file has 2×8 fields with annual upper and lower uncertainty bounds in % per each emission group and for all groups summed together on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell.
  - “Lower” – lower uncertainty bound (2.5^{th} percentile of log-normal distribution) for yearly emissions, in %;
  - “Upper” – upper uncertainty bound (97.5^{th} percentile of log-normal distribution) for yearly emissions, in %;
  - “Sector” – emission sector numerical name. “0” represents emission group ENERGY_S (with IPCC (2006) activity 1.A.1.a (subset)) standing for power industry emissions from super emitting power plants; “1” group ENERGY_A (1.A.1.a (rest), 4.C) – power industry emissions from average emitting power plants, & solid waste incineration; “2” group MANUFACTURING (1.A.2, 2.C.1, 2.C.2, 2.C.3, 2.C.4, 2.C.5, 2.C.6, 2.C.7, 2.D.1, 2.D.2, 2.D.4, 2.A.1, 2.A.2, 2.A.3, 2.A.4, 2.B.1, 2.B.2, 2.B.3, 2.B.4, 2.B.5, 2.B.6, 2.B.8) – combustion for manufacturing (including autoproducers), & iron and steel production, & non-ferrous metals production, & non energy use of fuels, & non-metallic minerals production, & chemical processes; “3” group SETTLEMENTS (1.A.4, 1.A.5.a, 1.A.5.b.i, 1.A.5.b.ii) – energy for buildings, residential heating; “4” group AVIATION (1.A.3.a_CRS, 1.A.3.a_CDS, 1.A.3.a_LTO) – aviation cruise, & climbing and descent, & landing and take off; “5” group TRANSPORT (1.A.3.b, 1.A.3.d, 1.A.3.c, 1.A.3.e) – road transportation, & shipping, & railways, pipelines, off-road transport; “6” group OTHER (1.A.1.b, 1.A.1.c, 1.A.5.b.iii, 1.B.1.c, 1.B.2.a.iii.4, 1.B.2.a.iii.6, 1.B.2.b.iii.3, 1.B.2.a.ii, 1.B.2.a.iii.2, 1.B.2.a.iii.3, 1.B.2.b.ii, 1.B.2.b.iii.2, 1.B.2.b.iii.4, 1.B.2.b.iii.5, 1.C, 1.B.1.a, 3.C.2, 3.C.3, 3.C.4, 3.C.7, 2.D.3, 2.B.9, 2.E, 2.F, 2.G) – oil refineries and transformation industry, & fuel exploitation, & coal production, & agricultural soils, & solvents and products use; “7” represents all groups summed together;
- Monthly_Upper_Lower_Uncertainties_Percentage_0.1_0.1.nc – file has 2×8×12 fields with monthly upper and lower uncertainty bounds in % per each emission group and for all groups summed together on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell. File stricture is identical to the file Annual_Upper_Lower_Uncertainties_Percentage_0.1_0.1.nc, but per month (1, 2, …, 12 correspond to January, February, …, December);
- Annual_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×8 fields with annual emissions, and upper and lower uncertainty bounds in $kg \cdot m^{-2} \cdot s^{-1}$ per each emission group and for all groups summed together on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell.
  - “Sup_lower” – lower uncertainty bound (2.5^{th} percentile of log-normal distribution) for yearly emissions of ENERGY_S group, in $kg \cdot m^{-2} \cdot s^{-1}$;
  - “Sup_upper” – upper uncertainty bound (97.5^{th} percentile of log-normal distribution) for yearly emissions of ENERGY_S group, in $kg \cdot m^{-2} \cdot s^{-1}$;
  - “Sup_flux” – yearly emissions of ENERGY_S group, in $kg \cdot m^{-2} \cdot s^{-1}$;
  - “Ene_lower”, “ene_upper”, “ene_flux” – same, but for ENERGY_A group, in $kg \cdot m^{-2} \cdot s^{-1}$;
  - “Man_lower”, “man_upper”, “man_flux” – same, but for MANUFACTURING group, in $kg \cdot m^{-2} \cdot s^{-1}$;

