



The PRIMAP-hist national historical emissions time series

Johannes Gütschow¹, M. Louise Jeffery¹, Robert Gieseke¹, Ronja Gebel¹, David Stevens¹,
Mario Krapp², and Marcia Rocha²

¹Potsdam Institute for Climate Impact Research, Telegraphenberg A 31, 14473 Potsdam, Germany

²Climate Analytics, Friedrichstraße 231, Haus B, 10969 Berlin, Germany

Correspondence to: Johannes Gütschow (johannes.guetschow@pik-potsdam.de)

Received: 15 April 2016 – Published in Earth Syst. Sci. Data Discuss.: 2 June 2016

Revised: 7 October 2016 – Accepted: 12 October 2016 – Published: 9 November 2016

Abstract. To assess the history of greenhouse gas emissions and individual countries' contributions to emissions and climate change, detailed historical data are needed. We combine several published datasets to create a comprehensive set of emissions pathways for each country and Kyoto gas, covering the years 1850 to 2014 with yearly values, for all UNFCCC member states and most non-UNFCCC territories. The sectoral resolution is that of the main IPCC 1996 categories. Additional time series of CO₂ are available for energy and industry subsectors. Country-resolved data are combined from different sources and supplemented using year-to-year growth rates from regionally resolved sources and numerical extrapolations to complete the dataset. Regional deforestation emissions are downscaled to country level using estimates of the deforested area obtained from potential vegetation and simulations of agricultural land. In this paper, we discuss the data sources and methods used and present the resulting dataset, including its limitations and uncertainties. The dataset is available from doi:10.5880/PIK.2016.003 and can be viewed on the website accompanying this paper (<http://www.pik-potsdam.de/primap-live/primap-hist/>).

1 Introduction

The question of responsibility for climate change and its impacts plays a significant role in the UNFCCC¹ negotiations around the global agreement to limit the global mean temperature increase and avoid dangerous climate change. It is interlinked with the discussion about equitable access to sustainable development, which forms the basis of different frameworks to assess whether climate targets put forward by countries reflect a “fair share” in the collective burden to reshape the economy towards emissions neutrality. The Brazilian delegation to the UNFCCC has put forward a framework that assesses a country's contribution to climate change by calculating the fraction of the total warming generated by that country's historical greenhouse gas emissions. This approach is explained in Miguez and Filho (2000) and has been quantified in Höhne et al. (2010), den Elzen et al. (2013), and

Matthews et al. (2014), among others. Other effort-sharing proposals use cumulative per capita emissions as a metric and thus also need a detailed record of historical emissions by individual countries (Winkler et al., 2011; Baer et al., 2008; Bode, 2004). In 2001 the MATCH² expert group was established by the UNFCCC to generate historical emissions time series for this purpose. The dataset which resulted from this effort proved very useful in the negotiations and to the scientific community (Höhne et al., 2010). It was updated in den Elzen et al. (2013) with data from the Emissions Database for Global Atmospheric Research v4.2 (EDGAR) to cover all gases and emissions until 2010. The Climate Analysis Indicator Tool (CAIT) also publishes a historical greenhouse gas emissions dataset that is a composite of other sources (World Resources Institute, 2016). However, non-CO₂ emissions are only covered for recent years (1990–2012) and it resolves

¹United Nations Framework Convention on Climate Change

²Ad hoc group for the modeling and assessment of contributions of climate change, <http://www.match-info.net>

either sectors or gases but not both at the same time. Most of the sources used in the CAIT composite dataset are also included in the dataset presented here. The Global Carbon Project publishes the Global Carbon Budget (Le Quéré et al., 2015), which covers the atmospheric concentration of CO₂ and its sources and sinks. The fossil fuel CO₂ emissions data used are taken directly from other sources; non-CO₂ emissions data are not included.

Here we present a historical emissions dataset with a finer sectoral resolution, newly available input data, and new and improved methods for the combination of datasets. Previous versions of the PRIMAP-hist (PRIMAP – Potsdam Real-time Integrated Model for probabilistic Assessment of emissions Paths) dataset have been used in the UNEP³ gap report 2015 (UNEP, 2015) and the INDC fact sheets published by the Australian-German Climate and Energy College (Meinshausen and Alexander, 2016). Predecessors of the dataset, especially the PRIMAP4 baseline⁴, have been used, for example, for the Climate Action Tracker⁵ and in Meinshausen et al. (2015). The dataset presented here has been improved in categorical resolution, time coverage, and country coverage compared to its predecessors. Methodological improvements include extrapolation with regional growth rates, more sophisticated downscaling methods (e.g., for land use emissions), and category and gas aggregation that automatically interpolates and extrapolates missing data.

We build our time series from a range of publicly available data sources (see Sect. 2), which are prioritized based on their completeness and reliability – an approach that has also been taken by the IPCC⁶ to compile the historical dataset for the 5th Assessment Report (IPCC, 2014, Annex.II.9, Historical data). For each time series (country-, gas-, and sector- resolved), the lower-priority sources are used as year-by-year growth rates⁷ to extend the higher-priority sources. Where no country data are available, we use regional growth rates, growth rates from superordinate sectors, and numerical extrapolation to complete the time series.

For land use emissions, we use the approach introduced in Matthews et al. (2014) and downscale a regional dataset using estimates of deforested areas derived from simulations of potential vegetation and agricultural land.

The PRIMAP-hist dataset covers the six Kyoto greenhouse gases and gas groups (Kyoto GHG). Independent time series are generated for carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), and sulfur hexafluoride (SF₆). For all gases

except CO₂, the sectoral resolution is that of the main IPCC 1996 categories. For CO₂, more detailed categories are used because some important datasets cover only subsectors of categories 1 and 2. For details and sector names, we refer the reader to Table 1.

NF₃ is not included as it has only been included in the group of Kyoto Protocol relevant gases for the second commitment period of the Kyoto Protocol, which started in 2013, and data availability is therefore still scarce. In the remainder of the paper we use the term fluorinated gases to refer to the combined group of gases HFCs, PFCs, and SF₆.

We use the IPCC 1996 categories instead of the new IPCC 2006 categories because almost all data sources are reported using the 1996 categories and we can avoid conversions between categorizations by using the 1996 categories. The UNFCCC is switching towards IPCC 2006 categories for data reported by countries; however, issues with the reporting software resulted in some countries delaying their emissions reporting and others asking the UNFCCC not to display the reported data. We plan to switch to the IPCC 2006 categories for a future release of the PRIMAP-hist dataset once these problems are solved.

The emissions time series cover the period of 1850 to 2014. This is achieved through the combination of various sources and extrapolation for some sectors, gases, and countries both into the past and into the future. The extent of the extrapolation needed varies between sectors, gases, and countries. Data coverage is very good for energy-related CO₂ emissions for the whole period. For other gases and sectors we have to rely on growth rates from regional data for the period before 1970 and on numerical extrapolation for the period after 2012. The data sources we use are described in Sect. 2, while the details of the combination process, including the prioritization, are described in Sect. 4.

The time series starts in 1850 for all sectors, including land use. Pre-1850 land use emissions have a small effect on cumulative emissions, and accounting for them would “results in a shift of attribution of global temperature increase from the industrialized countries to less industrialized countries, in particular South Asia and China, by up to 2–3 %” (Pongratz and Caldeira, 2012). On the other hand, uncertainties are especially high for early emissions, which limits the usefulness of the additional data. However, preindustrial land use change emissions could be included in a future version of this dataset.

As this dataset is designed to be used for global climate policy analysis, we provide data for all 196 member states of the UNFCCC as well as several countries and territories that are not UNFCCC members, not internationally recognized, or associated with a UNFCCC member state but not included in the state’s emissions reporting. We follow the territorial coverage of the countries’ submissions to the UNFCCC and use territorial accounting, which is in line with UNFCCC standards. Territorial accounting attributes emissions originating from a certain territory at any point in time

³United Nations Environment Programme

⁴<https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-flagship-projects/primap/emissionsmoduledocumentation/primap-baseline-reference>

⁵<http://www.climateactiontracker.org>

⁶Intergovernmental Panel on Climate Change

⁷Other publications use the term “rate of change”.

Table 1. Categorical detail of the PRIMAP-hist source for different gases. The categorical hierarchy uses IPCC 1996 terminology. The subcategories of categories 1 and 2 are only resolved for CO₂. Other gases are treated at the level of categories 1 and 2. For categories 2E and 2F of the industrial sector, there are no data for CO₂ because these categories only include the production and consumption of fluorinated gases.

Category	Sector name	Gases
0	National total	CO ₂ , CH ₄ , N ₂ O, HFCs, PFCs, SF ₆
0EL	National total excluding LULUCF	CO ₂ , CH ₄ , N ₂ O, HFCs, PFCs, SF ₆
1	Total energy	CO ₂ , CH ₄ , N ₂ O
1A	Fuel combustion activities	CO ₂ , CH ₄ , N ₂ O
1B1	Fugitive emissions from solid fuels	CO ₂
1B2	Fugitive emissions from oil and gas	CO ₂
2	Industrial processes	CO ₂ , CH ₄ , N ₂ O, HFCs, PFCs, SF ₆
2A	Mineral products	CO ₂
2B	Chemical industries	CO ₂
2C	Metal production	CO ₂
2D	Other production	CO ₂
2G	Other	CO ₂
3	Solvent and other product use	CO ₂ , N ₂ O
4	Agriculture	CO ₂ , CH ₄ , N ₂ O
5	Land Use, land use change, and forestry	CO ₂ , CH ₄ , N ₂ O
6	Waste	CO ₂ , CH ₄ , N ₂ O
7	Other	CO ₂ , CH ₄ , N ₂ O

to the state the territory currently belongs to. Emissions of former colonies are thus attributed to the now independent state and not to the former metropolitan state. Occupation of countries' territories is only taken into account if the occupying country reports the emissions from that territory.⁸ In Sect. 4.3 we present a list of territories included in the emissions of UNFCCC parties as well as information on the territories that are treated separately and how we deal with missing data and territorial changes.

The paper is organized as follows: we begin by describing the individual data sources we use in Sect. 2 and their prioritization in Sect. 3. In Sect. 4 we describe how the dataset is constructed from the individual sources, including the special treatment of land use data. In Sect. 5 we give information on how to obtain and use the data. Results are described in Sect. 6 with information on the uncertainties of emissions data in Sect. 7. Limitations are covered in Sect. 8. Methodological details and data sources that we did not use are described in the Appendices A, B, and C.

2 Data sources

In this section we describe the data sources used to create our composite source. We only use sources that are publicly available and give preference to sources that are not composites of other sources in order to avoid including original sources twice, once directly and once indirectly, through a composite source. However, it is likely that some sources

⁸This is the case for Israel and the Palestinian territories, for example.

share at least some input data, such as information on fossil fuel production or use the same emission factors. The sources are grouped into four categories. *Country-reported data* form the highest priority category as it can benefit from detailed knowledge about the specific situation in a country and is well accepted in the context of the UNFCCC negotiations. This is exemplified by the linking of the entry into force of the Paris Agreement to the latest country reported emissions and not to any third-party source (UNFCCC, 2015b). Where this data are not available, or do not meet our minimum requirements (see Sect. 2.1 below), we use *country-resolved data* provided by third parties, such as research institutions and international organizations. To extrapolate data into the past we use *region-resolved datasets*. Finally, we use two *gridded datasets* to calculate land use change patterns and subsequently country-resolved land use change emissions. Figure 1 gives an overview of the data sources described in detail in the remainder of this section. Detailed information on data preprocessing is available in Appendix B. In the text we refer to data sources using the abbreviations introduced in the source description below.

