



# Global anthropogenic CO<sub>2</sub> emissions and uncertainties as a prior for Earth system modelling and data assimilation

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

<sup>1</sup>Research Department, European Centre for Medium-Range Weather Forecasts,  
ECMWF, Reading, RG2 9AX, United Kingdom

<sup>2</sup>Joint Research Centre (JRC), European Commission, Ispra, 21027, Italy

<sup>3</sup>Department of Climate, Air and Sustainability, TNO, Utrecht, 3584 CB, the Netherlands

**Correspondence:** Margarita Choulga ([margarita.choulga@ecmwf.int](mailto:margarita.choulga@ecmwf.int))

Received: 17 March 2020 – Discussion started: 6 April 2020

Revised: 20 September 2021 – Accepted: 23 September 2021 – Published: 17 November 2021

**Abstract.** The growth in anthropogenic carbon dioxide (CO<sub>2</sub>) emissions acts as a major climate change driver, which has widespread implications across society, influencing the scientific, political, and public sectors. For an increased understanding of the CO<sub>2</sub> emission sources, patterns, and trends, a link between the emission inventories and observed CO<sub>2</sub> concentrations is best established via Earth system modelling and data assimilation. Bringing together the different pieces of the puzzle of a very different nature (measurements, reported statistics, and models), it is of utmost importance to know their level of confidence and boundaries well.

Inversions disaggregate the variation in observed atmospheric CO<sub>2</sub> concentration to variability in CO<sub>2</sub> emissions by constraining the regional distribution of CO<sub>2</sub> fluxes, derived either bottom-up from statistics or top-down from observations. The level of confidence and boundaries for each of these CO<sub>2</sub> fluxes is as important as their intensity, though often not available for bottom-up anthropogenic CO<sub>2</sub> emissions. This study provides a post-processing tool CHE\_UNC\_APP for anthropogenic CO<sub>2</sub> emissions to help assess and manage the uncertainty in the different emitting sectors. The postprocessor is available under <https://doi.org/10.5281/zenodo.5196190> (Choulga et al., 2021). Recommendations are given for regrouping the sectoral emissions, taking into account their uncertainty instead of their statistical origin; for addressing local hot spots; for the treatment of sectors with small budget but uncertainties larger than 100 %; and for the assumptions around the classification of countries based on the quality of their statistical infrastructure. This tool has been applied to the EDGARv4.3.2\_FT2015 dataset, resulting in seven input grid maps with upper- and lower-half ranges of uncertainty for the European Centre for Medium-Range Weather Forecasts Integrated Forecasting System. The dataset is documented and available under <https://doi.org/10.5281/zenodo.3967439> (Choulga et al., 2020). While the uncertainty in most emission groups remains relatively small (5 %–20 %), the largest contribution (usually over 40 %) to the total uncertainty is determined by the OTHER group (of fuel exploitation and transformation but also agricultural soils and solvents) at the global scale. The uncertainties have been compared for selected countries to those reported in the inventories submitted to the United Nations Framework Convention on Climate Change and to those assessed for the European emission grid maps of the Netherlands Organisation for Applied Scientific Research. Several sensitivity experiments are performed to check (1) the country dependence (by analysing the impact of assuming either a well- or less well-developed statistical infrastructure), (2) the fuel type dependence (by adding explicit information for each fuel type used per activity from the Intergovernmental Panel on Climate Change), and (3) the spatial source distribution dependence (by aggregating all emission sources and comparing the effect

against an even redistribution over the country). The first experiment shows that the SETTLEMENTS group (of energy for buildings) uncertainty changes the most when development level is changed. The second experiment shows that fuel-specific information reduces uncertainty in emissions only when a country uses several different fuels in the same amount; when a country mainly uses the most globally typical fuel for an activity, uncertainty values computed with and without detailed fuel information are the same. The third experiment highlights the importance of spatial mapping.

## 1 Introduction

Accurate assessment of anthropogenic carbon dioxide (CO<sub>2</sub>) emissions is important to better understand the global carbon cycle. Efforts towards a global anthropogenic CO<sub>2</sub> monitoring and verification support capacity as described by Janssens-Maenhout et al. (2020) rely on atmospheric modelling and atmospheric observations, like in situ (e.g. the Integrated Carbon Observation System, ICOS), airborne (e.g. aircraft campaigns), or spaceborne observations (e.g. the Orbiting Carbon Observatory, OCO-2, and the Greenhouse gases Observing Satellite, GOSAT). Atmospheric measurements of CO<sub>2</sub> and co-emitted species can be assimilated into flux inversion systems to provide top-down estimates of CO<sub>2</sub> fluxes at multiple spatiotemporal scales. The European Centre for Medium-Range Weather Forecasts (ECMWF), for example, aims to develop an operational inversion system to estimate CO<sub>2</sub> fluxes using observed atmospheric concentrations of CO<sub>2</sub> and other relevant species.

The global transport models require an initial best estimate of the CO<sub>2</sub> 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<sub>2</sub>. The uncertainty in these corrected CO<sub>2</sub> fluxes based on inverse modelling will be lower with the increase in CO<sub>2</sub> observations and their accuracy. The disentanglement of the fossil CO<sub>2</sub> emissions from the total atmospheric CO<sub>2</sub> emissions remains challenging. For example in 2018 total anthropogenic CO<sub>2</sub> concentrations ( $5.4 \pm 0.4$  ppm) represented only 1.3 % of the global atmospheric CO<sub>2</sub> concentration ( $407.4 \pm 0.1$  ppm) (Friedlingstein et al., 2019), which states the need for a high accuracy of measurements ( $\geq 1.0$  %).

Emission fields are often supplied through emission inventories. Bottom-up emission inventories start from human activity statistics. Emission factors are defined for each activity and provided at the international or country level (e.g. national greenhouse gas inventory report, NIR). Such inventories need to be gridded and characterized with uncertainties to represent a prior dataset useful for numerical modelling. Table 1 shows examples of most commonly used global gridded CO<sub>2</sub> emission datasets; for more details see Cong et al. (2018, Table 1), Janssens-Maenhout et al. (2019, Table 3), Andrew (2020), and Jones et al. (2021).

Only four datasets from Table 1 provide uncertainty estimates, namely CDIAC, FFDAS, PKU-FUEL, and GCP-GridFED. CDIAC uncertainties have no sectors and include contributions from the tabular fossil fuel CO<sub>2</sub> emissions (assigned per seven country types; values are constant over time), geography map (power plant location), and population map (has details in both time and space and used to distribute fossil fuel CO<sub>2</sub> emissions). Population map uncertainty strongly dominates in the generated gridded fossil fuel CO<sub>2</sub> uncertainties (Andres et al., 2016). CDIAC uncertainties have no sectoral distribution and are presented on a  $1.0^\circ \times 1.0^\circ$  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 characterization of uncertainty more complex if there were similarities in the model and observations used. PKU-FUEL uncertainty estimates of CO<sub>2</sub> 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 were randomly sampled 1000 times from an a priori normal uncertainty distribution with a certain coefficient of variation: for fuel consumptions from ships and aviation the sector coefficient of variation is set to be 20 %, for the wildfires sector 18 %, for all other fuel data 10 %, and for combustion rates 20 % (Marland et al., 2003; Marland et al., 2006; Wang et al., 2013; Oda et al., 2019). GCP-GridFED focusses strongly on the fuel disaggregation for the global CO<sub>2</sub> emissions, for which a detailed assessment of the uncertainty has not yet been published.

## 2 Methods

### 2.1 Purpose and UNFCCC context

Intercomparisons of global greenhouse gas (GHG) emission inventories were carried out (e.g. Cong et al., 2019; Petrescu et al., 2020) to better understand discrepancies and missing or lesser-known sources. The United Nations Framework Convention on Climate Change (UNFCCC) experts, reviewing national GHG inventories on a yearly basis, are keen to know which sectors or fuels need extra attention for an inventory that complies with the principles of transparency, accuracy, consistency, completeness, and comparability (TACCC principles). Discrepancies are often related to the different in-

**Table 1.** Examples of global gridded anthropogenic CO<sub>2</sub> emission bottom-up datasets.

Name	Resolution	Period	Main assumptions, uncertainties	Source
Carbon Dioxide Information Analysis Center (CDIAC)	Spatial: $1.0^{\circ} \times 1.0^{\circ}$ 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, 2016)
Open-Data Inventory for Anthropogenic Carbon dioxide (ODIAC)	Spatial: $1 \times 1 \text{ km}^2$ , $0.1^{\circ} \times 0.1^{\circ}$ Temporal: monthly Sectoral: 6	1979–2018	First introduce the combined use of nightlight data and individual power plant emission and location profiles.	Oda and Maksyutov (2011); Oda et al. (2018); ODIAC (2021)
Emissions Database for Global Atmospheric Research (EDGAR)	Spatial: $0.1^{\circ} \times 0.1^{\circ}$ 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: $0.1^{\circ} \times 0.1^{\circ}$ Temporal: annual Sectoral: 2	1997–2012	Provide gridded posterior uncertainty (version 2.2); in addition, provide monthly, weekly, and hourly fractions from annual CO <sub>2</sub> emissions.	Asefi-Najafabady et al. (2014)
Community Emissions Data System (CEDS)	Spatial: $0.1^{\circ} \times 0.1^{\circ}$ Temporal: annual, monthly Sectoral: 55	1750–2014	Provide emissions of CO <sub>2</sub> and other GHGs and pollutants.	Hoesly et al. (2018)
Peking University Fuel combustion inventory (PKU-FUEL)	Spatial: $0.1^{\circ} \times 0.1^{\circ}$ 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 Carbon Budget Gridded Fossil Emissions Dataset (GCP-GridFED)	Spatial: $0.1^{\circ} \times 0.1^{\circ}$ Temporal: monthly Sectoral: 28	1959–2018	National GHG inventories reported to UNFCCC are used for the GCP dataset, that is gridded with predefined grid maps following EDGARv4.3.2 spatial distribution proxies; also provide gridded sectoral uncertainties.	Jones et al. (2021)

interpretations of definitions or to missing information (statistics and/or measurements). When focussing on global emission datasets, which are calculated bottom-up following the Intergovernmental Panel on Climate Change (IPCC) 2006 Guidelines for National Greenhouse Gas Inventories, then the discrepancy using different definitions disappears, while the lack of information becomes strongly apparent for certain regions. More information costs time and effort when compiling a global dataset in a consistent way. Therefore, it is of paramount importance to prioritize the additional information needs and the weaknesses in the inventory with sources of large uncertainty in intensity or variability.

The IPCC has been addressing uncertainty from the beginning. Methodology, data, and data sources in this paper were taken from IPCC (2006) guidelines and their refinements (IPCC, 2019). Also, the assumptions are based on IPCC (2006), so all emissions are considered to be fully uncorrelated with activity (and so with sector and type) (i.e. all activities from IPCC (2006) are fully uncorrelated with each other) for the calculation of the uncertainty as well as of the covariance matrices.

While the UNFCCC sticks to national inventories, the atmospheric modelling community needs spatially distributed

data. This adds an extra uncertainty to the emission grid maps, not evaluated with the uncertainty in the proxy data but which needs an assessment of the representativeness of the selected proxies for distributing the emissions. The point sources, leading to large plumes, were prioritized for being treated separately with more data. These consisted of super power plants, which are defined as a large power plant or a group of closely located power plants (operating at maximum capacity and availability), causing CO<sub>2</sub> plumes from a single grid cell with a CO<sub>2</sub> flux  $\geq 7.9 \times 10^{-6} \text{ kg m}^{-2} \text{ s}^{-1}$ . According to expert knowledge, the upper-half range of uncertainty for super power plants is not larger than +3.0 %, whereas for small plants whose operation is decided based on day-to-day needs, this can reach up to +15.0 %. In this paper, 30 grid cells of  $0.1^{\circ} \times 0.1^{\circ}$  from 12 countries were identified, representing these super power generators (896.7 Mt of the energy sector) and including large plants from China, Russia, and India (for the detailed ranking of the power plant sites as a function of their emission intensity, refer to the Supplement, Sect. S1). The power plant coordinates were checked to avoid the need for an uncertainty related to their positioning. The remaining power plants (not super power generators), over

30 000, could not be checked to the same extent and therefore are recommended in a second emission group.

## 2.2 Generating uncertainty input for transport models

The uncertainty calculation methodology and initial uncertainty values (i.e. activity data and emission factor uncertainties per CO<sub>2</sub>-emitting activity) are both taken from IPCC (2006) and its refinements (IPCC, 2019). The following terminology is used to ease the explanation: “activity” – IPCC (2006) activities which result in anthropogenic CO<sub>2</sub> emissions in the yearly budget (a long-cycle carbon), “sector” – combination of different activities that are measured or reported together (that have emission budget data), “group” – combination of different sectors that have emission budget data purely for modelling or comparison needs.