- “Set_lower”, “set_upper”, “set_flux” – same, but for SETTLEMENTS group, in kg·m^{-2}·s^{-1};
- “Avi_lower”, “avi_upper”, “avi_flux” – same, but for AVIATION group, in kg·m^{-2}·s^{-1};
- “Tra_lower”, “tra_upper”, “tra_flux” – same, but for TRANSPORT group, in kg·m^{-2}·s^{-1};
- “Oth_lower”, “oth_upper”, “oth_flux” – same, but for OTHER group, in kg·m^{-2}·s^{-1};
- “All_lower”, “all_upper”, “all_flux” – same, but for all groups summed together, in kg·m^{-2}·s^{-1};

• Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per ENERGY_S emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell.

- “Sup_lower” – lower uncertainty bound (2.5th percentile of log-normal distribution) for monthly emissions of ENERGY_S group, in kg·m^{-2}·s^{-1};
- “Sup_upper” – upper uncertainty bound (97.5th percentile of log-normal distribution) for monthly emissions of ENERGY_S group, in kg·m^{-2}·s^{-1};
- “Sup_flux” – monthly emissions of ENERGY_S group, in kg·m^{-2}·s^{-1};
- “Month” – month numerical name, where 1, 2, …, 12 correspond to January, February, …, December;

• Monthly_Ene_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per ENERGY_A emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell. File stricture is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with “ene_lower”, “ene_upper”, “ene_flux” fields;

• Monthly_Man_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per MANUFACTURING emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell. File stricture is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with “man_lower”, “man_upper”, “man_flux” fields;

• Monthly_Set_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per SETTLEMENTS emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell. File stricture is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with “set_lower”, “set_upper”, “set_flux” fields;

• Monthly_Avi_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per AVIATION emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell. File stricture is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with “avi_lower”, “avi_upper”, “avi_flux” fields;

• Monthly_Tra_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in kg·m^{-2}·s^{-1} per TRANSPORT emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell. File stricture is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with “tra_lower”, “tra_upper”, “tra_flux” fields;

- Monthly_Oth_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in $kg \cdot m^{-2} \cdot s^{-1}$ per OTHER emission group on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell. File stricture is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with “oth_lower”, “oth_upper”, “oth_flux” fields;
- Monthly_All_Upper_Lower_Uncertainties_0.1_0.1.nc – file has 3×12 fields with monthly emissions, and upper and lower uncertainty bounds in $kg \cdot m^{-2} \cdot s^{-1}$ for all groups summed together on a regular grid with 1800 pixels along the latitude and 3600 pixels along the longitude, where values represent centre of the grid-cell. File stricture is identical to the file Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc, but with “all_lower”, “all_upper”, “all_flux” fields;
- CHE_EDGAR_2015.xlsx – file has 16 spreadsheets with listed information per country (metadata, emissions, uncertainties, statistical parameters).
  - “COUNTRY” – ISO Code (3-letter abbreviation of a geographical entity), Geographical name (name of a geographical entity), Type (development level of countries statistical system, meaning well-/less well-developed statistical system, WDS/LDS respectively), Main country (dependency, which country geographical entity in question belongs to), Full information (full name of a geographical entity, and what territory it occupies on this research map);
  - “GROUP” – № (number of anthropogenic CO_2 emission group), ECMWF group (group name), IPCC (2006) activity (IPCC activities that are included in each group), Note (short explanation of the group), Global emission budget 2015, Mton (total global emissions per group), Prior uncertainty bounds, % (initial, calculated purely based on assumptions from IPCC, lower and upper uncertainty bounds for countries with well-/less well-developed statistical systems);
  - “YEARLY” – ISO Code (3-letter abbreviation of a geographical entity), ECMWF group (group name), Budget, kton (yearly anthropogenic CO_2 emission budget per group and total per geographical entity), Uncertainty bounds, % (calculated based on Prior uncertainty bounds and Budgets yearly uncertainties per group and total per geographical entity, uncertainties lower/upper/symmetrical bounds), Contribution to total countries uncertainty, % (share of each group in geographical entities total yearly uncertainty, total contribution is always 100 %), Parameters of log-normal distribution (anthropogenic CO_2 emission distribution is assumed to be log-normal, so additionally for modelling purposes log-normal mean, log-normal standard deviation and log-normal variance were calculated);
  - “MONTHLY_01”, “MONTHLY_02”, …, “MONTHLY_12” – same explanation as for spreadsheet “YEARLY”, but for a month (01, 02, …, 12 correspond to January, February, …, December).