2.1 Minimum requirements for data

To be useful for our composite source, data have to meet some minimal requirements. Emissions data have inherent fluctuations due to weather (determining heating requirements), economic activity, and other factors. Not all sources model all these factors equally and therefore exhibit different fluctuations. When combining the sources, we use the year-by-year growth rates from the lower-priority source to

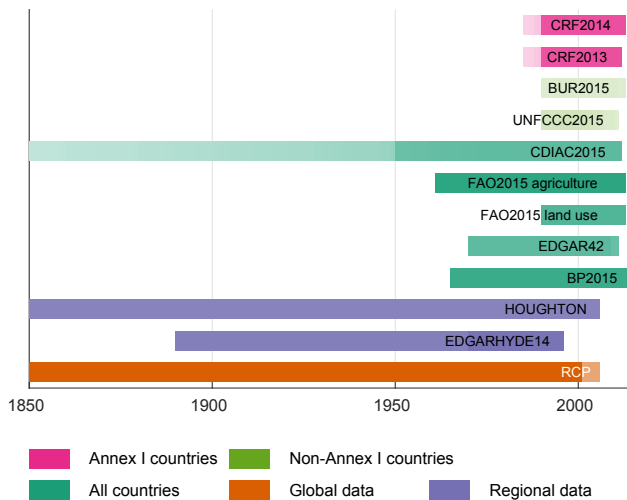


Figure 1. Coverage of years and countries in the sources used for the PRIMAP-hist dataset. The color indicates the country group covered or the regional resolution, while the intensities indicate the fraction of countries in the group covered by the source in each year. The fraction is taken over all gases and categories, which can be seen in the CDIAC time series, where the flaring time series only starts in 1950. The RCP time series for CH₄ ends in 2000, leading to the lower coverage after the year 2000.

extend a higher-priority source (for details see Sect. 4.1 and Appendices A4 and A5). To weaken the influence of these fluctuations, we use the trend of several years for the matching instead of a single year. We therefore require that each time series contain at least three data points spread over a period of at least 11 years. Furthermore, we need time series with the detail of sectors and gases listed in Table 1.

2.2 Country-reported data

Under the UNFCCC there are several requirements for reporting of greenhouse gas emissions data (see, e.g., Yamin and Depledge, 2005). Both developed (Annex I) and developing (non-Annex I) parties⁹ have to regularly submit communications that include an inventory of national GHG emissions and removals. Detailed requirements, however, differ strongly between Annex I and non-Annex I parties. Annex I parties have to submit an inventory that covers all sectors, gases, and years since 1990 annually. The submissions should consist of two parts, the common reporting format (CRF) tables with the data and a national inventory report (NIR), which gives background information like the rationale

⁹The term “parties” refers to the countries which have ratified the UNFCCC. Annex I parties refers to those countries listed in Annex I of the Kyoto Protocol (KP) which are the developed countries under the UNFCCC. The definition is now almost two decades old and does not represent the state of economic development any more. The distinction between developed and developing countries is thus subject to constant debate in the UNFCCC meetings.

behind the selection of emission factors and methodological questions. For details on the CRF tables, see Sect. 2.2.3 below. Annex I parties also submit national communications, which originally served the purpose to report on policies and measures to implement the party’s commitment to aim to return emissions to 1990 levels by the year 2000. The NIRs have recently (decided in 2011 at COP17¹⁰, Durban, South Africa) been complemented with biennial reports to enhance reporting. The emissions data contained should be consistent with the CRF data. Under the Kyoto Protocol (KP), Annex I parties have to regularly submit information needed to assess whether they are meeting their emissions targets. For our purpose, the CRF data are the most useful of these sources. The other sources do not provide additional information for the purpose of this paper and are not used.

Non-Annex I parties were required to submit an initial national communication within 3 years after the entry into force of the convention. The least developed countries (LDCs) could decide whether to submit an initial national communication. The submissions were required to contain an emissions inventory which covers the years 1990 to 1994 for most submissions. A time frame for subsequent national communications could not be agreed upon, and only a few countries submitted further national communications with updated inventories. The guidelines for national communications for non-Annex I parties are less stringent than the guidelines for Annex I parties; consequently, the coverage and detail in sectors and gases of the data differ strongly between countries. Since 2014, non-Annex I parties have been required to report GHG inventory information through biennial update reports (BURs). The first report was due by December 2014; however, only 24 of over 150 countries have actually submitted (as of January 2016). LDCs and SIDS¹¹ (94 countries in total) are exempted from the mandatory submission and can submit at their discretion.

The Paris Agreement requires regular national inventory reports by all parties, which might improve emissions reporting in the future (UNFCCC, 2015b, Article 7(a)).

2.2.1 National communications and national inventory reports for developing countries (UNFCCC2015)

Most developing countries reported historical emissions data at least once using national communications (UNFCCC, 2015a) and sometimes national inventory reports. However, several countries only reported data for the period of 1990 to 1994, sometimes only single years. Therefore, a lot of countries’ submissions do not meet our minimal data requirements and are consequently not used for the composite source. Where the data meet our requirements, we use them with high priority as they are prepared by in-country experts, which gives the results based on these data high cred-

¹⁰COP: Conference of the Parties to the UNFCCC

¹¹Small island developing states

ibility within the country and is beneficial for policy analysis. We compare the data with third-party data to identify whether there are differences that cannot be explained by uncertainties. National inventory reports give a more detailed overview over the emissions inventory than national communications but are not published by all countries. While developed-country parties also submit national communications and national inventory reports we only use these data for developing countries as we have the CRF data for developed countries (see Sect. 2.2.3 below). The data used here have been downloaded from the UNFCCC website using the “Detailed data by party” interface (UNFCCC, 2015c). The date of access was 25 September 2015. Some countries submit their data prepared according to IPCC 2006 guidelines. These data are not available through the interface (Andorra, Cook Islands, Jamaica, Kiribati, Malawi, Mauritania, Mexico, Namibia, Samoa, Swaziland, South Africa, and Tunisia). Furthermore, the UNFCCC greenhouse gas data interface seems to lag behind the submissions and misses some submissions from 2015 and 2016 (as of 1 February 2016). The source preprocessing is explained in Appendix B.

2.2.2 Biennial update reports (BUR2015)

Biennial update reports (BURs) are submitted to the UNFCCC by non-Annex I parties (UNFCCC, 2016). They contain greenhouse gas emissions information with varying detail in sectors, gases, and years. As of 1 February 2016, 24 countries have submitted data. Unfortunately, most of the submissions do not meet our minimal data requirements.

Argentina, Ghana, India, Namibia, Paraguay, Peru, Thailand, Tunisia, and Vietnam have submitted detailed values only for a single year. Bosnia and Herzegovina has published data for 2010 and 2011. Andorra and Macedonia have published only aggregate Kyoto greenhouse gas data.

Brazil and Singapore have published detailed information for 1994, 2000, and 2010; however, the level of detail is not sufficient for all sectors. Chile, Mexico, South Africa, Republic of Korea, and Uruguay have detailed information for a range of years in the annex to the BUR and the NIR. However, for South Africa the level of detail is not sufficient for all sectors and gases.

Colombia, Costa Rica, and Montenegro use the IPCC 2006 categorization, so we cannot include the data in the current version of this dataset. The Lebanon BUR was not accessible on the UNFCCC website, so we could not assess whether there are useful data in it (as of 1 February 2016).

The final PRIMAP-hist dataset uses BUR2015 data for Azerbaijan, Brazil, Chile, Republic of Korea, Mexico, Singapore, South Africa, and Uruguay. The source preprocessing is explained in Appendix B.

2.2.3 UNFCCC CRF (CRF2014, CRF2013)

CRF data, short for common reporting format, are reported by all Annex I parties every year on a mandatory basis. The data are very detailed, both in sectors and gases, and undergo review for consistency and compliance with reporting guidelines by expert teams from the UNFCCC roster of experts. We use the final version of the 2014 data (UNFCCC, 2014a), which contains information until the year 2012. The 2013 revision (UNFCCC, 2013) is used as a backup in case there are gaps in the 2014 data. The first year is 1990 with a few exceptions of data series starting in 1985. All Kyoto gases are covered and data are submitted using IPCC 1996 categories.¹²

The 2015 edition of the CRF data uses IPCC 2006 categories. This posed problems in data preparation for several countries such that publication was significantly delayed. To date (April 2016), not all countries have submitted their data, with large emitters missing.¹³ The gas NF_3 has been added as it is included in the Kyoto greenhouse gases for the second commitment period of the Kyoto Protocol. CRF2015 data will be included in a future update of this dataset together with a move to IPCC 2006 categorization.

Appendix B contains some additional information on the creation of the emissions pathways with individual fluorinated gases combined together.

2.3 Country-resolved data

2.3.1 BP Statistical Review of World Energy (BP2015)

The BP Statistical Review of World Energy is published every year and contains time series for CO_2 emissions from consumption of oil, gas, and coal (which corresponds to IPCC 1996 category 1A). Emissions data are derived on the basis of the carbon content of the fuels and statistics of fuel consumption. The 2015 edition (British Petroleum, 2015) contains information for 76 individual countries and 5 regional groups of smaller countries, which we downscale to country level. The first year in the time series is 1965, and the last is 2014. Appendix B gives details on the downscaling. We use the BP data additionally to sources covering similar gases and categories (e.g., CDIAC) because they offer emissions data for recent years which are not included in the other sources.

2.3.2 CDIAC fossil CO_2 (CDIAC2015)

The CDIAC fossil fuel and industrial CO_2 emissions dataset is published by the Carbon Dioxide Information Analysis

¹²When we write “all” there can still be a few exceptions where data are missing for single countries or sectors.

¹³No 2015 CRF data have been submitted by Belarus and Switzerland. Canada and the United States have submitted data but requested to not make them publicly available until problems with the CRF reporter software are solved.

Center (CDIAC) with regular updates (Boden et al., 2015). It covers emissions from fossil fuel burning, flaring, and cement production for 221 countries and territories. The first year is 1751 and the last year 2011. Emissions from 1751 to 1949 are computed using statistics of fossil fuel production and trade combined with information on the chemical composition of the fuels and assumptions on the use and combustion efficiency following the methodology presented in Andres et al. (1999). Emissions data for the years 1950 to 2011 are based primarily on the United Nations Energy Statistics Yearbook (United Nations, 2016) using the methodology presented in Marland and Rotty (1984). The data require some preprocessing to account for division and unification of countries. The preprocessing methodology and mapping of emissions categories are explained in Appendix B.