In general, uncertainties are calculated in three steps: (i) sector uncertainties (based on emission factors and activity data uncertainties), (ii) annual grouped uncertainties, and (iii) monthly grouped uncertainties. By default, all calculations are performed separately for upper- and lower-half ranges of uncertainties and sector and/or group combined uncertainties, where upper- and lower-half ranges of uncertainty are in percent.

### 2.2.1 Calculating sector uncertainties

The initial 92 IPCC (2006) activity uncertainties are combined into sectors for which the user has emission budget data<sup>1</sup>, following Eqs. (1) and (2):

$$UC_{activity\_i} = \sqrt{EF_{activity\_i}^2 + AD_{activity\_i}^2}, \quad (1)$$

where combined uncertainties  $UC_{activity\_i}$  per activity  $i$  were calculated using uncertainties for emission factors  $EF_{activity\_i}$  and activity data  $AD_{activity\_i}$  in percent provided in IPCC (2006) and its refinements (IPCC, 2019);

$$UC_{sector\_j} = \sqrt{UC_{activity\_1}^2 + UC_{activity\_2}^2 + \dots + UC_{activity\_n}^2}, \quad (2)$$

where combined uncertainties  $UC_{sector\_j}$  per sector  $j$  were calculated with the error propagation method, taking into account particularly for that sector activity combined uncertainties  $UC_{activity\_1}$ ,  $UC_{activity\_2}$ , ...,  $UC_{activity\_n}$  used in percent.

### 2.2.2 Group annual uncertainties

This concerns the further grouping of the combined IPCC (2006) sectors according to the user needs into groups and

<sup>1</sup> Often, emission budgets are provided not per IPCC (2006) activity but for several activities together (usually due to measuring or reporting limitations), for which the user then needs to assume a lump sum activity, emission factor, and uncertainties in those.

calculation of group yearly uncertainties. Usually, there are computational restrictions for operational modelling: the number of emission input fields read by the model cannot be too large, or emission values are too low to be distinguishable from a global or large regional modelling perspective, so some sectors need to be merged. In addition, instantaneous local emission data as an aggregated total might be rather uncertain and hard to evaluate for different emission types all over the world. IPCC (2006) and its refinement (IPCC, 2019) provide the best possible information on how certain emissions are reported on an annual national level.

Sector uncertainties have to be adjusted to consider a country's statistical system development level and its yearly emission budget and log-normal distribution of non-negative emissions and then further combined into group uncertainties for modelling and comparison purposes in the following way (by default all calculations are performed separately for upper- and lower-half ranges of uncertainties):

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

$$(UC_{sector\_j})_{corr} =$$

$$\begin{cases} UC_{sector\_j} \cdot FC_{sector\_j}, & 100\% \leq UC_{sector\_j} \leq 230\% \\ UC_{sector\_j}, & UC_{sector\_j} < 100\% \cup UC_{sector\_j} > 230\%, \end{cases} \quad (4)$$

where corrected uncertainties  $(UC_{sector\_j})_{corr}$  per sector  $j$  were calculated to take into account large combined uncertainty ( $100\% \leq UC_{sector\_j} \leq 230\%$ ) and underestimation by the error propagation method in comparison to a Monte Carlo simulation; correction factor  $FC_{sector\_j}$  is computed based on Frey (2003), and also log-normal adjustment of the emission distribution is computed based on Frey (2003) as detailed in the Supplement, Sect. S3;

$$UC_{group\_k} = \sqrt{\frac{(\{(UC_{sector\_1})_{corr}\}_{ln} \cdot E_{sector\_1})^2 + (\{(UC_{sector\_2})_{corr}\}_{ln} \cdot E_{sector\_2})^2 + \dots + (\{(UC_{sector\_n})_{corr}\}_{ln} \cdot E_{sector\_n})^2}{|E_{sector\_1} + E_{sector\_2} + \dots + E_{sector\_n}|}}, \quad (5)$$

$$E_{group\_k} = E_{sector\_1} + E_{sector\_2} + \dots + E_{sector\_n}, \quad (6)$$

where the combined uncertainties  $UC_{group\_k}$  and total emissions  $E_{group\_k}$  per group  $k$  were calculated taking into account specifically for that group sector log-normally transformed uncertainties  $\{(UC_{sector\_1})_{corr}\}_{ln}$ ,  $\{(UC_{sector\_2})_{corr}\}_{ln}$ , ...,  $\{(UC_{sector\_n})_{corr}\}_{ln}$  in percent.

Group upper- and lower-half range values of uncertainty are descriptive but not straightforward to use in numerical modelling (e.g. emission perturbations in ensemble runs, flux inversions), so mean  $\mu^{ln}$  and standard  $\sigma^{ln}$  deviation of the



group log-normal distribution are calculated starting from Eq. (7):

$$E_{\text{group}_k} = e^{\mu^{\text{ln}} + \sigma^{\text{ln}} \cdot z}, \quad (7)$$

where  $z$  is a standard normal variable, and parameters  $\mu^{\text{ln}}$  and  $\sigma^{\text{ln}}$  represent a natural logarithm of group emissions, not the emissions themselves. The lower and upper bounds of the 95 % probability range, which are the 2.5th and 97.5th percentiles, respectively, are calculated assuming a log-normal distribution based on a corrected estimated half range of uncertainty from the error propagation approach and are lower and upper uncertainty values. Taking this into account and using the Z table for 2.5th and 97.5th percentiles  $p$  ( $p_{2.5} = -1.96$ ,  $p_{97.5} = 1.96$ ), mean  $\mu^{\text{ln}}$  and standard deviation  $\sigma^{\text{ln}}$  of log-normal distribution can be calculated following Eq. (8):

$$Z_p = \frac{\ln([E_{\text{group}_k}]_p) - \mu^{\text{ln}}_{\text{group}_k}}{\sigma^{\text{ln}}_{\text{group}_k}}, \quad (8)$$

resulting in Eqs. (9) and (10).

$$\mu^{\text{ln}}_{\text{group}_k} = \ln(E_{\text{group}_k}) + \frac{1}{2} \ln \left( 1 + \frac{[\text{UC}_{\text{group}_k}]_{\text{low}}}{100\%} \right) + \frac{1}{2} \ln \left( 1 + \frac{[\text{UC}_{\text{group}_k}]_{\text{high}}}{100\%} \right), \quad (9)$$

$$\sigma^{\text{ln}}_{\text{group}_k} = \frac{\ln \left( 1 + \frac{[\text{UC}_{\text{group}_k}]_{\text{low}}}{100\%} \right) - \ln \left( 1 + \frac{[\text{UC}_{\text{group}_k}]_{\text{high}}}{100\%} \right)}{-3.92}, \quad (10)$$

where  $[\text{UC}_{\text{group}_k}]_{\text{low}}$  and  $[\text{UC}_{\text{group}_k}]_{\text{high}}$  are in percent.

Figure 1 shows a simplified roadmap for yearly uncertainty calculations.

### 2.2.3 Group monthly uncertainties

The group monthly uncertainties are calculated starting from the yearly uncertainties, which can provide a more appropriate variation than the yearly timescale for operational modelling. In this way, yearly sector uncertainties are adjusted to represent monthly variability (no correlation between months is assumed) and further combined into group monthly uncertainties by means of the following four steps.

1. The same steps as for annual uncertainty calculation are used but based on monthly emission budgets (i.e. uncertainties for IPCC activities are combined to sectors with the error propagation method, corrected for systematic underestimation by the error propagation method, and adapted to have log-normal distribution).
2. The correlation  $\alpha$  (an uncertainty-boosting parameter) between yearly and monthly uncertainties is based on an analysis of the variations over the different months following Eq. (11). It is computed to enhance obtained monthly uncertainties as they 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:

$$\begin{aligned} (E_{\text{YEAR}} \cdot \text{UC}_{\text{YEAR}})^2 = \\ \alpha^2 \cdot ((E_{\text{MONTH1}} \cdot \text{UC}_{\text{MONTH1}})^2 \\ + (E_{\text{MONTH2}} \cdot \text{UC}_{\text{MONTH2}})^2 + \dots \\ + (E_{\text{MONTH12}} \cdot \text{UC}_{\text{MONTH12}})^2), \end{aligned} \quad (11)$$

where  $E$  and  $\text{UC}$  correspond to sector emission budget and uncertainty in kilotonnes and percent, respectively; YEAR, MONTH1, MONTH2, ..., MONTH12 are yearly and monthly (January, February, ..., December) values. Equation (11) is based on the rule for combining uncorrelated uncertainties under the addition of the error propagation equation (see Eq. 5) and the assumption that each month's uncertainty should be enhanced (boosted) by the same value.

3. The prior yearly sector uncertainties are multiplied by the boosting parameter (specific per country and emission sector), and the results are used as a first guess of prior month sector uncertainties.
4. The calculation steps (1) to (3) are iterated to find the best boosting parameter as the best fit between yearly and combined 12-month uncertainties, with the incremental step below a given acceptable threshold from Eq. (11) for each country and emission sector. With this optimum boosting parameter, monthly uncertainties per sector are calculated and then merged into groups, with a log-normal distribution of CO<sub>2</sub> emissions.

Detailed information on each Unix shell script included in the anthropogenic CO<sub>2</sub> emission uncertainty calculation tool CHE\_UNC\_APP (Choulga et al., 2021) is provided in the Supplement, Sect. S4.

### 2.2.4 Remarks about the fuel dependence and assumptions concerning correlation

It should be noted that IPCC (2006) provides default emission factor values for different fuels in transport-related activities (e.g. railways, aviation). Detailed fuel consumption information per IPCC activity that results in a long-cycle carbon was not available, and instead the most typical and consumed (common) fuel type (or its emission factor value) was used:

- aviation cruise (1.A.3.a\_CRS), climbing and descent (1.A.3.a\_CDS), and landing and take-off (1.A.3.a\_LTO) – jet kerosene;
- road transportation (1.A.3.b) and pipelines, off-road transport (1.A.3.e) – most typical emission factor uncertainty;

	GROUP 1						GROUP k			GROUP C		
	SECTOR 1				SECTOR 2		SECTOR j			SECTOR B		
	ACTIVITY 1		ACTIVITY 2		ACTIVITY 3		...	ACTIVITY i		...	ACTIVITY A	
Uncertainty per ACTIVITY	Emission Factor <sub>1</sub>	Activity Data <sub>1</sub>	Emission Factor <sub>2</sub>	Activity Data <sub>2</sub>	Emission Factor <sub>3</sub>	Activity Data <sub>3</sub>	...	Emission Factor <sub>i</sub>	Activity Data <sub>i</sub>	...	Emission Factor <sub>A</sub>	Activity Data <sub>A</sub>
	Combined <sub>1</sub>		Combined <sub>2</sub>		Combined <sub>3</sub>		...	Combined <sub>i</sub>		...	Combined <sub>A</sub>	
Uncertainty per SECTOR	Combined <sub>1</sub>				Combined <sub>2</sub>		Combined <sub>j</sub>			Combined <sub>B</sub>		
	Corrected for large values <sub>1</sub>				Corrected for large values <sub>2</sub>		Corrected for large values <sub>j</sub>			Corrected for large values <sub>B</sub>		
	Log-normal transformation <sub>1</sub>				Log-normal transformation <sub>2</sub>		Log-normal transformation <sub>j</sub>			Log-normal transformation <sub>B</sub>		
Uncertainty per GROUP	Combined <sub>1</sub>						Combined <sub>k</sub>			Combined <sub>C</sub>		
	Parameters of log-normal distribution <sub>1</sub>						Parameters of log-normal distribution <sub>k</sub>			Parameters of log-normal distribution <sub>C</sub>		

**Figure 1.** Yearly uncertainty calculation simplified roadmap.

- shipping (1.A.3.d) – composition of 80 % diesel and 20 % residual fuel oil;
- railways (1.A.3.c) – diesel.

It should also be noted that some uncertainty ranges for emission factors and/or activity data in IPCC (2006) and its refinements (IPCC, 2019) are not symmetrical and have higher uncertainty values for the lower-half range than for the half-range (or vice versa) due to input from expert knowledge or available in situ data, which then leads to the same pattern in final prior uncertainty range.

It should finally be noted that according to the IPCC (2006), all anthropogenic CO<sub>2</sub> emissions are assumed to be fully uncorrelated; hence the prior error correlations between grid cell emissions from the same sector or group should be assumed negligible if country- and/or sector-specific information is lacking.

### 3 Uncertainty calculation application

The method explained above has been applied to the EDGARv4.3.2\_FT2015 dataset to prepare prior uncertainty information for the ECMWF Integrated Forecasting System (IFS) model.