Author contribution. All the authors participated in the EDGAR_CHE maps generation (methodology, data generation), model experiment set-up, and analysis of the result. Margarita Choulga and Greet Janssens-Maenhout wrote the manuscript with contributions from all the other authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank Glenn Carver (ECMWF) for editorial help and assistance; Anabel Bowen (ECMWF) for invaluable help with figure design. Margarita Choulga was funded by the CO_2 Human Emissions (CHE) project which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 776186.

Financial support. This research has been supported by CHE (grant no. 776186).

References

Agustí-Panareda, A., Massart, S., Chevallier, F., Boussetta, S., Balsamo, G., Beljaars, A., Ciais, P., Deutscher, N.M., Engelen, R., Jones, L., Kivi, R., Paris, J.-D., Peuch, V.-H., Sherlock, V., Vermeulen, A.T., Wennberg, P.O., Wunch, D.: Forecasting global atmospheric CO_2, Atmos. Chem. Phys., 14, 11959-11983, doi:10.5194/acp-14-11959-2014, 2014.

Agustí-Panareda, A., Diamantakis, M., Massart, S., Chevallier, F., Muñoz-Sabater, J., Barré, J., Curcoll, R., Engelen, R., Langerock, B., Law, R.M., Loh, Z., Morguí, J.A., Parrington, M., Peuch, V.-H., Ramonet, M., Roehl, C., Vermeulen, A.T., Warneke, T., Wunch, D.: Modelling CO_2 weather – why horizontal resolution matters, Atmos. Chem. Phys., 19, 7347-7376, doi:10.5194/acp-19-7347-2019, 2019.

Amann, M., Bertok, I., Borken-Kleefeld, J., Cofala, J., Heyes, C., Höglund-Isaksson, L., Klimont, Z., Nguyen, B., Posch, M., Rafaj, P., Sandler, R., Schöpp, W., Wagner, F., Winiwarter, W.: Cost-effective control of air quality and greenhouse gases in Europe: Modelling and policy applications, Environmental Modelling and Software, Vol. 26, pp. 1489-1501, 2011.

Andres, R.J., Marland, G., Fung, I., and Matthews, E.: A 1° × 1° distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950-1990. Glob. Biogeochem. Cycles, 10, 419-429, doi:10.1029/96GB01523, 1996.

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: Chemical and Physical Meteorology, 63:3, 309-327, doi:10.1111/j.1600-0889.2011.00530.x, 2011.

Andres, R.J., Boden, T.A., and Higdon, D.: A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission, Tellus B: Chemical and Physical Meteorology, 66:1, 23616, doi: 10.3402/tellusb.v66.23616, 2014.

Andres, R.J., Boden, T.A., and Marland, G.: Annual Fossil-Fuel CO_2 Emissions: Mass of Emissions Gridded by One Degree Latitude by One Degree Longitude, United States: N. p., (NDP-058.2016), doi:10.3334/CDIAC/ffe.ndp058.2016, 2016.

Arneth, A., Sitch, S., Pongratz, J., Stocker, B., Ciais, P., Poulter, B., Bayer, A., Bondeau, A., Calle, L., Chini, L., Gasser, T., Fader, M., Friedlingstein, P., Kato, E., Li, W., Lindeskog, M., Nabel, J.E.M.S., Pugh, T.A.M., Robertson, E., Viovy, N., Yue, C., and Zaehle, S.: Historical carbon dioxide emissions caused by land-use changes are possibly larger than assumed, Nature Geoscience, 10 (2), 79, 2017.