2.3.3 EDGAR versions 4.2 and 4.2 FT2010 (EDGAR42)

The EDGAR¹⁴ dataset is published by the European Commission Joint Research Centre (JRC) and Netherlands Environmental Assessment Agency (PBL). It undergoes regular updates. The current (1 February 2016) version is 4.2. It contains emissions data for all Kyoto greenhouse gases as well as other substances.¹⁵ It covers 233 countries and territories in all parts of the world, though not all countries have full data coverage. EDGAR version 4.2 covers the period 1970 to 2008 (JRC and PBL, 2011). Additionally the EDGAR v4.2 FT2010 covers the period 2000 to 2010 (JRC and PBL, 2013; Olivier and Janssens-Maenhout, 2012). EDGAR v4.2 FT2012 covers 1970 to 2012 but only for CO₂, CH₄, N₂O, and aggregate Kyoto GHG emissions with no sectoral resolution (JRC and PBL, 2014; UNEP, 2014). Version 4.3 covering the period until 2012 has been implicitly announced but is not yet available (as of 1 February 2016).¹⁶

EDGAR time series are calculated using activity data on a per sector, per gas, and per country basis. Emissions are calculated using a country, sector, and gas-specific technology mix with technology-dependent emission factors. The emission factors for each technology are determined by end-of-pipe measures, country-specific factors, and a relative emission reduction factor to incorporate installed emissions reduction technologies.

Appendix B contains information on the combination of EDGAR v4.2 and EDGAR v4.2 FT2010, as well as details on the category and gas basket aggregation and country preprocessing.

¹⁴Emissions Database for Global Atmospheric Research

¹⁵Some of the other substances in the EDGAR database are controlled under the Montreal Protocol (HCFCs), while others are not yet controlled (e.g., black carbon, organic carbon).

¹⁶In the “Trends in Global CO₂ emissions report” (Olivier et al., 2015), EDGAR v4.3 is referenced as forthcoming in 2015.

2.3.4 FAOSTAT database (FAO2015)

The Food and Agriculture Organization of the United Nations (FAO) publishes data with yearly values for emissions from agriculture and land use (Food and Agriculture Organization of the United Nations, 2015a). Over 200 countries and territories are included in the database.

The land use emissions are categorized into forestland, grassland, cropland, and biomass burning, where the first three categories contain information on CO₂ only, while biomass burning also contains information on N₂O and CH₄ emissions. To generate the time series, data from land use and forestry databases (both from FAO and other institutions) are used together with IPCC estimates of emission factors and the FAO “Global Forest Resources Assessment” database for carbon stock in forest biomass. For details we refer the reader to the methodology information on the FAOSTAT website (Food and Agriculture Organization of the United Nations, 2015b). The time series cover 1990 to 2012.

The land use emissions do not cover the emissions directly introduced by agriculture, but emissions from soil changes that are caused by agricultural use of the soil. For cropland FAOSTAT states that “greenhouse gas (GHG) emissions data from cropland are currently limited to emissions from cropland organic soils. They are those associated with carbon losses from drained histosols under cropland.” (Food and Agriculture Organization of the United Nations, 2016).

FAOSTAT data for agricultural emissions range from 1961 to 2012. They cover N₂O and CH₄ from various sources (e.g., rice cultivation, synthetic fertilizers, and manure management). Because the data cover a longer time period than other sources for the agricultural sector, we use them with highest priority after the country-reported data. The data are generated from activity data and emission factors following the tier 1 approach of the IPCC 2006 guidelines.

Appendix B gives details on the emissions categories and country processing.

2.4 Regionally resolved datasets

2.4.1 Houghton land use CO₂ (HOUGHTON2008)

This source covers land cover change CO₂ emissions from seven regions and three individual countries (USA, Canada, and China) for the years 1850 to 2005. The dataset is described in a series of papers by Houghton (2008, 2003, 1999). It is generated using a book-keeping model to track carbon in living vegetation, dead plant material, wood products, and soils. The carbon stock and its changes are taken from field studies. Information on changes in land use is mostly taken from agricultural and forestry statistics, historical records, and national handbooks. Emissions outside trop-

ical regions from 1990 onward are estimates (constants¹⁷). For our dataset the regional emissions have to be downscaled to country level. This is described in Sect. 4.2.2, while technical details are given in Appendix B.

The dataset covers only direct (deliberate) human-induced activities (Houghton, 2003, 1999). Thus, generally, forest fires are not included except for fire clearing. However, wild-fires and the effect of measures for fire suppression are included for the USA.

We use this dataset as a proxy for CO₂ emissions from land use, land use change, and forestry (LULUCF) as land cover change accounts for the majority of LULUCF emissions (Smith et al., 2014).

2.4.2 RCP historical data (RCP)

The representative concentration pathways (RCPs) were created for the CMIP5 intercomparison study of Earth system models that was organized by the World Climate Research Programme and used (among other models) in the IPCC's Fifth Assessment Report (AR5). They have a common historical emissions time series that covers all Kyoto gases but is only resolved at a coarse regional and sectoral level (Meinshausen et al., 2011b). For N₂O and fluorinated gases, only economy-wide global emissions are available. For CO₂, global emissions are split into land use and fossil and industrial emissions. CH₄ emissions are resolved into five regions with several subcategories of the IPCC 1996 categorization.

RCP historical data are compiled from a wide range of emissions sources and atmospheric concentration measurements. Where concentration data are used, inverse emissions estimates are computed using the MAGICC6 reduced-complexity climate model (Meinshausen et al., 2011a). RCP historical data can be used for extrapolation of country time series to the past using regional growth rates. RCP land use emissions data are not used in our dataset as they are based on the Houghton dataset, which we include directly (see previous section). Preprocessing of RCP data is explained in Appendix B.

2.5 EDGAR-HYDE 1.4 (EDGAR-HYDE14)

The EDGAR-HYDE 1.4 “Adjusted Regional Historical Emissions 1890–1990” dataset covers the gases CO₂, N₂O, and CH₄ for the years 1890 to 1995 (Olivier and Berdowski, 2001; Van Aardenne et al., 2001). The data are given for 13 regions, some of which are individual countries (USA, Canada, Japan). They are generated from the EDGAR v3.2 dataset (Olivier and Berdowski, 2001) and the “Hundred Year Database for Integrated Environmental Assessments” (HYDE v1.1) (Van Aardenne et al., 2001; Goldewijk and

Battjes, 1997). We use the EDGAR-HYDE dataset to extrapolate country emissions into the past. It has a relatively high sectoral detail, but the sectors differ from the IPCC 1996 definitions, so mapping to IPCC 1996 sectors is necessary. Details are presented in Appendix B.

2.6 Gridded datasets

The two gridded datasets included in the generation of the PRIMAP-hist dataset do not contain any emissions data. Instead, they contain data for potential vegetation and simulation data of past existing vegetation. By comparing these, we can determine areas where deforestation has occurred, which we use to downscale the Houghton land cover change emissions data to country level. More information on the use of these datasets is provided in Sect. 4.2.2.

2.6.1 HYDE land cover data (HYDE)

The HYDE land cover data (Klein Goldewijk et al., 2011, 2010; PBL, 2015) are generated using hindcast techniques and estimates on population development over the last 12 000 years. For the time period of interest here, they provide estimates of pasture and crop land on a 5 arcmin resolution grid for 10-year time steps. The data do not directly provide estimates for deforestation, but these can be computed by comparison with simulation data of potential vegetation (e.g., from SAGE; see below).

2.6.2 SAGE Global Potential Vegetation Dataset (SAGE)

This dataset is available in the SAGE¹⁸ database and is described in Ramankutty and Foley (1999) and available for download from Ramankutty and Foley (2015). It contains 5 arcmin resolution grid maps of potential vegetation (i.e., vegetation that potentially could be in a certain spot if there was no human interference) for a time period from 1700 to 1992. It has been used together with HYDE 3.1 in Matthews et al. (2014) to downscale CDIAC land use CO₂ emissions to country level. We use it for the same purpose here.

3 Source prioritization

To create a dataset covering all countries and gases for a period of over 150 years, multiple data sources need to be combined as no single source contains all the necessary data. We order sources such that the highest-quality sources are selected for each gas, category, and year, according to availability. Where possible, source prioritization is defined, and used, at a global level.

The source creation is carried out such that the absolute values are taken from the highest priority source, while

¹⁷The constant emissions outside tropical regions are obtained using the assumption that emissions calculated for 1990 are also valid for the subsequent years.

¹⁸Center for Sustainability and the Global Environment

lower-priority sources are used as year-to-year growth rates to extend the time series. The prioritization of the sources takes the completeness and reliability of the absolute values into account to use the most reliable absolute values and the year-by-year growth rates of the other sources to extend those data. A similar method is employed for the Global Carbon Budget (Le Quéré et al., 2015), where the motivation is that the growth rates are less uncertain than the absolute values. The details of the process of the combination of sources are described in Sect. 4.

3.1 Emissions from energy, industrial processes, solvent use, agriculture, and waste

For fossil emissions, our highest priority source is the UNFCCC CRF data as they are both accepted by the countries that report and by other countries because they are reviewed by experts. However, these data are only available for developed-country parties. We use CRF2014 and fill gaps with CRF2013 where necessary. For non-Annex I parties we use data from national communications and national inventory reports with highest priority (UNFCCC2015). For a few developing countries, data from the biennial update reports (BUR2015) are available and fulfill our minimal requirements. These are used to supplement the UNFCCC2015 data. UNFCCC2015 is prioritized over BUR2015 because the latter only contains a few data points for most countries, while the UNFCCC2015 data contain full time series for more countries. These sources of UNFCCC-reported data cover a wide range of gases and sectors. For most countries, CO₂, CH₄, and N₂O are available for all sectors at the level of detail needed for the composite source. Fluorinated gases are only contained for a few countries. For CO₂ related to fossil fuel burning, CO₂ from flaring, and CO₂ emissions from mineral products, we use CDIAC as the next priority source. For CO₂ from other sectors and all other gases we use a combination of EDGAR v4.2 FT2010 and EDGAR v4.2 as the next priority source. This combined EDGAR dataset is also used to complement CDIAC data where necessary (e.g., for small countries missing in the CDIAC source). BP2015 data are used to extend the energy CO₂ time series until 2014. Where no country-reported data are available, the country-resolved data sources are used as the first sources.

Sources without country-level information, i.e., RCP and EDGAR-HYDE, are used to extrapolate emissions into the past. As EDGAR-HYDE has a higher regional and sectoral resolution it is used as the first priority source for extrapolation of CO₂, CH₄, and N₂O emissions. Emissions from fluorinated gases for years before 1970 are only available from the RCP historical data and only on a global level.

The source prioritization for the individual gases is summarized in Tables 2, 3, 4, and 5. Details of the source creation methods are available in Sect. 4.1.

3.2 Land use, land use change, and forestry emissions

The first priority source for land use CO₂ is FAOSTAT. However, it does not contain information for the period before 1990. EDGAR42 does contain information starting in 1970 but excludes sinks from the calculation of CO₂ land use emissions, which is why we exclude EDGAR CO₂ land use data from our dataset. The period before 1990 is covered by the Houghton dataset on a regional level, which we downscale using estimates of historical deforestation (see Sect. 4.2).

For CH₄ and N₂O we use country-reported data, FAOSTAT, and EDGAR data on a per country basis. Regional growth rates from EDGAR-HYDE14 are used to extrapolate the time series.

Details of the source creation are presented in Sect. 4.2 and in Tables 6 and 7 within that section.