#### 3.1 Data input

In this example, 2015, the year of the Paris Agreement and reference for several Nationally Determined Contributions, is chosen as a base year to analyse anthropogenic CO<sub>2</sub> budgets (i.e. global, regional, national) from different sources (i.e. global statistics, national reports), benefitting the availability of observations (both in situ ground and spaceborne) as well as reported and verified emission inventories.

Following IPCC (2006) and its refinements (IPCC, 2019), starting from the global fossil CO<sub>2</sub> 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, an updated emission dataset CHE\_EDGAR-ECMWF\_2015<sup>2</sup> (Choulga et al., 2020) is derived. The EDGARv4.3.2 dataset is improved by correcting the allocation of the autoproducers to the manufacturing sector instead of the energy sector. Autoproducers are defined by the International Energy Agency (IEA) and include the energy (electricity and heat) generated by an industry for its own use, mostly for the manufacturing. An extra emission source of fugitive CO<sub>2</sub> from coal mines is also added, following the recommendations from IPCC (2019). Even though this emission source is not that large globally, usually the coal seam gas is composed dominantly of methane (CH<sub>4</sub>), but in some coal mines (in Australia and also in Brazil) seam gas consists predominantly (> 95 %) of CO<sub>2</sub> (Beamish and Vance, 1992), leading to significant atmospheric CO<sub>2</sub> concentration increases. An additional map for CHE\_EDGAR-ECMWF\_2015 with coal mining emissions from underground mines has been generated following the IPCC (2019) default values and the coal mining activity of CH<sub>4</sub> emission grid maps from hard and brown coal production in EDGARv4.3.2 (for more information refer to the Supplement, Sect. S2). For the update from 2012 to 2015 the fast-track approach of Olivier et al. (2016b) is used. The initial 92 IPCC activity uncertainties are combined into 20 EDGAR sectors for two distinct country types with well- and less well-developed statistical infrastructures (i.e. country's ability to register different emissions, meaning tabulate even very small emissions or only major ones, respectively). For the input to the IFS model the emission sectors are grouped in seven groups, with one group devoted to super power plants. Table 2 shows activity and sector grouping and emission budget differences between EDGARv4.3.2\_FT2015 and CHE\_EDGAR-ECMWF\_2015 datasets due to reallocation

<sup>2</sup>CHE stands for the CO<sub>2</sub> Human Emissions project (CHE, 2021).

**Table 2.** Grouping of anthropogenic long-cycle carbon CO<sub>2</sub> emission sectors into groups. Note provides main information and typical fuel type; global emission budgets for 2015 in megatonnes provides values for EDGARv4.3.2\_FT2015 (total sum 35 986.5 Mt) and CHE\_EDGAR-ECMWF\_2015 (total sum 35 995.2). Italics represent values with the biggest differences; asterisks (\*) represent values that were replaced from EDGARv4.3.2

No.	Group name	IPCC (2006) activities per sector	Note	Emission budget 2015, Mt			
				EDGARv4.3.2_FT2015		CHE_EDGAR-ECMWF_2015	
1	ENERGY_S	1.A.1.a (subset)	Power industry (without autoproducers): super-emitting power plants	<i>13 704.0</i>	<i>13841.2</i>	896.7	<i>12705.5</i>
2	ENERGY_A	1.A.1.a (rest)	Power industry (without autoproducers): standard-emitting power plants			<i>11 671.6</i>	
		4.C	Solid waste incineration	137.2		137.2	
3	MANUFACTURING	1.A.2	Combustion for manufacturing (including autoproducers)	<i>6182.8</i>	<i>8960.1</i>	<i>7320.4</i>	<i>10 096.0</i>
		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 metal 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	3321.9	3322.7	3322.7
5	AVIATION	1.A.3.a_CRS	Aviation cruise; typical fuel: jet kerosene	412.2	815.4	412.2	815.4
		1.A.3.a_CDS	Aviation climbing and descent; typical fuel: jet kerosene	305.5		305.5	
		1.A.3.a_LTO	Aviation landing and 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	6604.4	5530.6	6604.9
		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 transport – most typical emission factor uncertainty	255.2		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	1917.4	2443.5	1917.8	2450.6
		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	Solvent and product use	168.7*		168.3	

of the autoproducers from the energy sector (−8 %) to the manufacturing sector (+18 %) and due to the extra emission source of diffusive coal mine CO<sub>2</sub>.

### 3.2 Model constraints

The operational IFS model is used to provide global CO<sub>2</sub> forecasts using the gridded prior emissions previously described (Agusti-Panareda et al., 2014; Agusti-Panareda et al., 2019). A prototype 4D-Var inverse modelling system is currently under development to monitor anthropogenic CO<sub>2</sub> emission using the IFS. There is also an ongoing development to extend the window length beyond 24 h using an ensemble-based methodology.

The uncertainties derived for the seven groups described here have been used to generate an ensemble of forecasts for 2015 based on the operational IFS ensemble system (McNorton et al., 2020). This provides a representation of the model uncertainty and an estimation of the expected signal-to-noise ratio for a future inverse modelling system. Random seeds for each group and country were applied to the normalized log-normal mean  $\mu^{\ln}$  and standard deviation  $\sigma^{\ln}$  to generate emission scaling factors, which were then used for 50 ensemble members.

Primarily, the derived emission uncertainties presented here are envisaged for use as prior errors within atmospheric inversion frameworks. Aggregation of emission sectors into seven groups is required for computational efficiency and to reduce the dimensions of the inverse problem. To resolve collocated emissions, further information is required about spatial correlations and/or co-emitted species (e.g. nitrogen oxides, NO<sub>x</sub>). Within the IFS inversion prototype, the log-normal normalized standard deviation outlined in the previous section is used to provide the uncertainty values to prevent negative scaling factors.

### 3.3 CHE\_EDGAR-ECMWF\_2015 output

The new CHE\_EDGAR-ECMWF\_2015 dataset with anthropogenic fossil CO<sub>2</sub> emissions and their uncertainties was compiled and tested at ECMWF. The fossil CO<sub>2</sub> emissions include all long-cycle carbon emissions from human activities, such as fossil fuel combustion, industrial processes (e.g. cement), and product use, but excludes emissions from land-use change and forestry. Human CO<sub>2</sub> emission inventories were processed into gridded 0.1° × 0.1° resolution maps to provide an estimate of prior CO<sub>2</sub> emissions, aggregated in seven main emissions groups: (1) energy production by super-emitters, (2) energy production by standard emitters, (3) manufacturing, (4) settlements, (5) aviation, (6) other transport at ground level, and (7) others, with an estimation of their uncertainty and covariance. Aggregation of the IPCC activities and sectors into groups was based on similarities between the magnitude of uncertainty, the spatiotemporal correlation, and co-emission factors of each sector. It is

assumed that each emission group is fully correlated with itself and fully uncorrelated with any other group (only diagonal values of the 7 × 7 group covariance matrix for the atmospheric transport model are non-zero and equal to log-normal variance). The CHE\_EDGAR-ECMWF\_2015 data are freely available (<https://doi.org/10.5281/zenodo.3967439>; Choulga et al., 2020) and consist of 11 grid maps in NetCDF format and one Excel file with information on anthropogenic CO<sub>2</sub> emissions and their uncertainties. For detailed information on each file see Table 3.

### 3.4 Example of uncertainty calculation

Table 4 shows a step-by-step example of how yearly uncertainties are calculated, and Fig. 2 shows plotted probability density functions based on computed log-normal parameters. The example shows calculations for the TRANSPORT group that consists of several emission sectors. The example shows two countries with different statistical infrastructure development levels (the country with well-developed statistical infrastructure is Germany, and the country with less well-developed statistical infrastructure is the Russian Federation) and significant differences in emission budgets.

Calculated yearly and monthly uncertainties per country and emission group were assigned to each grid box on the global map. National uncertainties were applied uniformly across each country. Figure 3 shows an example of the upper and lower uncertainty limits of anthropogenic CO<sub>2</sub> emission flux for the 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 it is planned to address these uncertainties too, for example by following Oda et al. (2019) to characterize spatial patterns of the disaggregation errors in the emission maps.

## 4 Comparison and sensitivity

### 4.1 Comparison of total uncertainty in global CO<sub>2</sub> emission datasets

Calculated emissions and uncertainties in fossil CO<sub>2</sub> have been compared to other global datasets 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 a lowest value of −4.7 %/+9.6 %, or ±7.1 %; see Table 5. 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 the global level assuming no correlation following IPCC (2006); (ii) all calculations were done separately for upper- and lower-half ranges of uncertainty to pre-



**Table 3.** Detailed information on CHE\_EDGAR-ECMWF\_2015 data.

General note	Field/spreadsheet
Annual_Upper_Lower_Uncertainties_Percentage_0.1_0.1.nc	
File has $2 \times 8$ fields with annual upper- and lower-half ranges of uncertainty in percent per 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-half range of uncertainty (2.5th percentile of log-normal distribution) for yearly emissions, in percent
	“Upper” – upper-half range of uncertainty (97.5th percentile of log-normal distribution) for yearly emissions, in percent
	Sector – emission sector numerical name; “0”: emission group ENERGY_S (with IPCC (2006) activity 1.A.1.a (subset)) standing for power industry emissions from super-emitting power plants; “1”: ENERGY_A (1.A.1.a (rest), 4.C) – power industry emissions from standard-emitting power plants and solid waste incineration; “2”: 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 autoproductors), iron and steel production, non-ferrous metal production, non-energy use of fuels, non-metallic mineral production, and chemical processes; “3”: SETTLEMENTS (1.A.4, 1.A.5.a, 1.A.5.b.i, 1.A.5.b.ii) – energy for buildings, residential heating; “4”: AVIATION (1.A.3.a_CRS, 1.A.3.a_CDS, 1.A.3.a_LTO) – aviation cruise, climbing and descent, and landing and take-off; “5”: TRANSPORT (1.A.3.b, 1.A.3.d, 1.A.3.c, 1.A.3.e) – road transportation, shipping, railways, pipelines, and off-road transport; “6”: 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, and solvent and product use; “7”: all groups summed together
Monthly_Upper_Lower_Uncertainties_Percentage_0.1_0.1.nc	
File has $2 \times 8 \times 12$ fields with monthly upper- and lower-half ranges of uncertainty in percent per 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 $3 \times 8$ fields with annual emissions and upper- and lower-half ranges of uncertainty in $\text{kg m}^{-2} \text{s}^{-1}$ per 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-half range of uncertainty (2.5th percentile of log-normal distribution) for yearly emissions of ENERGY_S, in $\text{kg m}^{-2} \text{s}^{-1}$ ,
	“Sup_upper” – upper-half range of uncertainty (97.5th percentile of log-normal distribution) for yearly emissions of ENERGY_S, in $\text{kg m}^{-2} \text{s}^{-1}$ ,
	“Sup_flux” – yearly emissions of ENERGY_S, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Ene_lower”, “ene_upper”, “ene_flux” – same but for ENERGY_A, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Man_lower”, “man_upper”, “man_flux” – same but for MANUFACTURING, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Set_lower”, “set_upper”, “set_flux” – same but for SETTLEMENTS, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Avi_lower”, “avi_upper”, “avi_flux” – same but for AVIATION, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Tra_lower”, “tra_upper”, “tra_flux” – same but for TRANSPORT, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Oth_lower”, “oth_upper”, “oth_flux” – same but for OTHER, in $\text{kg m}^{-2} \text{s}^{-1}$
	“All_lower”, “all_upper”, “all_flux” – same but for all groups summed together, in $\text{kg m}^{-2} \text{s}^{-1}$
Monthly_Sup_Upper_Lower_Uncertainties_0.1_0.1.nc	
file has $3 \times 12$ fields with monthly emissions, and upper- and lower-half ranges of uncertainty in $\text{kg m}^{-2} \text{s}^{-1}$ 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-half range of uncertainty (2.5th percentile of log-normal distribution) for monthly emissions of ENERGY_S, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Sup_upper” – upper-half range of uncertainty (97.5th percentile of log-normal distribution) for monthly emissions of ENERGY_S, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Sup_flux” – monthly emissions of ENERGY_S, in $\text{kg m}^{-2} \text{s}^{-1}$
	“Month” – month numerical name, where 1, 2, ..., 12 correspond to January, February, ..., December

**Table 3.** Continued.

General note	Field/spreadsheet
<b>Monthly_Ene_Upper_Lower_Uncertainties_0.1_0.1.nc</b>	
File has 3 × 12 fields with monthly emissions and upper- and lower-half ranges of uncertainty in kg m <sup>-2</sup> s <sup>-1</sup> 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.
<b>Monthly_Man_Upper_Lower_Uncertainties_0.1_0.1.nc</b>	
File has 3 × 12 fields with monthly emissions and upper- and lower-half ranges of uncertainty in kg m <sup>-2</sup> s <sup>-1</sup> 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.
<b>Monthly_Set_Upper_Lower_Uncertainties_0.1_0.1.nc</b>	
File has 3 × 12 fields with monthly emissions and upper- and lower-half ranges of uncertainty in kg m <sup>-2</sup> s <sup>-1</sup> 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.
<b>Monthly_Avi_Upper_Lower_Uncertainties_0.1_0.1.nc</b>	
File has 3 × 12 fields with monthly emissions and upper- and lower-half ranges of uncertainty in kg m <sup>-2</sup> s <sup>-1</sup> 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.
<b>Monthly_Tra_Upper_Lower_Uncertainties_0.1_0.1.nc</b>	
File has 3 × 12 fields with monthly emissions and upper- and lower-half ranges of uncertainty in kg m <sup>-2</sup> s <sup>-1</sup> 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.
<b>Monthly_Oth_Upper_Lower_Uncertainties_0.1_0.1.nc</b>	
File has 3 × 12 fields with monthly emissions and upper- and lower-half ranges of uncertainty in kg m <sup>-2</sup> s <sup>-1</sup> 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.