Andrew, R.M: A comparison of estimates of global carbon dioxide emissions from fossil carbon sources, Earth Syst. Sci. Data, 12, 1437–1465, https://doi.org/10.5194/essd-12-1437-2020, 2020.

Asefi-Najafabady, S., Rayner, P.J., Gurney, K.R., McRobert, A., Song, Y., Coltin, K., Huang, J., Elvidge, C., Baugh, K.: A multiyear, global gridded fossil fuel CO_2 emission data product: Evaluation and analysis of results, J. Geophys. Res. Atmos., 119, 17, 10.213-10.231, doi:10.1002/2013JD021296, 2014.

Bastos, A., O'Sullivan, M., Ciais, P., Makowski, D., Sitch, S., Friedlingstein, P., Chevallier, F., Rödenbeck, C., Pongratz, J., Luijkx, I.T., Patra, P.K., Peylin, P., Canadell, J.G., Lauerwald, R., Li, W., Smith, N.E., Peters, W., Goll, D.S., Jain, A.K., Kato, E., Lienert, S., Lombardozzi, D.L., Haverd, V., Nabel, J.E.M.S., Poulter, B., Tian, H., Walker, A.P., and Zaehle, S.: Sources of Uncertainty in Regional and Global Terrestrial CO_2 Exchange Estimates, Global Biogeochemical Cycles, 34, 2, doi:10.1029/2019GB006393, 2020.

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.

Bergamaschi, P., Corazza, M., Karstens, U., Athanassiadou, M., Thompson, R. L., Pison, I., Manning, A. J., Bousquet, P., Segers, A., Vermeulen, A. T., Janssens-Maenhout, G., Schmidt, M., Ramonet, M., Meinhardt, F., Aalto, T., Haszpra, L., Moncrieff, J., Popa, M. E., Lowry, D., Steinbacher, M., Jordan, A., O'Doherty, S., Piacentino, S., and Dlugokencky, E.: Top-down estimates of European CH4 and N2O emissions based on four different inverse models, Atmos. Chem. Phys., 15, 715–736, https://doi.org/10.5194/acp-15-715-2015, 2015.

Bond, T. C., Streets, D. G., Yarber, K. F., Nelson, S. M., Woo, J.-H., and Klimont, Z.: A technology-based Global inventory of black and organic carbon emissions from combustion, J. Geophys. Res.,109, D14203, doi:10.1029/2003JD003697, 2004.

Bousserez, N.: Towards a Prototype Global CO_2 Emissions Monitoring System for Copernicus, arXiv preprint arXiv:1910.11727, 2019.

BP: BP Statistical Review of World Energy 2016, available at: http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html, last access: ~~26 February~~29 June 2020.

CarbonBrief, clear on carbon: Paris 2015: Tracking country climate pledges, available at: https://www.carbonbrief.org/paris-2015-tracking-country-climate-pledges, last access: ~~26 February~~29 June 2020.

Bradley, S.: Natural gas use features two seasonal peaks per year, September 2015, available at: https://www.eia.gov/todayinenergy/detail.php?id=22892, last access: 29 June 2020.

CHE: CO_2 Human Emissions (CHE) project official website, available at: https://www.che-project.eu, last access: ~~26 February~~29 June 2020.

Chen, H., Huang, Y., Shen, H., Chen, Y., Ru, M., Chen, Y., Lin, N., Su, S., Zhuo, S., Zhong, Q., Wang, X., Liu, J., Li, B., Tao, S.: Modelling temporal variations in global residential energy consumption and pollutant emissions, Applied Energy, 184, 0306-2619, 820-829, doi:10.1016/j.apenergy.2015.10.185, 2016.

Choulga, M., McNorton, J., Janssens-Maenhout, G.: CHE_EDGAR-ECMWF_2015 [Data set], Zenodo, ~~doi:10.5281/zenodo.3712339~~https://doi.org/10.5281/zenodo.3967439, 2020.

Comstock, O.: U.S. natural gas consumption has both winter and summer peaks, February 2020, available at: https://www.eia.gov/todayinenergy/detail.php?id=42815, last access: 29 June 2020.