4 Dataset construction

4.1 Emissions from energy, industrial processes, solvent use, agriculture, and waste

The generation of the emissions time series is carried out using the composite source generator (CSG) of the PRIMAP emissions module described in Nabel et al. (2011). Data are aggregated on a per country, per gas, and per category level taking into account source prioritization (see Sect. 3). The result is one time series for each country, category, and gas. The source creation is organized in the four steps described below.

Source preprocessing First, each dataset undergoes source-specific preprocessing, e.g., category mapping and country preprocessing, which is explained in Appendix B. This is followed by category aggregation: if data are defined on a more detailed level of gases (in the case of HFCs and PFCs) or categories (e.g., categories 4A and 4B), they are aggregated to the resolution described above for all sources individually. The aggregate time series covers the union of all years of the individual gas or sector series. If data are missing for some years in any of the individual gas or subcategory time series, they are interpolated to close gaps and extrapolated to fill missing data at the boundaries before aggregation. After aggregation, the information that a subcategory or gas was missing is lost. If data are missing at the gas and category level we are working with in the PRIMAP-hist dataset, they are not interpolated in preprocessing as they can be filled from other sources.

Composite source generator The composite source generator (CSG) works on every country, gas, and category individually. Its core is the priority algorithm, which combines the sources following a given prioritization. The algorithm starts with the highest priority source. Missing time series are copied from lower-

Table 2. Source prioritization and extrapolation for fossil and industrial CO₂. In Fig. 3 we show the individual steps using the example of category 1 for the Republic of Korea. Years are maximal values. Some countries have less coverage. In CRF a few countries have data starting a few years before 1990. Category names refer to IPCC 1996 categories.

Step	Source	Categories	Countries	Years	Type of operation
1	CRF2014	all	Annex I	1990–2012	CSG
2	CRF2013	all	Annex I	1990–2011	CSG
3	UNFCCC2015	all	35 non-Annex I	1990–2010	CSG
4	BUR2015	all	8 non-Annex I	1990–2012	CSG
5	CDIAC2015	1A, 1B2, 2A	almost all	1850–2011	CSG
6	EDGAR42	all	almost all	1970–2010	CSG
7	BP2015	1A	almost all	1965–2014	CSG
8	EDGAR-HYDE14	1A, 1B1-2, 2A-G	regions	1890–1995	growth rates extrap.
9	RCP	1A	global	1850–2005	growth rates extrap.
10	PRIMAP-hist CAT1A	all but 1A	all	1850–2014	growth rates extrap.
11	numerical	all	all	1850–2014	linear extrapolation

Table 3. Source prioritization for fossil and industrial CH₄. Years are maximal values. Some countries have less coverage. In CRF a few countries have data starting a few years before 1990. Category names refer to IPCC 1996 categories. Note that there are no CH₄ emissions data in category 3 (solvent and other product use).

Step	Source	Categories	Countries	Years	Type of operation
1	CRF2014	all	Annex I	1990–2012	CSG
2	CRF2013	all	Annex I	1990–2011	CSG
3	UNFCCC2015	all	35 non-Annex I	1990–2010	CSG
4	BUR2015	all	7 non-Annex I	1990–2012	CSG
5	FAO2015	4	almost all	1961–2012	CSG
6	EDGAR42	all	almost all	1970–2010	CSG
7	EDGAR-HYDE14	1, 2, 4, 6	regions	1890–1995	growth rates extrap.
8	RCP	1, 2, 4, 6	global	1850–2000	growth rates extrap.
9	numerical	7	all	1850–2010	linear to zero in 1850
10	numerical	all	all	1850–2014	linear extrapolation

priority sources. After this step the priority algorithm fills gaps in the time series using lower-priority sources and extrapolates using year-to-year growth rates from lower-priority sources. The composite source time series for each gas, category, and country is checked for gaps and whether or not it covers the full time period. If that is the case, the second-highest priority source is checked for data that could fill gaps and extend the time series. If that time series itself contains gaps or needs extension, the default behavior of the CSG is to parse the hierarchy downwards recursively and to use the resulting time series to extend the composite source. For this study we add one source at a time and therefore do not parse the sources recursively but rather add what is present in the next priority source and then see whether the resulting time series needs further extension. For details on the harmonization of the lower-priority sources, see Appendix A4. If there are data missing after the end of this process, the CSG can numerically interpolate gaps and extrapolate missing data at the boundaries. For this dataset we only use the interpolation by the CSG,

because we use regional growth rates from other sources to extrapolate the country data. A schematic of the composite source generator within the PRIMAP emissions module is shown in Fig. 2.

The rationale underlying this combination method is that the absolute values are taken from the highest priority source, while lower-priority sources are only used for the dynamics of emissions. By scaling the lower-priority sources to match the higher-priority source, we retain the year-to-year growth rates of the lower-priority sources but adjust the absolute values to the highest priority source. For details see Appendix A4. Other options for harmonization are discussed in Sect. 8.

Extrapolation Missing years in the past are extrapolated using growth rates from regional data or data from other sectors and numerical extrapolation. The details depend on gas and sector and are described later in this section. Missing data in the future are extrapolated linearly using a 15-year trend. This usually affects up to 4 years, with very few exceptions where extrapolation is used

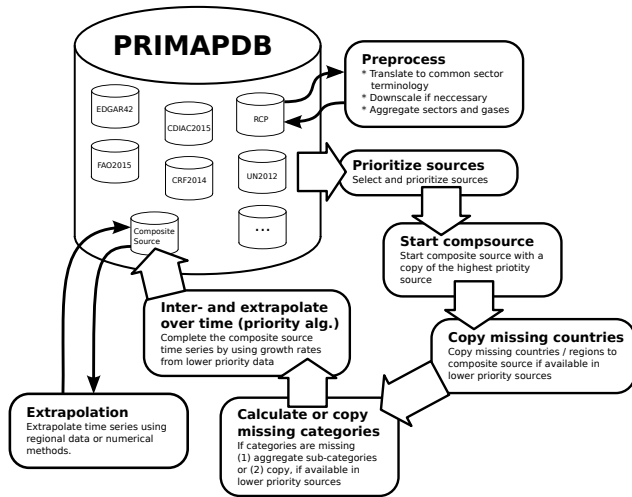


Figure 2. The composite source generator (CSG) is used to assemble time series from different sources into one time series covering all countries, sectors, gases, and years. The source prioritization in the figure is illustrative and does not represent the source prioritization for the dataset described here. In this study the internal category aggregation of the composite source generator is not used, but categories are aggregated before the generation of the composite source to enable extrapolation of subcategories. For the PRIMAP-hist dataset we always combine only two sources at a time instead of recursively filling missing data. Section 4.1 and Appendix A describe the use of the CSG for this dataset. Figure 3 shows the individual steps for an example time series.

for longer periods. We also offer a dataset without this numerical extrapolation.

When using extrapolation with growth rates from regions or other sectors, we make the assumption that these time series share growth rates with the unknown time series we want to determine through extrapolation. This assumption seems crude, but it is much more transparent than, for example, building a more complicated model to compute the time series. A more sophisticated model will likely also need some input data, such as population or economic data, to estimate an extrapolated time series. Such input data are also scarce for the time periods we need the extrapolation for (i.e., before 1960/1970). Numerical extrapolation, on the other hand, does not require any information for the time period for which we want to build our time series, and it only uses data from a range of years before or after the time period to be computed. It thus makes the assumption that we can deduce emissions in one time period from emissions in another time period, which is often not true. As an example, consider the Second World War, when emissions changed drastically, and that a numerical extrapolation would not model when using, for example, 1960–1980 as input data. However, a regional time series for Europe, for example, would have this

feature and would model emissions for European countries more realistically than numerical extrapolation. We still use numerical extrapolation for the PRIMAP-hist dataset, but only when it is the only option because no national or regional data exist.

Postprocessing After extrapolation the individual gas and category time series are aggregated to build the higher categories and the Kyoto GHG basket. For details on the aggregation, see Appendix A1.2.

Sudan needs a special treatment as the split into Sudan and South Sudan was so recent that no separate emissions data are available yet. We downscale the Sudan emissions time series to Sudan and South Sudan using UN population statistics (UN Population Division, 2015) as a downscaling key. We also aggregate country data for some regional groups.

Figure 3 shows an example of how we build a pathway from different time series.

In the following we describe the availability and use of datasets in detail for the different gases and sectors.

CO₂ Data coverage for CO₂ is, in general, very good. The largest emissions sources are the consumption and production of fossil fuels and the production of cement. Both are covered by CDIAC, which extends the country-reported data back to 1850 for 31 countries, to 1900 for 65 countries, to 1950 for 168 countries, and to 1990 for 196 countries. For other sectors EDGAR42 extends the time series back to 1970. BP data complete the fossil fuel consumption time series until 2014.

To further extend time series into the past we use EDGAR-HYDE regional growth rates (starting in 1890). For categories 1A, 1B1, and 1B2, explicit time series are available, while we use category 2 time series as a proxy for the subcategories of category 2. Other categories are not available. RCP CO₂ data that range back until 1850 are only available for total emissions excluding LULUCF on a global level. As total CO₂ emissions are dominated by fossil fuel burning, we use the RCP data as growth rates to extrapolate category 1A emissions for those countries that were not covered by CDIAC and EDGAR-HYDE from 1850 onwards. This does not affect any major emitter at the time for which data are extrapolated. For categories 3, 4, 6, and 7, no source for extrapolation is available, so the first year is 1970 from EDGAR. We use growth rates of the fossil fuel consumption time series for each country as a proxy to extend the time series of all other sectors to 1850.

The source prioritization and extrapolation is summarized in Table 2. Details of the growth rate extrapolation are discussed in Appendix A5.1.

CH₄ We have data on a per country level from 1990 to 2010 or 2012 from the country-reported data. For agriculture