Table 3. Continued.

General note	Field/spreadsheet
Monthly_All_Upper_Lower_Uncertainties_0.1_0.1.nc	
File has 3 × 12 fields with monthly emissions and upper- and lower-half ranges of uncertainty in kg m <sup>−2</sup> s <sup>−1</sup> 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 (three-letter abbreviation of a geographical entity), geographical name (name of a geographical entity), type (development level of country’s statistical infrastructure, meaning with well- or 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 – no. (number of anthropogenic CO <sub>2</sub> emission group), group (group name), IPCC (2006) activity (IPCC activities that are included in each group), note (short explanation of the group), global emission budget 2015 (total global emissions per group in megatonnes), prior upper- and lower-half ranges of uncertainty (in percent; initial, calculated purely based on assumptions from IPCC, lower- and upper-half ranges of uncertainty for countries with well- or less well-developed statistical infrastructures)  “YEARLY” – ISO code (three-letter abbreviation of a geographical entity), group (group name), budget (yearly anthropogenic CO <sub>2</sub> emission budget per group and total per geographical entity in kilotonnes), uncertainty range (in percent; calculated based on prior uncertainty range and yearly budgets per group and total per geographical entity; lower- and upper-half ranges of uncertainty and averaged uncertainty), contribution to country’s total uncertainty (in percent; share of each group in geographical entities’ total yearly uncertainty; total contribution is always 100 %), parameters of log-normal distribution (anthropogenic CO <sub>2</sub> 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)

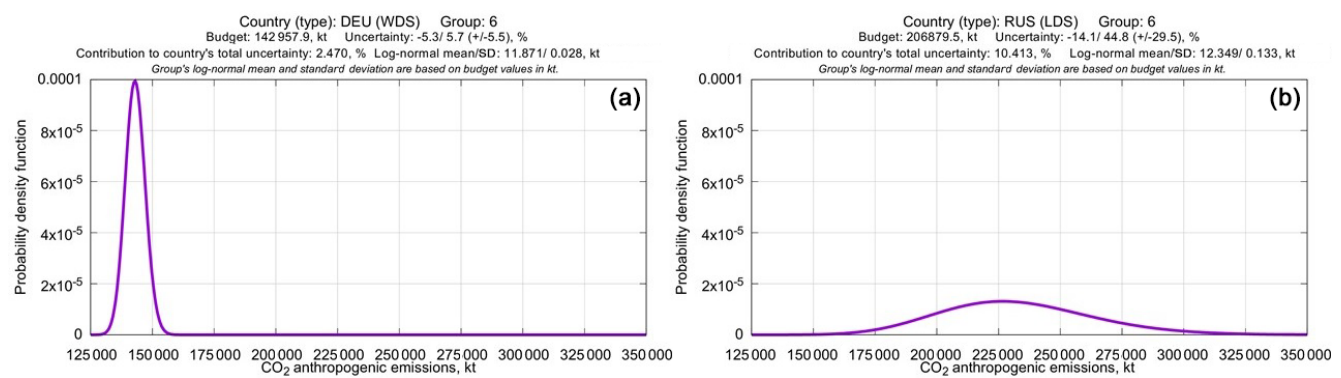


Figure 2. Probability density functions (for Germany **a** and the Russian Federation **b**) based on computed log-normal mean and standard deviation for the TRANSPORT group.

serve original information with asymmetric confidence intervals for large uncertainties (not required for the Approach 1 described in IPCC (2006), in which only the higher uncertainty value of the asymmetric interval should be used, leading to artificial inflation of uncertainty upper or lower limit);

and (iii) in this study proxy grid map uncertainties are not considered.

The contribution of each emission group to the total uncertainty per grid cell is assessed. Figures 4–7 show which group contributes the most to the total uncertainty per grid cell. The TRANSPORT group contributes most to the grid cell uncer-

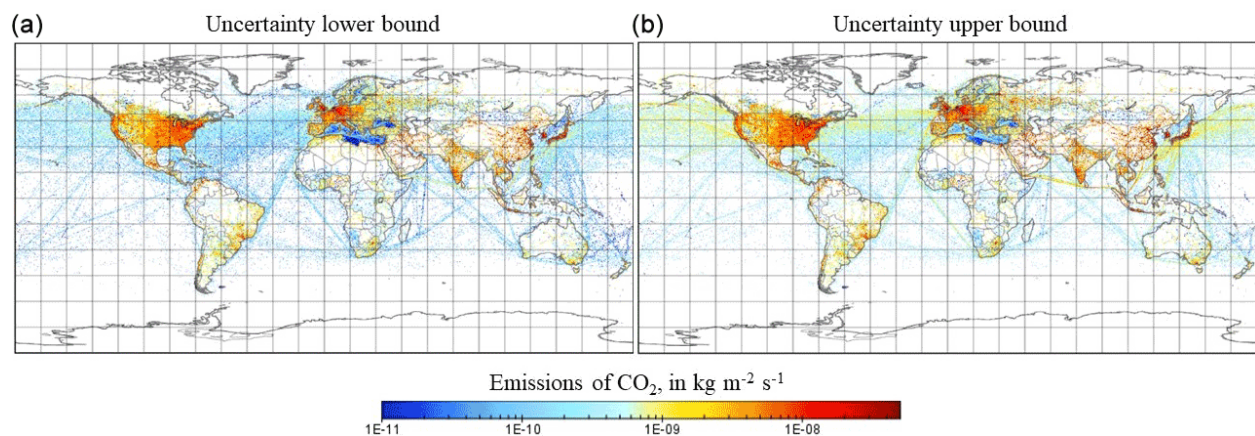
**Table 4.** Yearly uncertainty calculation steps. Example shows TRANSPORT group uncertainty calculations for Germany (DEU) and the Russian Federation (RUS), countries with a well- (WDS) and less well-developed statistical infrastructure (LDS), respectively. **(a)** Preparatory step (data collection) – same values are applied for all countries with the same development level of statistical infrastructure. **(b)** First step – same values are applied for all countries with the same development level of statistical infrastructure. **(c)** Second step – values are specific per geographical entity considering countries' development level of statistical infrastructure and emission budget (values are from CHE\_EDGAR-ECMWF\_2015); SD stands for standard deviation.

(a)											
IPCC (2006) activities per sector	IPCC (2006) activity	Note	Typical fuel	Uncertainty (%)							
				Emission factor				Activity data			
				DEU (WDS)		RUS (LDS)		DEU (WDS)		RUS (LDS)	
				Low	Up	Low	Up	Low	Up	Low	Up
1.A.3.b	1.A.3.b	Road transportation	Most typical emission factor	2.0	2.0	5.0	5.0	5.0	5.0	5.0	5.0
1.A.3.d	1.A.3.d	Waterborne navigation	Composition of 80 % diesel and 20 % residual fuel oil	2.1	1.1	2.1	1.1	5.0	5.0	50.0	50.0
1.A.3.c, 1.A.3.e	1.A.3.c	Railways	Diesel	2.0	0.9	2.0	0.9	5.0	5.0	5.0	5.0
	1.A.3.e	Other transportation – pipeline	None (suggested to neglect)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Other transportation – off-road	Most typical emission factor	2.0	2.0	5.0	5.0	50.0	100.0	50.0	100.0

(b)													
IPCC (2006) activities per sector	IPCC (2006) activity	Combined uncertainty per IPCC (2006) activity (%; see Eq. 1)				Combined uncertainty per sector (%; see Eq. 2)				Corrected combined uncertainty per sector (%; see Eqs. 3 and 4)			
		DEU (WDS)		RUS (LDS)		DEU (WDS)		RUS (LDS)		DEU (WDS)		RUS (LDS)	
		Low	Up	Low	Up	Low	Up	Low	Up	Low	Up	Low	Up
1.A.3.b	1.A.3.b	5.4	5.4	7.1	7.1	5.4	5.4	7.1	7.1	5.4	5.4	7.1	7.1
1.A.3.d	1.A.3.d	5.4	5.1	50.0	50.0	5.4	5.1	50.0	50.0	5.4	5.1	50.0	50.0
1.A.3.c, 1.A.3.e	1.A.3.c	5.4	5.1	5.4	5.1	50.3	100.1	50.5	100.3	50.3	106.9	50.5	107.0
	1.A.3.e	0.0	0.0	0.0	0.0								
		50.0	100.0	50.2	100.1								

(c)																
IPCC (2006) activities per sector	Emission budget 2015 per sector (× 10 <sup>3</sup> kt)		Uncertainty with assumed log-normal distribution per sector (%)				Emission budget 2015 per group (× 10 <sup>3</sup> kt; see Eq. 6)		Grouped uncertainty with assumed log-normal distribution per group (%; see Eq. 5)				Log-normal parameters of grouped uncertainty with assumed log-normal distribution per group (see Eqs. 9 and 10)			
	DEU (WDS)	RUS (LDS)	DEU (WDS)		RUS (LDS)		DEU (WDS)	RUS (LDS)	DEU (WDS)		RUS (LDS)		DEU (WDS)		RUS (LDS)	
			Low	Up	Low	Up			Low	Up	Low	Up	Mean	SD	Mean	SD
1.A.3.b	139.6	131.7	5.4	5.4	7.1	7.1	143.0	206.9	5.3	5.7	14.1	44.8	11.9	0.0	12.3	0.1
1.A.3.d	1.0	7.4	5.4	5.1	40.1	57.2										
1.A.3.c, 1.A.3.e	2.3	67.9	40.3	135.5	40.5	135.7										





**Figure 3.** CO<sub>2</sub> emission flux uncertainties (a lower- and b upper-half ranges of uncertainty) for the TRANSPORT group in kg m<sup>−2</sup> s<sup>−1</sup>.

**Table 5.** Comparison of global anthropogenic CO<sub>2</sub> 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	Hoesly et al. (2018)
CHE_EDGAR-ECMWF_2015	±7.1 % (−4.7/+9.6 %)	<i>Current study</i>
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 %*	Oda et al. (2018)

\* The difference between ODIAC and CDIAC gridded data is 3.3 %–5.7 % (Oda et al., 2018).

tainty over the Unites States of America (due to road and off-road transport) and over the ocean (due to shipping). The AVIATION group contributes most over main flight routes all over the globe. The OTHER group contributes the most over agricultural areas and regions with oil refineries and transformation industry and fuel exploitation. The MANUFACTURING group contributes most over industrial areas (e.g. in Vietnam and Bangladesh). The ENERGY\_A (and ENERGY\_S) group contributes the most over power plant (and super power plant) location grid cells (e.g. South Africa). The SETTLEMENTS group contributes the most to the grid cell uncertainty over either very densely or very sparsely populated areas.