Cong, R., Saitō, M., Hirata, R., Ito, A., and Maksyutov, S.: Uncertainty Analysis on Global Greenhouse Gas Inventories from Anthropogenic Sources, Proceedings of the 2nd International Conference of Recent Trends in Environmental Science and Engineering (RTESE'18), Niagara Falls, Canada 10-12.06.2018, Paper No. 141, doi:10.11159/rtese18.141, 2018.

Crippa, M., Oreggioni, G., Guizzardi, D., Muntean, M., Schaaf, E., Lo Vullo, E., Solazzo, E., Monforti-Ferrario, F., Olivier, J.G.J., Vignati, E.: Fossil CO_2 and GHG emissions of all world countries - 2019 Report, EUR 29849 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-11100-9, doi:10.2760/687800, JRC117610, 2019.

Crippa, M., Solazzo, E., Huang, G., Guizzardi, D., Koffi, E., Muntean, M., Schieberle, C., Friedrich, R., Janssens-Maenhout, G.: High resolution temporal profiles in the Emissions Database for Global Atmospheric Research, Sci Data 7, 121 https://doi.org/10.1038/s41597-020-0462-2, 2020.

EIA: International Energy Statistics, U.S. Energy Information Administration, Washington DC, USA, available at: https://www.eia.gov/international/overview/world, last access: ~~26 February~~29 June 2020.

Frey, H.C.: Evaluation of an Approximate Analytical Procedure for Calculating Uncertainty in the Greenhouse Gas Version of the Multi-Scale Motor Vehicle and Equipment Emissions System, Prepared for Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Ann Arbor, MI, May 30, 2003.

Friedlingstein, P., Jones, M.W., O'Sullivan, M., Andrew, R.M., Hauck, J., Peters, G.P., Peters, W., Pongratz, J., Sitch, S., Le Quere, C., Bakker, D.C.E., Canadell, J.G., Ciais, P., Jackson, R.B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L.P., Currie, K.I., Feely, R.A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D.S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R.A., Hurtt, G., Ilyina, T., Jain, A.K., Joetzjer, E., Kaplan, J.O., Kato, E., Klein Goldewijk, K., Korsbakken, J.I., Landschutzer, P., Lauvset, S.K., Lefevre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P.C., Melton, J.R., Metzl, N., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S.-I., Neill, C., Omar, A.M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rodenbeck, C., Seferian, R., Schwinger, J., Smith, N., Tans, P.P., Tian, H., Tilbrook, B., Tubiello, F.N., van der Werf, G.R., Wiltshire, A.J., Zaehle, S.: Global Carbon Budget 2019, Earth System Science Data, 11, 4, 1783-1838, doi:10.5194/essd-11-1783-2019, 2019.

Hoesly, R.M., Smith, S.J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J.J., Vu, L., Andres, R.J., Bolt, R.M., Bond, T.C., Dawidowski, L., Kholod, N., Kurokawa, J.I., Li, M., Liu, L., Lu, Z., Moura, M.C.P., O'Rourke, P.R., Zhang, Q.: Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11, 369-408, doi:10.5194/gmd-11-369-2018, 2018.

IEA: Energy Balances of OECD and non-OECD countries, International Energy Agency, Paris, Beyond 2020 Online Database, available at: http://data.iea.org, last access: ~~26 February~~29 June 2020.

IPCC: 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Eggleston, S., Buendia, L., Miwa, K., Ngara, T., and Tanabe, K. (eds.). IPCC-TSU NGGIP, IGES, Hayama, Japan. www.ipcc-nggip.iges.or.jp/public/2006gl/index.html, 2006.

IPCC-TFI, Calvo Buendia, E., Guendehou, S., Limmeechokchai, B., Pipatti, R., Rojas, Y., Sturgiss, R., Tanabe, K., Wirth, T., Romano, D., Witi, J., Garg, A., Weitz, M.M., Bofeng, C., Ottinger, D.A., Dong, H., MacDonald, J.D., Ogle, S.M., Theoto Rocha, M., Sanz Sanchez, M.J., Bartram, D.M., and Towprayoon, S.: 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Gomez, D. and Irving, W. (ed.), Vol1. Ch.8, May 2019.