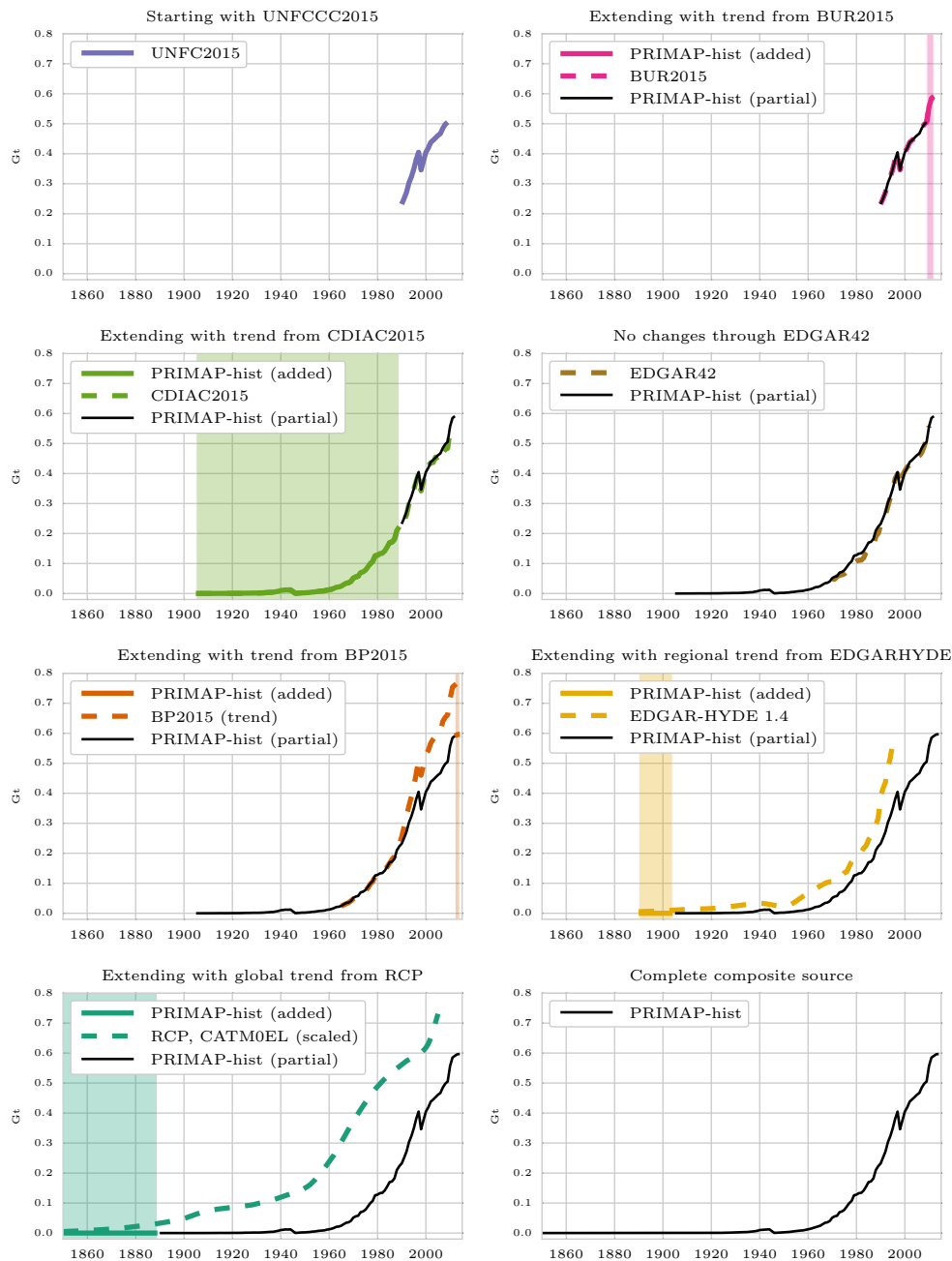


Figure 3. Example for the work of the composite source generator: the creation of the category 1A, CO₂ pathway for the Republic of Korea. The buildup starts with the UNFCCC source as there are no CRF data for the Republic of Korea. Extrapolation is not needed in this case, so the step is omitted from the figure. Details on the methodologies for the individual steps are given in Sects. 4.1 and A4 and A5 of the Appendix. The individual steps shown here correspond to the steps shown in Table 2.

(category 4) we have FAOSTAT data where the first year is 1961 and the last year 2012. For all other sectors and missing countries we use EDGAR42, which covers 1970 to 2010 for almost all countries. Categories 1, 2, 4, and 6 are extrapolated back to 1890 using the regional growth rates from EDGAR-HYDE. The regional growth rates defined in the RCP historical database are

used to extrapolate emissions in categories 1, 2, 4, and 6 back to 1850. Emissions in category 7 are extrapolated backward using a linear decline to zero in 1850 from the last year with data starting from a 21-year linear trend. In category 3 there are no CH₄ emissions reported. The source prioritization and extrapolation is summarized in Table 3.

N₂O Country-reported data cover 1990 to 2012 for all Annex I countries and some non-Annex I countries. Using EDGAR42 we obtain per country data from 1970 until at least 2010 for all sectors and countries. For agriculture (category 4), the first available year is 1961 from the FAOSTAT dataset and the last year is 2012 for all countries. For the period 1890 to 1970 we use the regional growth rates from the EDGAR-HYDE dataset to extrapolate categories 1, 2, 4, and 6. For the period prior to 1890, the RCP database provides data, but only at a global level and without sectoral detail. We know of no source that provides regionally or sectorally resolved N₂O emissions prior to 1890. The main contribution to N₂O emissions comes from the agricultural sector, especially the use of manure and nitrogen fertilizers (Davidson, 2009). N₂O emissions are therefore not well correlated with CO₂ or CH₄ emissions as these have different sources and thus they cannot be used as a proxy for N₂O emissions. Data on fertilizer use are only available for a few countries for years earlier than 1961 (Federico, 2008). This is not sufficient for downscaling of agricultural N₂O emissions. We therefore use the RCP global growth rates, which are computed from atmospheric concentration measurements to extend the country time series into the past for all sectors. The source prioritization and extrapolation is summarized in Table 4.

Fluorinated gases Country-reported data cover 1990 to 2012 for all Annex I countries and some non-Annex I countries. Other countries are added from EDGAR 42, which also extends existing time series to start in 1970. To extrapolate the data to 1850 we use RCP global growth rates. RCP data and global emissions from EDGAR data are in very good agreement for the time of overlap of the two sources for SF₆, HFCs, and PFCs. The time series are obtained using different methods: EDGAR from activity data and emission factors, and RCP from inverse emissions estimates based on atmospheric concentration measurements. This is a good sign with respect to the uncertainty in the datasets. Because of the similarity in absolute emissions, using RCP growth rates to extend EDGAR data does not significantly alter the global emissions compared to the RCP and is a safe method to obtain emissions for the first years of use of fluorinated gases. Emissions from fluorinated gases are generally very low before 1950 as their large-scale production and use only started in the second half of the 20th century. Technology for large-scale production of HFCs was developed in the late 1940s. For PFCs, a major breakthrough in industrial production was the Fowler process, which was published in 1947 and industrial production of SF₆ began in 1953 (Levin et al., 2010). The IPCC “Special Report on Safeguarding the Ozone Layer and the Global Climate Sys-

tem” (Metz et al., 2007) estimated emissions from most HFCs to be zero in 1990, with a steep rise afterwards. However, this is not in agreement with other sources like EDGAR and RCP, which show significant HFC emissions before 1990. As EDGAR and RCP agree at the HFC emissions levels, we use the nonzero emissions before 1990. The source prioritization and extrapolation is summarized in Table 5.

Data for individual fluorinated gases will be provided in a future release of this dataset. Currently this is not possible as some of the sources we use only provide aggregate HFC and PFC emissions (UNFCCC2015).

4.2 Emissions from land use

The largest share of emissions from land use, land use change, and forestry (LULUCF) is in the form of CO₂ originating from deforestation.¹⁹ We therefore focus on CO₂ emissions and use a simpler method for CH₄ and N₂O emissions. The preparation of the LULUCF pathways follows the same steps as for the fossil fuel and industry pathways. However, due to the high fluctuations in LULUCF data, the harmonization of sources is problematic (e.g., when one source shows a sink while another source shows emissions for the same period of time). We therefore use the time series from different datasets directly without harmonization. In the pre-processing, the Houghton source needs to be downscaled, which is described below.

4.2.1 Composition of the land use CO₂ pathways

We do not use country-reported data for CO₂ as there are several ways to calculate anthropogenic land use emissions. Developed countries in particular use this freedom to choose an accounting method that results in high CO₂ removals which are in contrast to third-party sources. We use FAOSTAT data, which are available for almost all countries for 1990 to 2012. The period before 1990 is covered by the Houghton dataset, which uses 10 regions. In general, the period from 1850 to 2005 is covered, but for non-tropical regions the latest years are estimates based on constant extrapolation of the last data point. In some cases this starts as early as 1990. As the period from 1990 on is covered by FAOSTAT data, this is no problem for us. If countries still have missing data, we extrapolate into the past using a linear pathway to zero emissions in 1850. The starting point of the extrapolation is a 21-year average. We use 0 in 1850 to rather under- than overestimate emissions when extrapolating. Linear or even exponential extrapolation is difficult for land use because of the strong fluctuations, which strongly influence the trend that is needed

¹⁹The IPCC AR5 WG3 states that “fluxes resulting directly from anthropogenic FOLU (forestry and other land use) activity are dominated by CO₂ fluxes, primarily emissions due to deforestation, but also uptake due to reforestation/regrowth”. (Smith et al., 2014)

Table 4. Source prioritization for fossil and industrial N₂O. Years are maximal values. Some countries have less coverage. In CRF a few countries have data starting a few years before 1990. Category names refer to IPCC 1996 categories.

Step	Source	Categories	Countries	Years	Type of operation
1	CRF2014	all	Annex I	1990–2012	CSG
2	CRF2013	all	Annex I	1990–2011	CSG
3	UNFCCC2015	all	35 non-Annex I	1990–2009	CSG
4	BUR2015	all	8 non-Annex I	1994–2010	CSG
5	FAO2015	4	almost all	1961–2012	CSG
6	EDGAR42	all	almost all	1970–2010	CSG
7	EDGAR-HYDE14	1, 2, 4, 6	regions	1890–1995	growth rates extrap.
8	RCP	all	global	1850–2005	growth rates extrap.
9	numerical	all	all	1850–2014	linear extrapolation

Table 5. Source prioritization for fluorinated gases. Years are maximal values. Some countries have less coverage. In CRF a few countries have data starting a few years before 1990. Category names refer to IPCC 1996 categories. Fluorinated gas emissions are only reported in category 2. For some countries, data in the BUR and UNFCCC sources are only available for SF₆.

Step	Source	Categories	Countries	Years	Type of operation
1	CRF2014	2	Annex I	1990–2012	CSG
2	CRF2013	2	Annex I	1990–2011	CSG
3	UNFCCC2015	2	7 non-Annex I	1990–2009	CSG
4	BUR2015	2	2 non-Annex I	1990–2012	CSG
5	EDGAR42	2	almost all	1970–2010	CSG
6	RCP	2	global	1850–2005	growth rates extrap.
7	numerical	2	all	1850–2014	linear extrapolation

for the extrapolation. This extrapolation is only used for very few small countries. Extrapolation to the future uses a constant derived from the average emissions of the last 15 years. Table 6 summarizes the source creation.

4.2.2 Downscaling of HOUGHTON2008

The Houghton source only resolves 10 regions: Canada, China, Europe, former USSR, northern Africa and the Middle East, Pacific developed countries, South and Central America, South and Southeast Asia, tropical Africa, and the USA. Data for all countries except Canada, China, and the USA therefore have to be computed using downscaling of regional emissions.

As land use emissions are not correlated well with emissions from other sectors we cannot use fossil and industrial emissions as a proxy. Instead, we use estimates of the conversion of forests into cropland and pasture (deforestation), which is the main source of land use emissions. The methodology we use is based on an approach recently published by Matthews et al. (2014). Estimates of historical deforestation can be computed starting from models of the amount of cropland and pasture required to feed the population in a certain area at a certain time. This time series gives estimates of the land converted to cropland or pasture in that area. Using a dataset of potential natural vegetation (i.e., simulated vegetation in the absence of human interference like deforestation),

we compute the fraction of land that was likely covered by forests before the conversion. This gives us a time series of deforested areas on a grid map of the world. The gridded data are transferred into country data using country masks.