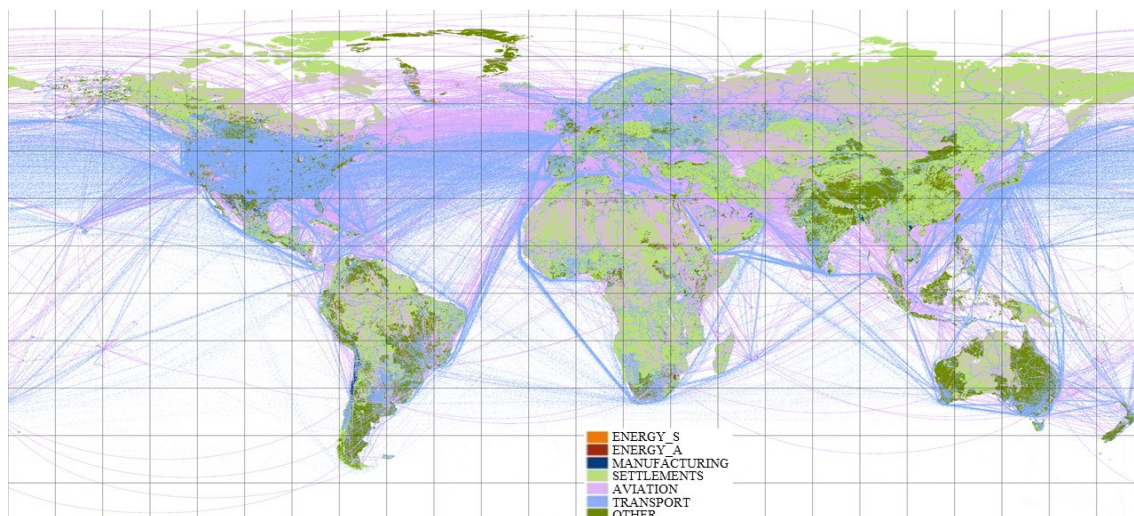
4.2 Dependence of the country-specific statistical infrastructure

Also, some specific geographical areas are analysed: chosen to be among the most emitting in total or per emission group and the most typical or most influential for a certain region. A list of these geographical entities and development levels of their statistical infrastructures is presented in Table 6.

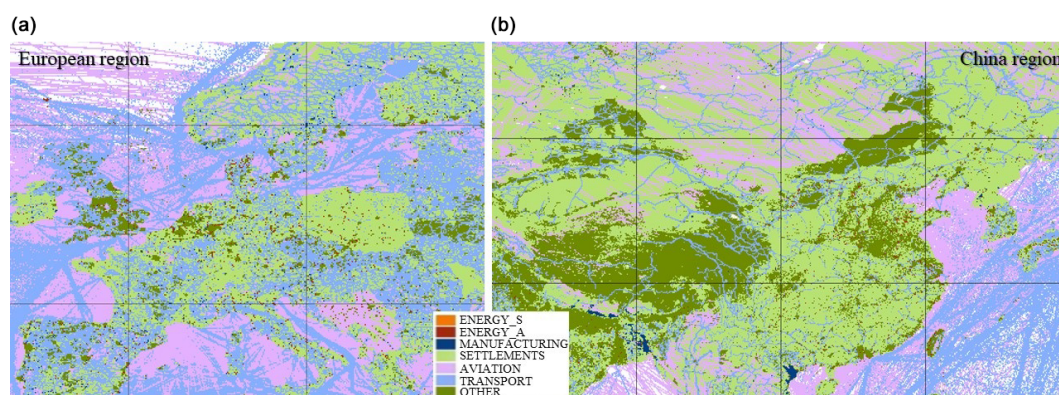
In order to see how the development level of country’s or geographical entity’s statistical infrastructure influences the

emission uncertainty in that country or geographical entity itself and (possibly) the globe, uncertainty calculations for selected entities were performed twice – with their original and switched types (i.e. a country with a well-developed statistical infrastructure becomes a country with a less well-developed statistical infrastructure and vice versa). More details on a geographical entity’s statistical infrastructure development level (e.g. how it was determined) are given in the Supplement, Sect. S5. Figure 8 shows sectoral emission budgets, uncertainties, and contributions in percentage to the total uncertainty in a country or geographical entity with its original and switched statistical infrastructure development levels. The biggest impact of development level change occurs for countries with larger emission budgets. On average, total uncertainties in selected countries (see Table 6) changed by 1 %–2 %; group uncertainties changed in line with prior uncertainties and countries’ emission budgets, as reported in Table 7.

Alterations in some countries’ (e.g. Germany, France) statistical infrastructure’s development levels lead to changes in uncertainties in Europe (28 members until end of 2019), with the most substantial change for the SETTLEMENTS group (e.g. 2.5 % and 1.0 %, respectively). Huge changes (> 10.0 %) in Europe’s (28 members until end of 2019) AVI-



**Figure 4.** Main emission group that contributes to the total uncertainty per grid cell – global region.



**Figure 5.** Main emission group that contributes to the total uncertainty per grid cell – European (a) and China (b) regions.

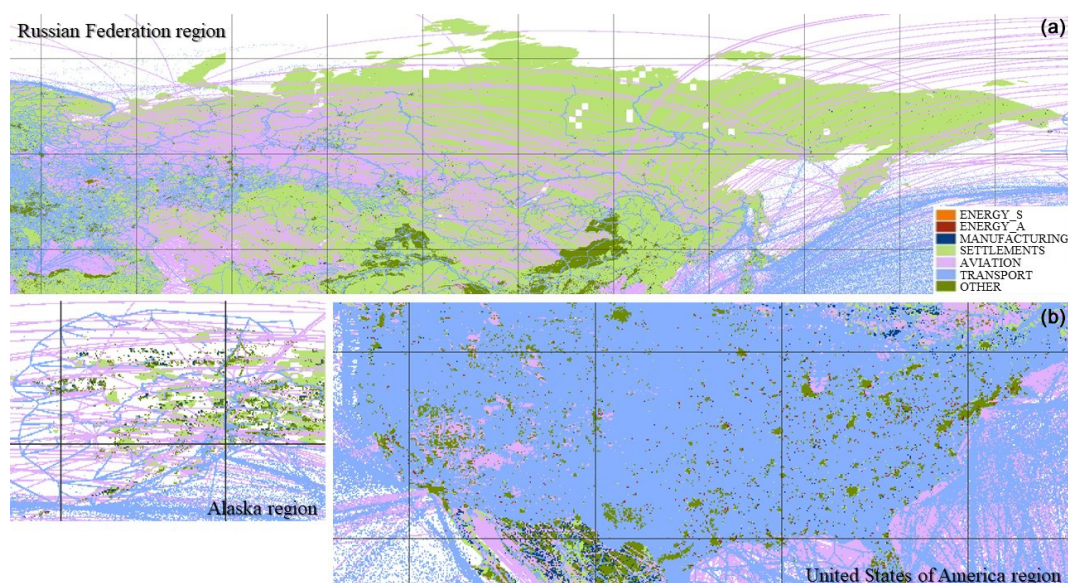
ATION group's uncertainty percentage value can be due to the variation in statistical infrastructure development level for Germany, United Kingdom, France, or Spain, though this group's contribution to Europe's total uncertainty remains negligible. Alterations in statistical infrastructure development levels for China or the United States of America modify even global uncertainties because these countries substantially contribute to the total global emission budget; e.g. China emits  $\sim 1/3$  of the global anthropogenic CO<sub>2</sub> budget and can change global total uncertainty up to 0.5 %.

#### 4.3 Effect of increasing temporal resolution from yearly to monthly

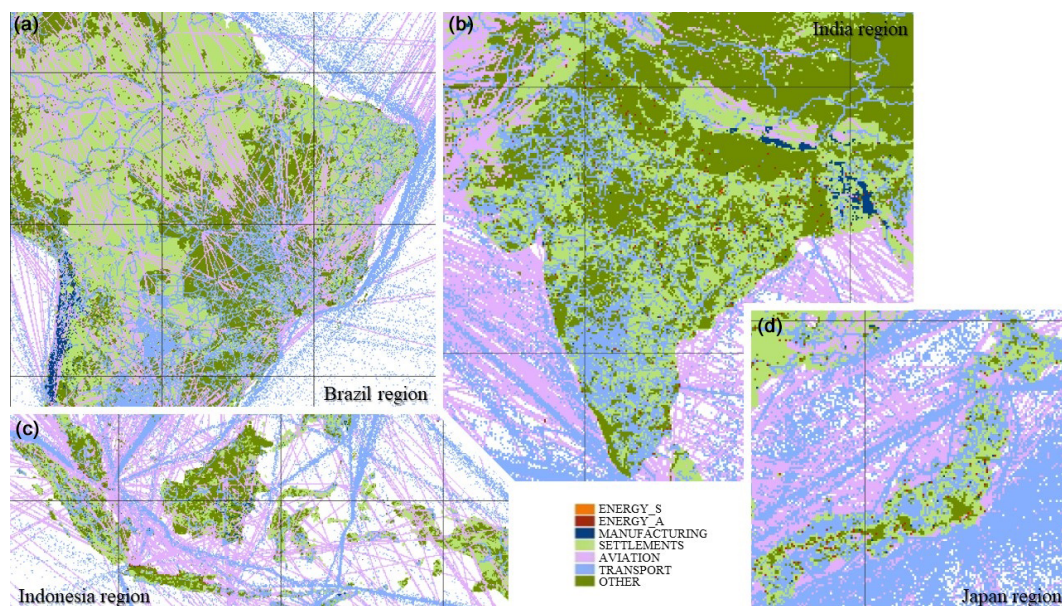
To increase the emission temporal resolution, monthly emissions and their uncertainties were calculated combining yearly emissions, monthly multiplication factors, and adapted uncertainty calculation methodology (see Sect. 2.2). Prior yearly uncertainties were multiplied by a dimensionless uncertainty-boosting parameter  $\alpha$  (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) with monthly emission budgets (weighted with the total number of days in a month). The dimensionless uncertainty-boosting parameter  $\alpha$  is applied; see Table 8 for most common values for countries with well- and less well-developed statistical infrastructures per sector. Boosting parameters become active ( $\alpha \neq 1$ ) when absolute uncertainty values are  $\geq 25.0\%$ , and  $\alpha$  increases with the increase in absolute uncertainty following a third-order polynomial. For lower-half ranges of uncertainty,  $\alpha$  has larger values and steeper growth than for upper-half ranges of uncertainty (e.g.  $-25.0\% \hat{=} \alpha = 1.5$  and  $-124.0\% \hat{=} \alpha = 2.6$ ,  $+25.0\% \hat{=} \alpha = 0.8$  and  $+124.0\% \hat{=} \alpha = 1.2$ ;  $\hat{=}$  means “corresponds to”), and  $\alpha$  behaves in the same way for countries with well- and less well-developed statistical infrastructures. Discrepancies in a different geographical entity's (country's





**Figure 6.** Main emission group that contributes to the total uncertainty per grid cell – the Russian Federation (a) and the United States of America (b) regions.



**Figure 7.** Main emission group that contributes to the total uncertainty per grid cell – Brazil (a), India (b), Indonesia (c), and Japan (d) regions

boosting parameters might be for several reasons. The main ones are (i) sector emissions were zero (e.g. super power plant emissions of the energy sector had no emissions), and (ii) sector uncertainties were  $\geq 50.0\%$  and needed to be adapted accordingly to log-normal distribution (this is the case for the agricultural soils sector with prior uncertainties  $-70.7/+0.0\%$  for countries with well- and less well-developed statistical infrastructures; discrepancies from Table 8 for agricultural soils are France –  $\alpha = 1.8/3.1$ , UK –

1.8/7.2, China – 1.8/8.4, Japan – 1.8/10.8, Brazil – 1.8/0.0, and the Russian Federation – 1.8/5.6, where the first value is for the lower-half range of uncertainty, and the second value is for the upper-half range of uncertainty).

In general, Brazil, Indonesia, and India have a very weak yearly cycle with quite high monthly uncertainties throughout the year. The globe, Europe (28 members until end of 2019), Germany, Spain, France, United Kingdom, Poland, China, Japan, the Russian Federation, and the United States

**Table 6.** List of selected geographical entities with their statistical infrastructure's development levels.

ISO Code	Geographical name	Type
GLB	All world countries	Mixed-developed statistical infrastructure
E28	Europe (28 members until end of 2019)	Well-developed statistical infrastructure
DEU	Germany	Well-developed statistical infrastructure
ESP	Spain	Well-developed statistical infrastructure
FRA	France	Well-developed statistical infrastructure
GBR	United Kingdom	Well-developed statistical infrastructure
POL	Poland	Well-developed statistical infrastructure
BRA	Brazil	Less well-developed statistical infrastructure
CHN	China	Well-developed statistical infrastructure
IDN	Indonesia	Less well-developed statistical infrastructure
IND	India	Well-developed statistical infrastructure
JPN	Japan	Well-developed statistical infrastructure
RUS	Russian Federation	Less well-developed statistical infrastructure
USA	United States of America	Well-developed statistical infrastructure

of America have more pronounced yearly cycles, most significant for the SETTLEMENTS and ENERGY\_A (and ENERGY\_S where present) groups and less significant for the 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 the Equator; see Janssens-Maenhout et al. (2019). In the summer months for the northern temperate zone, a strong decrease in SETTLEMENTS and ENERGY\_A (and ENERGY\_S where present) group emissions was observed, with a light decrease in MANUFACTURING group emissions and a light increase in AVIATION and TRANSPORT group emissions. This corresponds rather well to the assumption that most of the population in the Northern Hemisphere heat their houses during winter and take holidays and travel more during summer.