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., Petrescu, A. M. R., Solazzo, E., and Oreggioni, G. D.: EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012, Earth Syst. Sci. Data, 11, 959-1002, https://doi.org/10.5194/essd-11-959-2019, 2019.

Janssens-Maenhout, G., Pinty, B., Dowell, M., Zunker, H., Andersson, E., Balsamo, G., Bézy, J.-L., Brunhes, T., Bösch, H., Bojkov, B., Brunner, D., Buchwitz, M., Crisp, D., Ciais, P., Counet, P., Dee, D., Denier van der Gon, H., Dolman, H., Drinkwater, M., Dubovik, O., Engelen, R., Fehr, T., Fernandez, V., Heimann, M., Holmlund, K., Houseling, S., Husband, R., Juvyns, O., Kentarchos, A.,. Landgraf, J., Lang, R., Löscher, A., Marshall, J., Meijer, Y., Nakajima, M., Palmer, P., Peylin, P., Rayner, P., Scholze, M., Sierk, B., and Veefkind, P.: Towards an operational anthropogenic CO_2 emissions monitoring and verification support capacity, Bull. Amer. Meteor. Soc., 0, doi:10.1175/BAMS-D-19-0017.1, 2020.

Kuenen, J.J.P., Visschedijk, A.J.H., Jozwicka, M., and Denier van der Gon, H.A.C.: TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling, Atmos. Chem. Phys., 14, 10963-10976, https://doi.org/10.5194/acp-14-10963-2014, 2014.

Le Quéré, C., Andrew, R.M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P.A., Korsbakken, J.I., Peters, G.P., Canadell, J.G., Arneth, A., Arora, V.K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L.P., Ciais, P., Doney, S.C., Gkritzalis, T., Goll, D.S., Harris, I., Haverd, V., Hoffman, F.M., Hoppema, M., Houghton, R.A., Hurtt, G., Ilyina, T., Jain, A.K., Johannessen, T., Jones, C.D., Kato, E., Keeling, R.F., Goldewijk, K.K., Landschutzer, P., Lefevre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rodenbeck, C., Schuster, U., Schwinger, J., Seferian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P.P., Tian, H., Tilbrook, B., Tubiello, F.N., van der Laan-Luijkx, I.T., van der Werf, G.R., Viovy, N., Walker, A.P., Wiltshire, A.J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, Earth System Science Data, 10, 4, 2141-2194, doi:10.5194/essd-10-2141-2018, 2018.

Liu, Z., Guan, D., Wei, W., Davis, S.J., Ciais, P., Bai, J., Peng, S., Zhang, Q., Hubacek, K., Marland, G., Andres, R.J., Crawford-Brown, D., Lin, J., Zhao, H., Hong, C., Boden, T.A., Feng, K., Peters, G.P., Xi, F., Liu, J., Li, Y., Zhao, Y., Zeng, N., He, K.: Reduced carbon emission estimates from fossil fuel combustion and cement production in China, Nature, 524, 7565, 335-338, doi:10.1038/nature14677, 2015.

McNorton, J., Bousserez, N., Agusti-Panareda, A., Balsamo, G., Choulga, M., Dawson, A., Engelen, R., Kiping, Z., and Lang, S.: Representing Model Uncertainty for Global Atmospheric CO_2 Flux Inversions Using ECMWF-IFS-46R1, Geoscientific Model Development Discussions, 2020, 1-30, doi:10.5194/gmd-2019-314, 2020.

~~Mitchell, J.F.B.: Carbon dioxide review: 1982, Edited by William C.C. Oxford University Press New York 1982, 1-469, Quarterly Journal of the Royal Meteorological Society, 110, 464, 568-569, doi:10.1002/qj.49711046421, 1984.~~

NIR: National Inventory Submissions 2018, available at: https://unfccc.int/process-and-meetings/transparency-and-reporting/reporting-and-review-under-the-convention/greenhouse-gas-inventories-annex-i-parties/submissions/national-inventory-submissions-2018, last access: ~~26 February~~29 June 2020.