The cropland and pasture data are taken from the History Database of the Global Environment (HYDE). We use the SAGE Global Potential Vegetation Dataset to get an estimate of historical forest cover in the absence of human interference. The potential vegetation in this dataset is representative of what vegetation cover would be if anthropogenic interference were removed from the climate and vegetation state observed in the mid-1990s. It therefore does not account for any historic changes in forest area driven by changing climate or atmospheric CO₂ concentrations. The SAGE dataset contains 15 separate plant function types (PFTs), of which eight forest/woodland types were combined to generate a simple forest cover mask. The SAGE dataset also includes a PFT for savanna, which we included in the “non-forest” category. Although loss of biomass from savanna land has contributed to historical CO₂ emissions, we chose to exclude it from this dataset because the carbon density is substantially different to that of forest or woodland areas occurring in the same region. The CO₂ emissions downscaling scheme assumes uniform carbon density of vegetation throughout each region, so savanna was excluded to avoid skewing results. While the different forest PFTs also have different carbon contents, the

Table 6. Source prioritization for CO₂ from LULUCF. Years are maximal values. Some countries have less coverage. Category names refer to IPCC 1996 categories. Linear to zero extrapolation is only used for the Netherlands Antilles and Pitcairn Islands.

Step	Source	Categories	Countries	Years	Type of operation
1	FAOSTAT	5	almost all	1990–2010	copy
2	Houghton downsc.	5	almost all	1850–2005	copy
3	numerical	5	see caption	1850–2000	linear to zero in 1850
4	numerical	5	all	1850–2014	linear extrapolation

Table 7. Source prioritization for CH₄ and N₂O from LULUCF. Years are maximal values. Some countries have less coverage. In CRF a few countries have data starting a few years before 1990. Category names refer to IPCC 1996 categories.

Step	Source	Categories	Countries	Years	Type of operation
1	CRF2014	5	Annex I	1990–2012	copy
2	CRF2013	5	Annex I	1990–2011	copy
3	UNFCCC2015	5	16 non-Annex I	1990–2009	copy
4	BUR2015	5	3 non-Annex I	1990–2012	copy
5	FAOSTAT	5	almost all	1990–2012	copy
6	EDGAR42	5	almost all	1970–2010	copy
7	EDGAR-HYDE14	5	global	1850–2000	growth rates extrap.
8	numerical	5	global	1850–2000	linear extrap. (past)
9	numerical	5	all	1850–2014	linear extrap. (future)

variability within a region is much smaller than the difference between forest PFTs and savanna within one region. See, for example, Fig. 1 of Liu et al. (2015).

The area converted to agricultural land, defined as the sum of cropland and pasture, and that coincides with land that would otherwise be forested is calculated to determine the areal extent of deforestation, as well as reforestation, over 10-year time steps for each grid cell. Spatial data are converted to country time series using an area-weighted summation according to the country boundaries data of the Food and Agriculture Organization of the United Nations (2015c). See also Fig. 4.

To downscale the regional emissions data, we make the assumption that forests in a region have the same average carbon content. Therefore, for any two countries in a region, we assume that converting 1 ha of forest into cropland in one country releases the same amount of CO₂ to the atmosphere as converting 1 ha of forest in the other country. The time-resolved data exhibit strong fluctuations, which do not necessarily coincide with fluctuations in the emissions data. One reason for this is the different methodological approaches used to create the two datasets. While the Houghton dataset models actual emissions from deforestation in detail, the method to calculate deforested area uses datasets that are of more theoretical nature. The HYDE dataset models the need for agricultural area in a region and does not represent the agricultural area that was actually present at that time. When population changes, the need for agricultural area changes with it, but the actual agricultural area changes more slowly. This is especially visible in Europe during the Second World

War. Population, and thus the need for agricultural area, declined rapidly, leading to afforestation in the SAGE-HYDE model. In reality, agricultural area will remain unused for some time until it is actively afforested or natural vegetation returns and takes up carbon from the atmosphere. This leads to situations where the Houghton source has positive emissions, while the SAGE-HYDE calculation shows an increase in forest cover indicating CO₂ removals. This sign discrepancy causes problems for downscaling (e.g., instability if some countries in a region show afforestation and some deforestation and a general problem of interpreting the shares in afforestation to calculate shares in deforestation emissions). To solve this problem, we do not use yearly shares but instead cumulative shares in deforestation for the whole period of 1850 to the last data year in the Houghton source in order to downscale the regional emissions to country level. This approach is also taken in Matthews et al. (2014). Details are given in Appendix B.

4.2.3 Composition of the land use CH₄ and N₂O pathways

For non-CO₂ emissions from land use we use country-reported data, which are complemented by FAOSTAT and EDGAR42 for the period from 1970 to 2010. Regional data from EDGAR-HYDE14 are used to extrapolate the time series into the past until 1890 starting from the 1969 value of the 30-year linear trend from 1970 to 1999. For 1850 to 1890 we use a simple linear extrapolation for each gas (CH₄, N₂O)

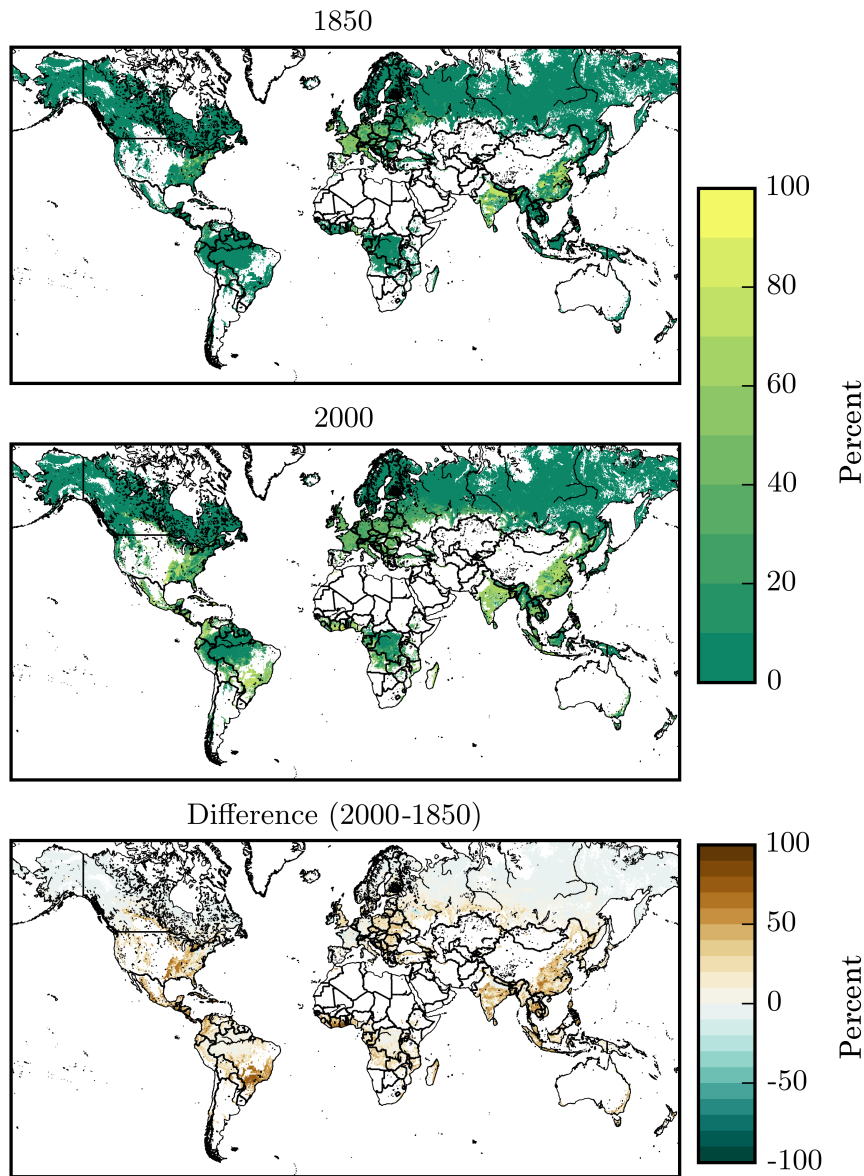


Figure 4. Calculating deforested areas: the two upper plots show the area potentially covered by forests (colored) and the fraction that has been cut until 1850 and 2000 according to the SAGE and HYDE datasets. The third plot shows the difference between the 1850 and 2000 deforestation and thus the area deforested or reforested between 1850 and 2000, which we use to downscale the Houghton dataset.

using the 21-year linear trend of the emissions from 1890 to 1910.

4.3 Territorial definitions, changes, and missing data

The dataset provides emissions time series for all UNFCCC member states. Some territories are associated with states but have partial independence, while other territories claim independence but are not internationally recognized, or have another special status. We include the emissions from these territories in the country emissions if, and only if, the country includes the emissions when reporting under the UN-

FCCC. Territories not included in the country reporting are treated independently. However, we cannot provide time series for all such territories. Territories which are uninhabited or have only very few inhabitants, e.g., in a research station, and with no significant emissions are completely excluded from the dataset (Bouvet Island, South Georgia and the South Sandwich Islands). In Table 8 we show which territories are included in countries, which are treated independently and whether data are available for those territories treated independently. The only territory that is not somehow associated with a single UNFCCC party is Antarctica. It is included in

Table 8. Territorial definitions of countries used in the dataset. The territorial definitions are based on country emissions reporting under the UNFCCC and do not imply any political judgment.

Country	Countries/territories/dependencies included	Countries/territories/dependencies with independent data	Countries/territories/dependencies without data
Australia	Norfolk Island; Christmas Island; Cocos Islands; Heard and McDonald Islands		
China		Hong Kong; Macao; Taiwan	
Denmark	Faroe Islands; Greenland		
Israel	Palestinian territories		
France	Saint Barthélemy; Guadeloupe; French Guiana; Saint Martin; Martinique; Mayotte; New Caledonia; French Polynesia; Réunion; Saint Pierre and Miquelon; Wallis and Futuna; French Southern and Antarctic Lands		
Finland	Åland Islands		
Morocco	Western Sahara		
Netherlands		Aruba; Netherlands Antilles (Bonaire; Curacau; Saba; Sint Eustatius; Sint Maarten)	
New Zealand			Tokelau
Norway	Svalbard		
United Kingdom	Bermuda; Cayman Islands; Channel Islands; Falkland Islands (Malvinas); Gibraltar; Guernsey; Isle of Man; Jersey; Montserrat	Anguilla; British Indian Ocean Territory; Pitcairn Islands; Saint Helena, Ascension and Tristan da Cunha; Turks and Caicos Islands; British Virgin Islands	
United States	Guam; Northern Mariana Islands; Puerto Rico; American Samoa; US Virgin Islands		

the dataset despite its negligible anthropogenic greenhouse gas emissions.

As a result of the Ukraine crisis, parts of the (former) Ukrainian territory are currently claimed by both Russia and Ukraine. The UN has not recognized any changes to the Ukrainian territory, so we do not make any adjustments to the Ukrainian emissions. There are no country-reported data recent enough to be influenced by the crisis.