#### 4.4 Comparison for selected European countries with UNFCCC and TNO data

The CHE\_EDGAR-ECMWF\_2015 dataset containing seven global gridded fossil CO<sub>2</sub> emission flux maps and country- and 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 the same country-level information as primary input, though differences in inclusion, interpretation, and treatment of that data lead to diverse results in emissions. It is necessary to try to harmonize data inclusion or omission across datasets to have more clarity in the discrepancies.

For Europe (28 members until end of 2019), Germany, Spain, France, United Kingdom, Poland, Japan, the Russian Federation, and the United States of America, emission and uncertainty data were collected from UNFCCC NIR. The ag-

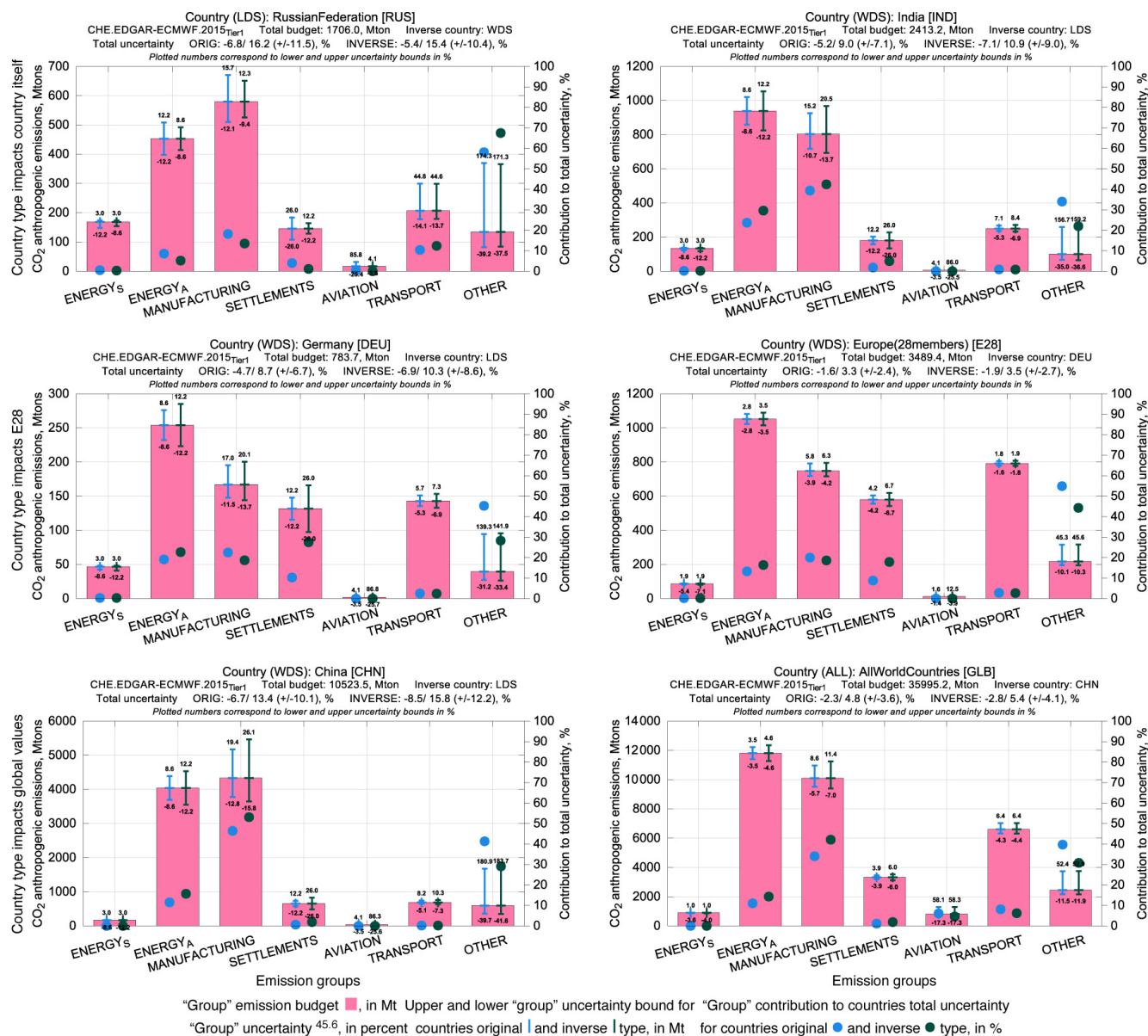
gregation of the IPCC (2006) activity-specific emissions and uncertainties into seven groups was done assuming no correlation, following IPCC (2006). Although IPCC (2006) has a standard table to report GHG emissions, uncertainties can be reported in less detail by a more general category (e.g. 2.D only 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 by the authors of this paper.

The Netherlands Organisation for Applied Scientific Research (TNO) has prepared the first version of their GHG and co-emitted species emission database (TNO\_GHGco\_v1.1) that covers the entire European domain (at  $0.1^\circ \times 0.05^\circ$  resolution), including CO<sub>2</sub> (distinguishing between fossil fuel and biofuel). Initial emission data are from the UNFCCC (common reporting format, CRF, tables) and the European Monitoring and Evaluation Programme (EMEP) of the Centre on Emission Inventories and Projections (CEIP) for air pollutants. These data were harmonized; checked for gaps, errors, and inconsistencies; and (where needed) replaced or completed using emission data from the Greenhouse Gas and Air Pollution Interactions and Synergies (GAINS) model (Amann et al., 2011). Moreover, inland shipping emissions were replaced with the 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 regarding their absolute value and trends (Kuenen et al., 2014). In this study emission budgets from 30 TNO sectors (Ingrid Super, Jeroen Kuenen, Antoon Visschedijk, and Hugo Denier van der Gon, personal communication, February 2020), and prior uncertainties calculated from IPCC (2006) and its refinements (IPCC, 2019) are used. In addition, the TNO has provided Tier 2 (Monte Carlo approach) uncertainties based on the same budgets and uncertainties from submitted NIR



**Table 7.** Influence of country's statistical infrastructure (countries with well- and less well-developed statistical infrastructures – WDSs and LDSs, respectively) on emission uncertainty.

Impact on the uncertainty	Group name	Cause description
Most substantial	SETTLEMENTS	<ul style="list-style-type: none"> <li>– Consists only of residential heating emissions</li> <li>– High differences in prior uncertainties for WDS and LDS: <math>\pm 12.2\%</math> and <math>\pm 26.0\%</math>, respectively</li> </ul>
Strong	MANUFACTURING	<ul style="list-style-type: none"> <li>– Budget usually makes a significant part of country's total emission budget</li> <li>– Globally mainly composed of combustion for manufacturing with rather low prior uncertainty (<math>\pm 8.6\%</math> and <math>\pm 12.2\%</math> for WDS and LDS, respectively) and non-metallic mineral production with much higher uncertainties (<math>\pm 70.9\%</math> and <math>\pm 93.0\%</math> for WDS and LDS, respectively)</li> <li>– Also contains emissions from very uncertain non-energy use of fuels (<math>\pm 121.7\%</math> and <math>\pm 124.0\%</math> for WDS and LDS, respectively) and chemical processes (<math>-107.8/+89.9\%</math> for both WDS and LDS) emissions, though their global share in this group is only <math>\sim 7.0\%</math></li> </ul>
	ENERGY_A	<ul style="list-style-type: none"> <li>– Budget usually makes a significant part of country's total emission budget</li> <li>– Composed of emissions from standard power plants with rather low uncertainties (<math>\pm 8.6\%</math> and <math>\pm 12.2\%</math> for WDS and LDS, respectively) and solid waste incineration with much higher uncertainties (<math>\pm 40.3\%</math> and <math>\pm 41.2\%</math> for WDS and LDS, respectively)</li> <li>– For the globe, the ratio of solid waste incineration to energy emissions is <math>\sim 1/100</math>, which keeps the total group prior uncertainty quite low (<math>\pm 3.5\%</math>)</li> <li>– Note: geographical entities with higher ratios will have higher uncertainties</li> </ul>
	ENERGY_S	<ul style="list-style-type: none"> <li>– Composed of emissions from super power plants only with rather low prior uncertainties (<math>-8.6/+3.0\%</math> and <math>-12.2/+3.0\%</math> for WDS and LDS, respectively) for all geographical entities</li> </ul>
Mild	TRANSPORT	<ul style="list-style-type: none"> <li>– Globally mainly composed of road transportation with rather low uncertainty (<math>\pm 5.4\%</math> and <math>\pm 7.1\%</math> for WDS and LDS, respectively) and shipping emissions with low uncertainties (<math>-5.4/+5.1\%</math>) for WDS and high uncertainties (<math>\pm 50.0\%</math>) for LDS</li> <li>– Also contains rather uncertain railways, pipelines, and off-road transport emissions (<math>\sim -50.4/+107.0\%</math> for both WDS and LDS), though their global share in this group is <math>\sim 16.0\%</math> only</li> <li>– Note: all international shipping is included in “all world countries” geographical entity</li> </ul>
Small	AVIATION	<ul style="list-style-type: none"> <li>– Extremely high differences in prior uncertainties for WDS and LDS (<math>-5.5/+6.4\%</math> and <math>-50.1/+106.8\%</math>, respectively), though this group's share in global emissions is only <math>2.3\%</math></li> <li>– Note: all international aviation is included in “all world countries” geographical entity</li> </ul>
Negligible	OTHER	<ul style="list-style-type: none"> <li>– Composed of very uncertain components with usually almost the same prior uncertainties for WDS and LDS</li> <li>– Main composite globally (<math>\sim 78.0\%</math>) are emissions from oil refineries and the transformation industry with prior uncertainties of <math>-54.4/+149.3\%</math> and <math>-57.7/+151.4\%</math> for WDS and LDS, respectively</li> <li>– Also usually has the highest contribution to the country's total uncertainty</li> </ul>



**Figure 8.** Emission budgets, uncertainties, and contributions in percentage to the total uncertainty in the country with their original and switched (inverse) types (countries with well- and less well-developed statistical infrastructures – WDSs and LDSs, respectively): impacting mainly the country itself, e.g. the Russian Federation (RUS) and India (IND); impacting also Europe (E28), e.g. Germany (DEU); impacting even global values, e.g. China (CHN).

reports based on a Tier 1 approach. The Monte Carlo simulations were done at the highest detail level (nomenclature for reporting (NFR) sector and fuel type) assuming correlations between certain sectors (for more information see Super et al., 2020), and then emissions were aggregated to groups assuming no correlation.

Figure 9 shows emission budgets and uncertainties in megatonnes and contributions in percent to the total geographical entity’s uncertainty for Europe (28 members until end of 2019), Germany, France, and United Kingdom with their original statistical infrastructure development types