NOAA State of the Climate Report 2018, Chapter 2 – Global Climate, Ed. Dunn, R.J.H., Stanitski, D.M., Gobron, N., and Willett, K.M. https://www.ametsoc.net/sotc2018/Chapter_02.pdf, 2019.

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 night-time lights, Atmospheric Chemistry and Physics, 11, doi: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., Bun, R., Kinakh, V., Topylko, P., Halushchak, M., Marland, G., Lauvaux, T., Jonas, M., Maksyutov, S., Nahorski, Z., Lesiv, M., Danylo, O., and Horabik-Pyzel, J.: Errors and uncertainties in a gridded carbon dioxide emissions inventory. Mitigation and Adaptation Strategies for Global Change, Vol. 24, 6, 1007-1050, doi:10.1007/s11027-019-09877-2, 2019.

ODIAC: ODIAC Fossil Fuel CO_2 Emissions Dataset, available at: http://www.nies.go.jp/doi/10.17595/20170411.001-e.html, last access: ~~26 February~~29 June 2020.

Olivier, J.G.J. and Janssens-Maenhout, G.: CO_2 Emissions from Fuel Combustion - 2016 Edition, IEA CO_2 report 2016, Part III, Greenhouse-Gas Emissions, ISBN 978-92-64-25856-3, 2016a.

Olivier, J.G.J., Janssens-Maenhout, G., Muntean, M., and Peters, J.A.H.W: Trends in global CO_2 emissions: 2016 report, JRC 103425, https://edgar.jrc.ec.europa.eu/news_docs/jrc-2016-trends-in-global-co2-emissions-2016-report-103425.pdf, 2016b.

Paris Agreement, official website of the European Union, available at: https://ec.europa.eu/clima/policies/international/negotiations/paris_en, last access: ~~26 February~~29 June 2020.

Paris Agreement - Status of Ratification, official website of the United Nations Framework Convention on Climate Change, available at: https://unfccc.int/process/the-paris-agreement/status-of-ratification, last access: ~~26 February~~29 June 2020.

Pinty, B., Janssens-Maenhout, G., Dowell, M., Zunker, H., Brunhes, T., Ciais, P., Dee, D., Denier van der Gon, H., Dolman, H., Drinkwater, M., Engelen, R., Heimann, M., Holmlund, K., Husband, R., Kentarchos, A., Meijer, Y., Palmer, P., Scholze, M.: An operational anthropogenic CO_2 emissions monitoring & verification support capacity - Baseline requirements, Model components and functional architecture, European Commission Joint Research Centre, EUR 28736 EN, doi:10.2760/39384, 2017.

Quilcaille, Y., Gasser, T., Ciais, P., Lecocq, F., Janssens-Maenhout, G., and Mohr, S.: Uncertainty in projected climate change arising from uncertain fossil-fuel emission factors, Environmental Research Letters, 13, 44017, http://doi.org/10.1088/1748-9326/aab304, 2018.

Super, I., Dellaert, S.N.C., Visschedijk, A.J.H., and Denier van der Gon, H.A.C.: Uncertainty analysis of a European high-resolution emission inventory of CO_2 and CO to support inverse modelling and network design, Atmospheric Chemistry and Physics, 20, 3, 1795-1816, doi:10.5194/acp-20-1795-2020, 2020.

Wang, R., Tao, S., Ciais, P., Shen, H.Z., Huang, Y., Chen, H., Shen, G.F., Wang, B., Li, W., Zhang, Y.Y., Lu, Y., Zhu, D., Chen, Y.C., Liu, X.P., Wang, W.T., Wang, X.L., Liu, W.X., Li, B.G., and Piao, S.L.: High-resolution mapping of combustion processes and implications for CO_2 emissions, Atmos. Chem. Phys., 13, 5189-5203, doi:10.5194/acp-13-5189-2013, 2013.

Wu, L., Bocquet, M., Chevallier, F., Lauvaux, T., and Davis, K.: Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions, Tellus B: Chemical and Physical Meteorology 65, no. 1: 20894, 2013.