We use territorial accounting in this dataset, meaning that emissions that originated from a territory that is now part of country A are always counted as emissions from country A even if the territory belonged to country B in the year the emissions took place. However, we can only be as precise as the datasets we are working with. Unfortunately, many sources are not very precise with respect to the methodology used. CDIAC CO₂ and, to a lesser extent, FAO data are

somewhat of an exception, where splitting up and merging of countries is made transparent by issuing different country codes. We sum and downscale the data to match the current countries according to the methodology described in Appendix A3. The CDIAC dataset also tries to account for land exchanges between countries. The CDIAC publication Andres et al. (1999) states that “land exchanges between countries were also accommodated, when possible. For example, the emissions from Alsace-Lorraine were included with Germany or France, reflecting which political unit governed these lands at any given time. This maintained the integrity of political entities despite changes in national borders.” This is not reflected in the country codes and thus remains in the final PRIMAP-hist dataset, in contrast to the territorial accounting used in our methodology. We cannot quantify the influence of this accounting discrepancy, because we do not

Table 9. Uncertainties for fossil fuel and industrial CO₂ emissions for different country groups. All values from Andres et al. (2014).

Country class	OECD	European countries outside of OECD	OPEC	Developing countries with stronger statistical bases (e.g., India)	Former USSR and eastern Europe	China and centrally planned Asia	Developing countries with weaker statistical bases (e.g., Mexico)
Uncertainty (95 % confidence)	4 %	6.7 %	9.4 %	12.1 %	14.8 %	17.5 %	20.2 %

know which regions were affected. However, as the land exchange including large emitters has been small in the recent decades and emissions were relatively low before the recent decades, the influence will likely be small. CRF2014, UN-FCCC2015, and BUR2015 data are reported by countries and do not require preprocessing as we use the territorial definitions of the UNFCCC reporting as a basis. For EDGAR data, the rules regarding how emissions are assigned to countries in the case of territorial changes are not clear from the methodology description and we assume that territorial accounting is used.

For some small countries and countries that recently became independent, no emissions data are currently available. In this case we have to construct time series using other countries' emissions data. Emissions data for San Marino and the Vatican are included in Italian emissions data and downscaled using population shares.²⁰ Downscaling is performed on the individual sources during preprocessing (for preprocessing details, see also Appendix B). For details on the downscaling methodology see Appendix A3. Sudan and South Sudan are also downscaled from emissions of former Sudan using UN population data (UN Population Division, 2015).

5 Data availability

The dataset is available from the GFZ Data Services under doi:10.5880/PIK.2016.003 (Gütschow et al., 2016). When using this dataset or one of its updates, please cite this paper and the precise version of the dataset used. Please also consider citing the relevant original sources when using this dataset. Any use of this dataset should also comply with the usage restrictions of the original data sources used for this project.

6 Results

In this section we show some key results of our analysis. Details for additional countries, sectors, and gases can be explored online on our companion website

²⁰GDP data not available.

<http://www.pik-potsdam.de/primap-live/primap-hist/>. Here we focus on major emitters and global emissions.

6.1 Sectoral distribution of aggregate Kyoto greenhouse gas emissions for major emitters

Globally, production and consumption of fossil fuels is responsible for about two-thirds of current aggregate Kyoto greenhouse gas emissions²¹, which is an increase from about 50 % in 1950 and a negligible contribution in 1850. This is shown in the upper left panel of Fig. 5. Before the Industrial Revolution, deforestation was the major emissions source followed by agriculture. Currently, these sectors are the second- and third-largest sources. Roughly 10 % of emissions come from waste and industrial processes. Industrial processes increased their share in yearly emissions after 1950, while the share of waste-related emissions stayed relatively constant.

The sectoral profile differs strongly among countries (Fig. 5). Land use emissions reached almost zero or even negative values in the 1950s to 1970s in industrialized countries (USA, EU, Japan) and a few decades later in China. For all these countries, fossil fuel use and production are by far the largest contributors to total emissions. While the industrialized countries have decreasing (USA, EU) or stagnating (Japan) fossil fuel emissions, China has rapidly increasing emissions. The increase in emissions from China may have slowed down in the last years, but more time is needed to say whether this is more than a temporary effect (Korsbakken et al., 2016).

India still has a large share of LULUCF emissions with no clear increase or decrease in the last two decades. Agriculture and LULUCF have similar emissions both in trends and absolute values, which have only recently (roughly 1990) been surpassed by the steeply increasing fossil-fuel-related emissions. For Brazil the largest sector is land use, followed by agriculture. Land use emissions show a decreasing trend, but total emissions do not follow this trend due to a rise in agricultural emissions and fossil-fuel-related emissions.

²¹In the remainder of this section the term “emissions” refers to aggregate Kyoto GHG emissions.

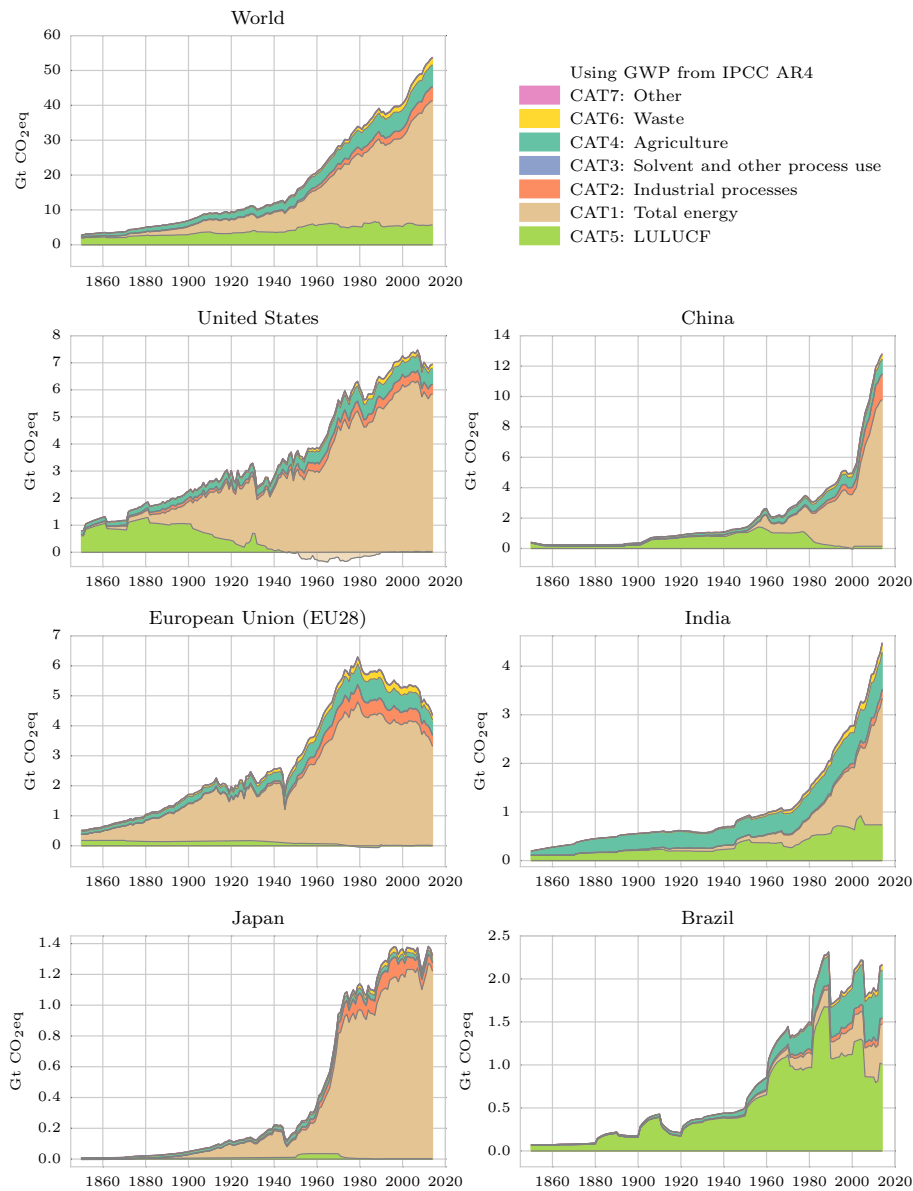


Figure 5. Aggregate Kyoto greenhouse gas emissions by sector for major emitters and the world. Where land use emissions are negative, the stacked emissions of the other sectors start at this negative value. International shipping and aviation emissions are not included. The figure is discussed in Sect. 6.1.

Waste gives a small contribution, differing by country without a clear split between developed and developing countries. The contribution of industrial processes is larger in industrialized countries, but especially large in China.

6.2 Gas distribution of economy-wide emissions for major emitters

The contribution of individual gases and gas groups to (global warming potential weighted) economy-wide (IPCC 1996 category 0) emissions is shown in Fig. 6. It is clearly visible that CO₂ is by far the largest contributor, fol-

lowed by CH₄ and N₂O, both globally and for individual countries. The contribution of fluorinated gases is, in general, small and negligible for developing countries. Again, China's emissions profile is closer to that of an industrialized country than to other major developing-country emitters. Economy-wide methane emissions are high for countries with a large agricultural sector (India and Brazil). Japan is somewhat of an exception with almost all emissions from CO₂.

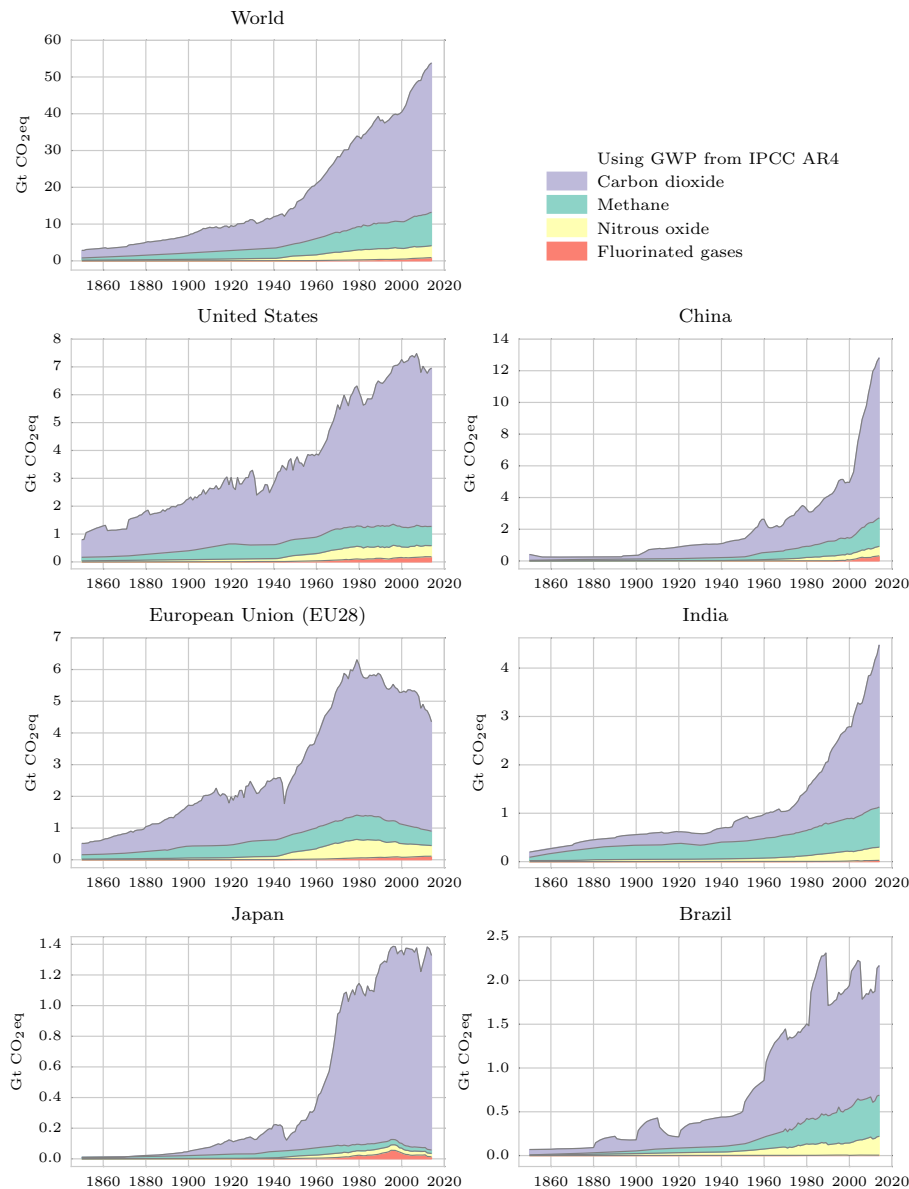


Figure 6. Economy wide (IPCC 1996 category 0) emissions by gas for major emitters and the world. International shipping and aviation emissions are not included. The figure is discussed in Sect. 6.2.