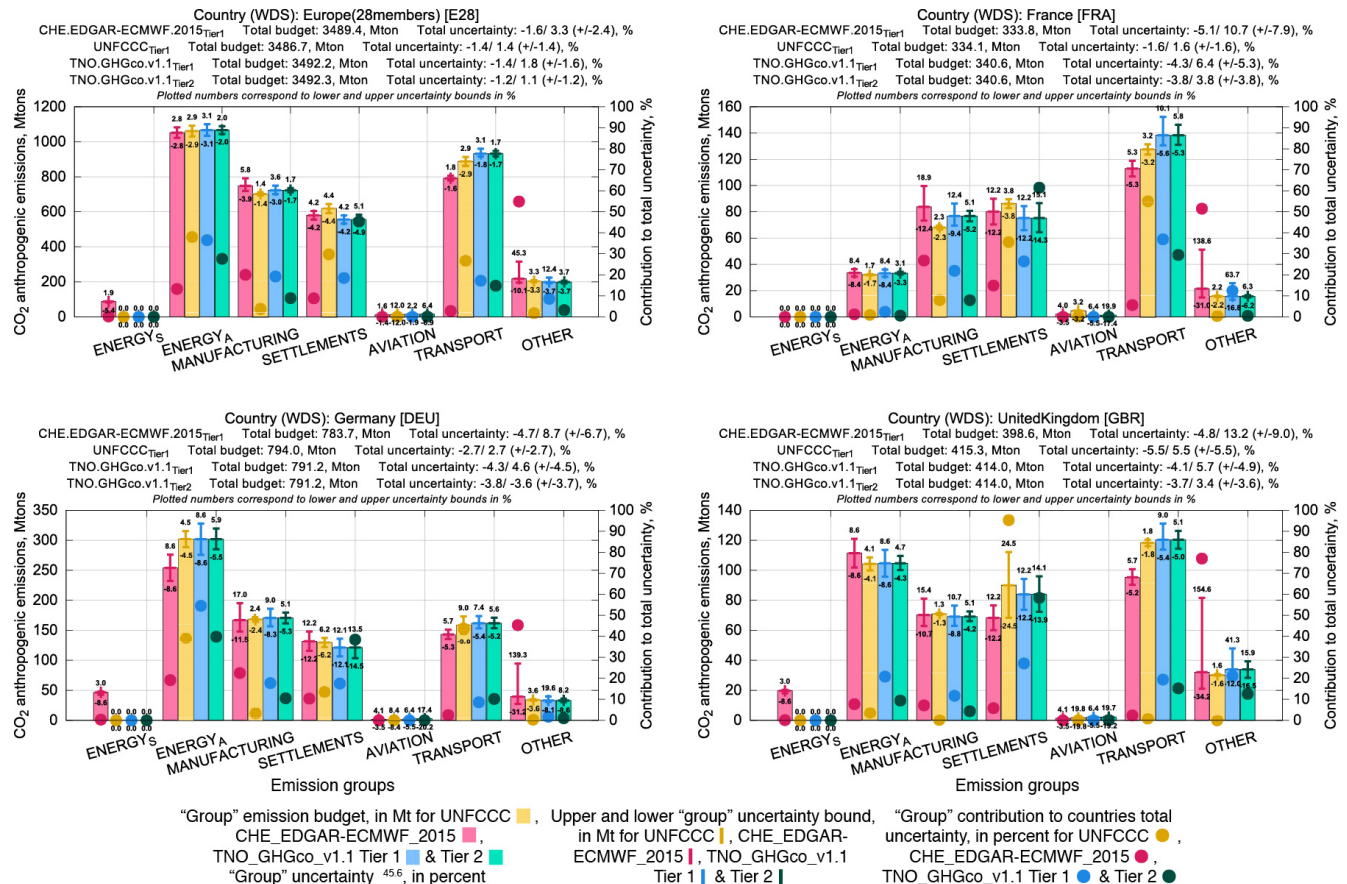
based on data from CHE\_EDGAR-ECMWF\_2015 (in pink), UNFCCC (in yellow), and TNO\_GHGco\_v1.1 Tier 1 (in blue) and Tier 2 (in green); plots for Spain and Poland are not shown here. Out of the four different sources, usually UNFCCC and TNO\_GHGco\_v1.1 Tier 2 uncertainties are the lowest ones and CHE\_EDGAR-ECMWF\_2015 the highest one. It should be noted that (i) UNFCCC uncertainties were aggregated to groups individually per country as uncertainties are reported in a rather free form and thus could be aggregated from different levels of precision; (ii) uncertainties for Europe (28 members until end of 2019) from CHE\_EDGAR-

**Table 8.** Dimensionless (DN) lower- and upper-half-range boosting parameter for countries with well- and less well-developed statistical infrastructures – WDSs and LDSs, respectively.

No.	Group name	IPCC (2006) activities per sector	Uncertainty-boosting parameter (DN)			
			WDS countries		LDS countries	
			Low	Up	Low	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

ECMWF\_2015 are rather low as they were calculated by aggregating information from 28 countries; and (iii) differences in uncertainties in CHE\_EDGAR-ECMWF\_2015 with other sources, especially in fuel-dependent emission groups, might be due to biofuels or 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. where absence of the uncertain glass production activity in the non-metallic mineral production sector decreases overall uncertainty). Differences in uncertainties between TNO\_GHGco\_v1.1 Tier 1 and TNO\_GHGco\_v1.1 Tier 2 show additional value in an 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.

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. The main reason for that is the use of country-specific emission data and activity data uncertainties, which are lower than default values suggested by IPCC (2006) and its refinements (IPCC, 2019). Only for the fuel-dependent groups (e.g. AVIATION) might UNFCCC uncertainties be higher than in this study as rather uncertain biofuels might be taken into account (note: CHE\_EDGAR-ECMWF\_2015 does not take biofuels into account). Also, emission budgets reported to the UNFCCC show some differences from the ones from CHE\_EDGAR-ECMWF\_2015. For Japan, group budgets agree rather well, and the total budget difference is  $\sim 1.0\%$ . For the Russian Federation, ma-



**Figure 9.** Emission budgets, uncertainties, and contributions in percentage to the total uncertainty for Europe (E28), Germany (DEU), France (FRA), and United Kingdom (GBR).

major differences are in the ENERGY\_A (and ENERGY\_S) and MANUFACTURING groups, which results in a  $\sim 6.0\%$  higher total budget of CHE\_EDGAR-ECMWF\_2015. For the United States of America, major differences are  $\sim 200$  Mt and  $\sim 100$  Mt for the SETTLEMENTS and OTHER groups, respectively, which results in a  $\sim 4.0\%$  higher total budget than based on UNFCCC data. Recent comparison of different gridded global datasets by Andrew (2020) pointed out that only a few of these datasets provide quantitative uncertainty assessment; see the summary in Table 5. Compared to other global emission uncertainty values, CHE\_EDGAR-ECMWF\_2015 shows the lowest values mainly due to the aggregation technique.

#### 4.5 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 emission factor uncertainty (for road and off-road transport) were used because detailed fuel consumption information per IPCC activity was not available for this study. The EDGAR dataset development team do have specific fuel information globally, which could be used

for uncertainty calculation. The EDGAR dataset with incorporated fuel-specific activity data and emission factor uncertainties and Tier 1 approach for uncertainty calculation (see Supplement, Sect. S6) is hereinafter referred to as EDGAR-JRC. Country budget uncertainties were calculated by considering “full fuel” splitting and by taking into consideration the assumption that the emission factors, 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 in activity data was set in accordance with IPCC (2006) guidelines, in the range of 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 marine transport; and 35.0 % for industrial processes of cement, lime, glass, and ammonia (the range of uncertainty values refers to the 95 % confidence interval of the mean, assigned separately to countries with well- and less well-developed statistical infrastructures). Uncertainties from the EDGAR-JRC dataset aggregated to the group level were compared with the ones from CHE\_EDGAR-ECMWF\_2015; see Table 9 for Europe (28 members until end of 2019) and all world countries and



Table S8 from the Supplement, Sect. S6, for all the remaining geographical entities from Table 6. 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. Uncertainties calculated with fuel-specific data are usually smaller; when prevailing fuel coincides with a typical fuel type from CHE\_EDGAR-ECMWF\_2015, emission group uncertainties from both sources are quite similar. It should be noted that (i) countries' total uncertainty is higher in EDGAR-JRC due to the aggregation technique (full correlation is assumed), and (ii) AVIATION group uncertainties are higher in EDGAR-JRC due to prior aggregation of all three aviation connected sectors (cruise, climbing and descent, and landing and take-off).

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 percentage values, they are usually quite small in megatonnes.

#### 4.6 Atmospheric sensitivity to nationally disaggregated emissions

The gridded emissions are required input to the ECMWF IFS model used to simulate atmospheric CO<sub>2</sub> globally (Agusti-Panareda et al., 2014; Agusti-Panareda et al., 2019). Ideally, uncertainties at a grid cell level would be preferred by the models in general, which is a difficult time-consuming task. To check the usefulness of the information-intensive derivation of uncertainties at a grid cell level, it was decided to run some experiments. High-resolution ( $\sim 25$  km horizontal resolution, 137 vertical levels) simulations with the ECMWF IFS model have been performed to assess the atmospheric sensitivity to fully resolved emissions compared to nationally smoothed (global emission budget is conserved); see Fig. 10.

Model simulations were performed for January 2015 with 3-hourly output. Anthropogenic, fire, ocean, and biogenic fluxes (large-scale model bias mitigated by the biogenic CO<sub>2</sub> flux adjustment scheme, BFAS) were considered. For the full model configuration description see McNorton et al. (2020). It was noted that point sources (e.g. power plants, factories) can be easily detected if they comprise a 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 difficult to detect one extra or missing emitting hotspot (e.g. in Germany). China is a very good example for both cases as its western part has very few hotspots, and they are easy to detect over the low-emitting background. Its eastern part, however, has lots of hotspots and high-emitting areal sources, making it almost impossible to disentangle emissions from a

single power plant or factory from the high-emitting background. Differences of several parts per million are detected over multiple regions, highlighting the importance of using high-resolution spatially resolved emissions. With increase in both flux and transport model resolutions, these differences are expected to increase further with steeper atmospheric CO<sub>2</sub> gradients.

## 5 Data availability

EDGARv4.3.2 data are open-access and available at [http://data.europa.eu/89h/jrc-edgar-edgar\\_v432\\_ghg\\_gridmaps](http://data.europa.eu/89h/jrc-edgar-edgar_v432_ghg_gridmaps) (last access: 29 June 2021, Janssens-Maenhout et al., 2017) and are documented in Janssens-Maenhout et al. (2019). CHE\_EDGAR-ECMWF\_2015 data are freely available <https://doi.org/10.5281/zenodo.3967439> (Choulga et al., 2020) and documented in this paper. The CHE\_UNC\_APP anthropogenic CO<sub>2</sub> emission uncertainty calculation tool is freely available <https://doi.org/10.5281/zenodo.5196190> (Choulga et al., 2021) and documented in this paper.

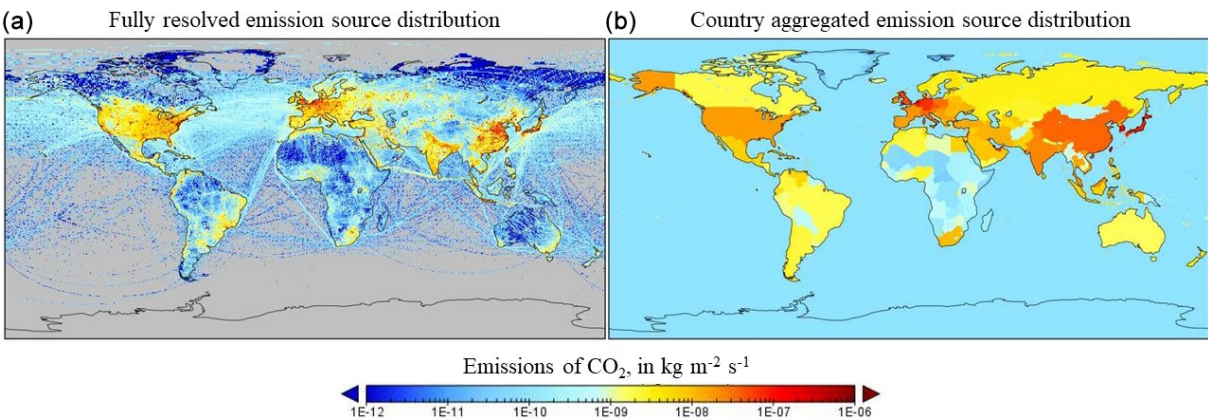
## 6 Recommendations and conclusion

A pre-processor has been created that allows derivation of the upper- and lower-half range of uncertainty grid maps while making use of an appropriate classification of more certain and uncertain sectors. These grid maps allow assessment of the error propagation of country emission budgets following the IPCC 2006 Guidelines for National Greenhouse Gas Inventories. It is a first step in evaluating where to provide more effort in reducing the propagated error budget that can be taken up in any global or regional atmospheric model as a first step. The method has been applied using EDGARv4.3.2\_FT2015 and was tested as input to the ECMWF IFS ensemble spread to characterize the carbon dioxide (CO<sub>2</sub>) atmospheric concentrations' uncertainties in the prototype of the Copernicus CO<sub>2</sub> Monitoring and Verification Support Capacity. At the country level the CHE\_EDGAR-ECMWF\_2015 dataset provides generally larger uncertainty ranges, reduced when more detailed information is available. In summary, using the information uniformly available for all countries, a coherent uncertainty representation is obtained.

The application in the ECMWF IFS Earth system model sheds light on the spatial representativeness of the emissions. While the emission-intensive point sources were checked with reference to their spatial location, the diffuse emission sources are gridded using spatial proxy data. With CHE\_EDGAR-ECMWF\_2015 implemented in the IFS model it was demonstrated that the choice of the spatial proxy data has a strong influence on the model results. As such, it is proposed that this is analysed in comparison to other datasets, going beyond the evaluation of the probability density of the spatial proxy itself. Contribution of repre-

**Table 9.** Aggregated to the group level uncertainties (lower- and upper-half ranges of uncertainty) in percent and contributions in percent to the total uncertainty (CV) for Europe (E28) and the globe (GLB) from EDGAR-JRC (with extra fuel type knowledge) and CHE\_EDGAR-ECMWF\_2015 (with typical fuel only).

Country	Group name	EDGAR-JRC			CHE_EDGAR-ECMWF_2015		
		Low (%)	Up (%)	CV (%)	Low (%)	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



**Figure 10.** Anthropogenic CO<sub>2</sub> flux source distribution at ~ 25 km resolution – fully resolved (a), country aggregated (b).

sentativeness errors to uncertainties and time correlation will need to be assessed in successive future studies, as foreseen under the Prototype System for a Copernicus CO<sub>2</sub> Service (CoCO<sub>2</sub>) project, following up on the CO<sub>2</sub> Human Emissions (CHE) project.

The use of an ensemble technique to estimate CO<sub>2</sub> uncertainties is recommended. The optimal number of ensemble members is bound by practical considerations on computational costs. Leutbecher (2018) found a minimum of an 8-member ensemble can mimic some of the skill of larger ensembles, with a 20-member ensemble being a typical value used by several modelling systems and with a 50-member ensemble being a desirable target. Further grouping of an-

thropogenic emissions into, for example, one to reduce the dimensions of the problem is also possible with the tool CHE\_UNC\_APP (Choulga et al., 2021).

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<sub>2</sub> monitoring mission CO<sub>2</sub>M 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 computation of priors presented in

this paper, are an important preparatory step for an ensemble-based CO<sub>2</sub> monitoring and verification system prototype, such as the one developed within the CHE project.

**Supplement.** The supplement related to this article is available online at: <https://doi.org/10.5194/essd-13-5311-2021-supplement>.

**Author contributions.** All the authors participated in the uncertainty calculation tool CHE\_UNC\_APP design and CHE\_EDGAR-ECMWF\_2015 map 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.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Acknowledgements.** The authors thank Glenn Carver (ECMWF) for editorial help and assistance and Vladimir Tupoguz for invaluable support during the preparation of the paper and numerous discussions. Margarita Choulga was funded by the CO<sub>2</sub> Human Emissions (CHE) project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 776186, and by the Prototype System for a Copernicus CO<sub>2</sub> Service (CoCO<sub>2</sub>) project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 958927.

**Financial support.** This research has been supported by the CO<sub>2</sub> Human Emissions (CHE) project (grant no. 776186) and the Prototype System for a Copernicus CO<sub>2</sub> Service (CoCO<sub>2</sub>) project (grant no. 958927).

**Review statement.** This paper was edited by David Carlson and reviewed by three anonymous referees.

## 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., and Wunch, D.: Forecasting global atmospheric CO<sub>2</sub>, *Atmos. Chem. Phys.*, 14, 11959–11983, <https://doi.org/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., Lange-
- rock, B., Law, R. M., Loh, Z., Morguí, J. A., Parrington, M., Peuch, V.-H., Ramonet, M., Roehl, C., Vermeulen, A. T., Warneke, T., and Wunch, D.: Modelling CO<sub>2</sub> weather – why horizontal resolution matters, *Atmos. Chem. Phys.*, 19, 7347–7376, <https://doi.org/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., and Winiwarter, W.: Cost-effective control of air quality and greenhouse gases in Europe: Modelling and policy applications, *Environ. Modell. Softw.*, 26, 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, *Global Biogeochem. Cy.*, 10, 419–429, <https://doi.org/10.1029/96GB01523>, 1996.
- Andres, R. J., Boden, T. A., and Marland, G.: Annual Fossil-Fuel CO<sub>2</sub> Emissions: Mass of Emissions Gridded by One Degree Latitude by One Degree Longitude, United States: N. p., (NDP-058.2016), ESS-DIVE [data set], <https://doi.org/10.3334/CDIAC/ffe.ndp058.2016>, 2016.
- 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 multi-year, global gridded fossil fuel CO<sub>2</sub> emission data product: Evaluation and analysis of results, *J. Geophys. Res.-Atmos.*, 119, 10.213–10.231, <https://doi.org/10.1002/2013JD021296>, 2014.
- Beamish, B. B. and Vance, W. E.: Greenhouse gas contributions from coal mining in Australia and New Zealand, *J. Roy. Soc. New Zeal.*, 22:2, 153–156, <https://doi.org/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 CH<sub>4</sub> and N<sub>2</sub>O 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, <https://doi.org/10.1029/2003JD003697>, 2004.
- CHE: CO<sub>2</sub> Human Emissions (CHE) project official website, available at: <https://www.che-project.eu>, last access: 29 June 2021.
- 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., and Tao, S.: Modelling temporal variations in global residential energy consumption and pollutant emissions, *Appl. Ener.*, 184, 0306–2619, 820–829, <https://doi.org/10.1016/j.apenergy.2015.10.185>, 2016.
- Choulga, M., McNorton, J., and Janssens-Maenhout, G.: CHE\_EDGAR-ECMWF\_2015, Zenodo [data set], <https://doi.org/10.5281/zenodo.3967439>, 2020.
- Choulga, M., Janssens-Maenhout, G., and McNorton, J.: Anthropogenic CO<sub>2</sub> emission uncertainty

- calculation tool CHE\_UNC\_APP, Zenodo [code], <https://doi.org/10.5281/zenodo.5196190>, 2021.
- Cong, R., Saitō, M., Hirata, R., Ito, A., and Maksyutov, S.: Uncertainty Analysis on Global Greenhouse Gas Inventories from Anthropogenic Sources, in: 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, <https://doi.org/10.11159/rtese18.141>, 2018.
- Cong, R., Saitō, M., Hirata, R., Ito, A., and Maksyutov, S.: Uncertainty Analysis on Global Greenhouse Gas Inventories from Anthropogenic Sources, *International Journal of Environmental Pollution and Remediation (IJEPR)*, 7, 1–8, <https://doi.org/10.11159/ijepr.2019.001>, 2019.
- 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, 30 May 2003, 2003.
- Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, 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., Joetzer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, 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., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, *Earth Syst. Sci. Data*, 11, 1783–1838, <https://doi.org/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., and 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, <https://doi.org/10.5194/gmd-11-369-2018>, 2018.
- IPCC: 2006 IPCC Guidelines for National Greenhouse Gas Inventories, edited by: Eggleston, S., Buendia, L., Miwa, K., Ngara, T., and Tanabe, K., IPCC-TSU NGGIP, IGES, Hayama, Japan, available at: <https://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html> (last access: 29 June 2021), 2006.
- IPCC: 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, 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. (authors), edited by: Gomez, D. and Irving, W., Vol. 1. Ch. 8, Task Force on National Greenhouse Gas Inventories (TFI), IPCC's 49th Session, 12 May 2019, Kyoto, Japan, 2019.
- Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., and Schaaf, E.: Emissions Database for Global Atmospheric Research, version v4.3.2 part I Greenhouse gases (gridmaps), European Commission, Joint Research Centre (JRC) [data set], available at: [http://data.europa.eu/89h/jrc-edgar-edgar\\_v432\\_ghg\\_gridmaps](http://data.europa.eu/89h/jrc-edgar-edgar_v432_ghg_gridmaps) (last access: 29 June 2021), 2017.
- 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., Björk, 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., Hoesung, S., Husband, R., Juvyns, O., Kentarchos, A., Landgraf, J., Lang, R., Löschner, A., Marshall, J., Meijer, Y., Nakajima, M., Palmer, P., Peylin, P., Rayner, P., Scholze, M., Sierk, B., and Veeckind, P.: Towards an operational anthropogenic CO<sub>2</sub> emissions monitoring and verification support capacity, *B. Am. Meteorol. Soc.*, 101, E1439–E1451, <https://doi.org/10.1175/BAMS-D-19-0017.1>, 2020.
- Jones, M. W., Andrew, R. M., Peters, G. P., Janssens-Maenhout, G., De-Gol, A. J., Ciais, P., Patra, P. K., Chevallier, F., and Le Quéré, C.: Gridded fossil CO<sub>2</sub> emissions and related O<sub>2</sub> combustion consistent with national inventories 1959–2018, *Sci. Data*, 8, 2, <https://doi.org/10.1038/s41597-020-00779-6>, 2021.
- 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.
- 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., and He, K.: Reduced carbon emission estimates from fossil fuel combustion and cement production in China, *Nature*, 524, 7565, 335–338, <https://doi.org/10.1038/nature14677>, 2015.
- Leutbecher, M.: Ensemble size: How suboptimal is less than infinity?, *Q. J. Roy. Meteor. Soc.*, 145, 107–128, <https://doi.org/10.1002/qj.3387>, 2018.
- Marland, G., Pielke Sr., R., Apps, M., Avissar, R., Betts, R., Davis, K., Frumhoff, P., Jackson, S., Joyce, L., Kauppi, P., Katzenberger, J., Macdicken, K., Neilson, R., Niles, J., Niyogi, D., Norby, R., Pena, N., Sampson, N., and Xue, Y.: The climatic impacts of land surface change and carbon management, and the implications for climate-change mitigation policy, *Clim. Policy*, 3, 149–157, <https://doi.org/10.3763/cpol.2003.0318>, 2003.
- Marland, G., Boden, T. A., and Andres, R. J.: Global, regional, and national fossil fuel CO<sub>2</sub> emissions, in: Trends: A Compendium of Data on Global Change, US Department of Energy, Carbon



- Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, 2006.
- McNorton, J. R., Bousserez, N., Agustí-Panareda, A., Balsamo, G., Choulga, M., Dawson, A., Engelen, R., Kipling, Z., and Lang, S.: Representing model uncertainty for global atmospheric CO<sub>2</sub> flux inversions using ECMWF-IFS-46R1, *Geosci. Model Dev.*, 13, 2297–2313, <https://doi.org/10.5194/gmd-13-2297-2020>, 2020.
- Oda, T. and Maksyutov, S.: A very high-resolution (1 km × 1 km) global fossil fuel CO<sub>2</sub> emission inventory derived using a point source database and satellite observations of nighttime lights, *Atmos. Chem. Phys.*, 11, 543–556, <https://doi.org/10.5194/acp-11-543-2011>, 2011.
- Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO<sub>2</sub>, version 2016 (ODIAC2016): a global monthly fossil fuel CO<sub>2</sub> 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, *Mitig. Adapt. Strat. Gl.*, 24, 1007–1050, <https://doi.org/10.1007/s11027-019-09877-2>, 2019.
- ODIAC: ODIAC Fossil Fuel CO<sub>2</sub> Emissions Dataset, ODIAC [data set], <https://doi.org/10.17595/20170411.001>, 2021.
- Olivier, J. G. J. and Janssens-Maenhout, G.: CO<sub>2</sub> Emissions from Fuel Combustion – 2016 Edition, IEA CO<sub>2</sub> report 2016, Part III, Greenhouse-Gas Emissions, OECD – IEA, ISBN 9789264258563, EU Science Hub, [https://doi.org/10.1787/co2\\_fuel-2016-en](https://doi.org/10.1787/co2_fuel-2016-en), 2016a.
- Olivier, J. G. J., Janssens-Maenhout, G., Muntean, M., and Peters, J. A. H. W.: Trends in global CO<sub>2</sub> emissions: 2016 report, PBL Netherlands Environmental Assessment Agency, The Hague, the Netherlands, PBL publication number: 2315, European Commission, Joint Research Centre, Directorate Energy, Transport & Climate, JRC Science for Policy Report: 103428, 1–86, available at: [https://www.pbl.nl/sites/default/files/downloads/pbl-2016-trends-in-global-co2-emissions-2016-report-2315\\_4.pdf](https://www.pbl.nl/sites/default/files/downloads/pbl-2016-trends-in-global-co2-emissions-2016-report-2315_4.pdf) (last access: 29 June 2021), 2016b.
- Petrescu, A. M. R., Peters, G. P., Janssens-Maenhout, G., Ciais, P., Tubiello, F. N., Grassi, G., Nabuurs, G.-J., Leip, A., Carmona-Garcia, G., Winiwarter, W., Höglund-Isaksson, L., Günther, D., Solazzo, E., Kiesow, A., Bastos, A., Pongratz, J., Nabel, J. E. M. S., Conchedda, G., Pilli, R., Andrew, R. M., Schelhaas, M.-J., and Dolman, A. J.: European anthropogenic AFOLU greenhouse gas emissions: a review and benchmark data, *Earth Syst. Sci. Data*, 12, 961–1001, <https://doi.org/10.5194/essd-12-961-2020>, 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., and Scholze, M.: An operational anthropogenic CO<sub>2</sub> emissions monitoring & verification support capacity – Baseline requirements, Model components and functional architecture, European Commission Joint Research Centre, Publications Office of the European Union, Luxembourg, JRC107499, EUR 28736 EN, ISBN 978-92-79-72101-4, <https://doi.org/10.2760/08644>, 2017.
- 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<sub>2</sub> and CO to support inverse modelling and network design, *Atmos. Chem. Phys.*, 20, 1795–1816, <https://doi.org/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<sub>2</sub> emissions, *Atmos. Chem. Phys.*, 13, 5189–5203, <https://doi.org/10.5194/acp-13-5189-2013>, 2013.