7 Uncertainties

In this paper we do not assess the uncertainties of the dataset in detail. Of the individual datasets used, uncertainty information is available for some, while for others it is not provided. Where it is available, the level of detail is very different. Some datasets give per country or per regional group uncertainty estimates, while others only provide global estimates. Individual uncertainty estimates can be over 100% (Olivier et al., 1999). To calculate uncertainty estimates for all countries, gases, and sectors for the composite source, one has to transform the information given for the individual sources to a common methodology and level of detail

and combine it in line with the creation of the composite source. As most datasets come without an uncertainty estimate and third-party estimates are scarce for some datasets, it is difficult to find a consistent set of uncertainty estimates. Furthermore, different studies use different sectoral resolutions, confidence intervals, etc., which makes it difficult to compare and combine the results to arrive at an estimate for our aggregate source. We leave this task for a future publication. In the following, we give a broad overview of the uncertainties of individual sources and present an indicative uncertainty range for individual gases and sectors based on literature values. We plot this source together with input data

and the indicative uncertainty range to reveal differences between sources and identify possible problems (Figs. 7 and 8).

7.1 Uncertainties from individual sources

Uncertainty estimates for the CDIAC dataset of global CO₂ emissions from fossil fuels and industry have varied since the first assessment made by Marland and Rotty (1984), which resulted in an uncertainty range between 6 and 10 % (using a 90 % confidence interval). In a recent publication, a single global fossil fuel CO₂ emissions uncertainty of 8.4 % (using a 95 % confidence interval) is offered as a reasonable combination of data (Andres et al., 2014), in an attempt to simplify the different assessments and to make the best of the qualitative and quantitative knowledge developed since the first study of 1984.

Different approaches examine CDIAC global uncertainty as the aggregate of the uncertainties associated with fossil fuel CO₂ emissions from individual countries. In Andres et al. (1996) a country grouping was introduced that uses seven classes of countries with “similar perceived uncertainty”. Andres et al. (2014) calculated uncertainty estimates for these groups, which are presented in Table 9.

The authors of the EDGAR dataset have stated that it was not feasible to go beyond the uncertainty tables compiled for EDGAR v2.0, where uncertainties are indicated in terms of ranges ranking from small (10 %) to very large (> 100 %) (PBL, 2010). However, other institutions, such as UNEP (UNEP, 2012), estimated an uncertainty range of ± 10 % (for a 95 % confidence interval) for total CO₂ (including LULUCF). For global emissions of CH₄, N₂O, and fluorinated gases, uncertainties are estimated to be ± 25 , ± 30 , and ± 20 %, respectively (using a 95 % confidence interval) (UNEP, 2012).

FAOSTAT land use emissions estimates are limited to only two carbon pools (above- and belowground biomass) out of six identified by the IPCC guidelines (above- and belowground, dead wood, litter, soil organic carbon, and harvested wood products). Therefore, FAOSTAT estimates greenhouse gas emissions and removals from land use are likely underestimated (Federici et al., 2015).

Tubiello et al. (2015) provides overall uncertainty estimates of the FAOSTAT database, where global emissions estimates from crop and livestock carry ± 30 % uncertainty ranges. Uncertainties in the land use sector are even larger, with a ± 50 % range.

Table 10 gives an overview of available uncertainty estimates for the individual sectors and gases included in the PRIMAP-hist dataset and how we calculated the indicative uncertainty range used in Figs. 7 and 8 for different sectors and gases.

7.2 Comparison with other data sources

A different approach at uncertainty estimates is to compare different datasets. If they were completely independent, the distribution of emissions for the same category and gas should represent the uncertainties. This approach also captures uncertainties from different definitions of sectors, which are not included in the uncertainties of individual datasets. Some sources used in the PRIMAP-hist dataset depend on each other or may use common underlying data, so we cannot determine an upper bound on uncertainty but rather a lower bound. Adding independent sources would likely increase uncertainty. In Figs. 7 and 8 we plot the composite source alongside some of the individual sources and other composite sources for individual gases and sectors at a global level. To compare the lower bound of the inter-source uncertainty to the individual source uncertainty, we also plot an indicative uncertainty range from Table 10 around the PRIMAP-hist dataset. For most categories and gases, it is apparent that the inter-source uncertainty is lower than the uncertainty estimated for the individual sources. However, as we have some source interdependence, we cannot conclude that the individual uncertainties are overestimated. Additionally, the number of sources is too small to reliably sample the 95 % confidence interval of the individual source uncertainty.

In the following, we investigate discrepancies between sources for total emissions, as well as individual gases and sectors, to analyze whether the discrepancies result from different assumptions and underlying data or lack of data for subsectors or individual gases. The EDGAR-HYDE data have relatively low total Kyoto GHG values. The sector plots show that this is due to low values for industrial processes and land use emissions. The low industrial process emissions can partly be explained by the lack of data for fluorinated gases in the EDGAR-HYDE dataset, but emissions of CH₄ and CO₂ are also low. Land use CO₂ emissions in the EDGAR-HYDE dataset are only about half of the emissions of all other datasets assessed and outside of the sizable uncertainty range applied to the PRIMAP-hist time series. We should note that RCP, MATCH, and PRIMAP-hist include HOUGHTON data in their land use time series and are therefore not independent. The HOUGHTON-based time series are consistent with EDGAR42 and FAO, while the EDGAR-HYDE time series is not similar to any of the time series for more recent emissions.

A further major discrepancy is the RCP CH₄ time series, which differs strongly from all other sources. Emissions are significantly higher than in other sources but show a steep decline between 1990 and 2000. No other source used in this analysis shows this effect. RCP CH₄ emissions are based on Lamarque et al. (2010), which uses EDGAR-HYDE but adds

Table 10. Uncertainties for gases and sectors covered in the PRIMAP-hist dataset. “NA” indicates that there are no emissions from this gas and sector combination. “–” indicates that we have no uncertainty estimate for the gas–sector combination. Where different uncertainty estimates exist, we calculate both upper and lower bound aggregate uncertainties. Calculations have been carried out according to the IPCC tier 1 methodology using average global emissions gas shares of the period 1990–2014. All calculated values are rounded to the nearest multiple of 5 % except for CO₂ values, which are rounded to 1 % with the exception of category 1, where rounding is to 0.1 %. For category 5 the high and low uncertainty cases are the same within rounding, so only one number is given.

Category	CO ₂	CH ₄	N ₂ O	Fl. gases	Kyoto GHG
0	10 % ^b –20 % ^e	25 % ^b –70 % ^e	30 % ^b –90 % ^e	20 % ^b	25 % ^e –35 % ^e
0EL	8.4 % ^a –14 % ^e	45 % ^e –55 % ^e	35 % ^e –65 % ^e	20 % ^b	20 % ^e –30 % ^e
1	12.5 % ^e	25 % ^g	25 % ^g	NA	15 % ^e
1A	12.6 % ^a	–	–	NA	–
1B1	6 % ^f	–	–	NA	–
1B2	6 % (25 % for 1B2C2) ^a	–	–	NA	–
2	23 % ^e	10 % ^e	50 % ^e	20 % ^b	25 % ^e
2A	23 % ^a	10 % ^c	NA	NA	–
2B	–	10 % ^c	50 % ^c	NA	–
2C	–	10 % ^c (2C1)	–	–	–
2D	–	–	–	NA	–
2E	NA	NA	NA	–	–
2F	NA	NA	NA	–	–
2G	–	–	–	–	–
3	10 % ^g	NA	30 % ^g	NA	15 % ^e
4	30 % ^d –100 % ^c	30 % ^d –100 % ^c	30 % ^d –100 % ^c	NA	30 % ^e –100 % ^e
5	50 % ^d	50 % ^d –75 % ^c	50 % ^d –100 % ^c	NA	50 % ^e
6	100 % ^f	100 % ^c	100 % ^f	NA	100 % ^e
7	100 % ^f	100 % ^f	100 % ^f	NA	100 % ^e

The references are ^a Andres et al. (2014); ^b UNEP (2012); ^c Olivier et al. (1999); ^d Tubiello et al. (2015); ^e calculated from available data for subsectors and gases; ^f estimated, no data available; ^g category 0 uncertainty value from ^b used.

information for some sectors missing in EDGAR-HYDE14, namely grassland and forest fire emissions.²² However, the discrepancies cannot fully be explained by this as they are also present in other sectors than land use.

For N₂O, MATCH and EDGAR42 economy-wide emissions are lower than the PRIMAP-hist dataset while EDGAR-HYDE14 and RCP are higher. MATCH is based on EDGAR-HYDE growth rates prior to 1990, which explains the very similar pathway profiles and leads to very low emissions before 1970.

Finally, the estimates of emissions of fluorinated gases are higher for EDGAR42 than for our aggregate dataset in the period 2000–2014. This indicates that, for recent years, country reported fluorinated gas emissions are significantly lower than what EDGAR calculates the emissions to be.

Not all discrepancies between sources could be explained, and some are larger than the indicative uncertainty range for an individual source. This indicates that the actual uncertainties of emissions data could be even higher than what is assessed for individual sources.

7.3 Uncertainties from methodology

The creation of this composite dataset implies several decisions on source prioritization, extrapolations, and downscaling options. These questions usually do not have one “correct” solution but rather different options with individual benefits and drawbacks. Different options (e.g., linear or constant extrapolation) have different implications for the calculated emissions, so the decisions introduce an “expert judgment uncertainty” to the final dataset. A further source of uncertainty is the use of regional growth rates for extrapolation. This assumes that all countries within that region shared the same growth rates, which is a simplification. Similarly, downscaling uses simplifications such as constant emissions shares or the use of another source as a proxy. We only use these methods if no individual country data are available and have to accept the uncertainty to fill gaps in data. See also Sect. 8 below.

The scaling of one source to another also increases the uncertainties associated with the final time series compared to the individual time series. The uncertainty of the final time series due to scaling can be calculated using standard error propagation formulas. For a scaling f , the standard deviation of a scaled time series $C = f \cdot B$ would be $s_c = \sqrt{s_f^2 + s_B^2}$, where s_B is the standard deviation of the time series B and s_f

²²International shipping and aviation emissions are also added, but they are not included in this study.

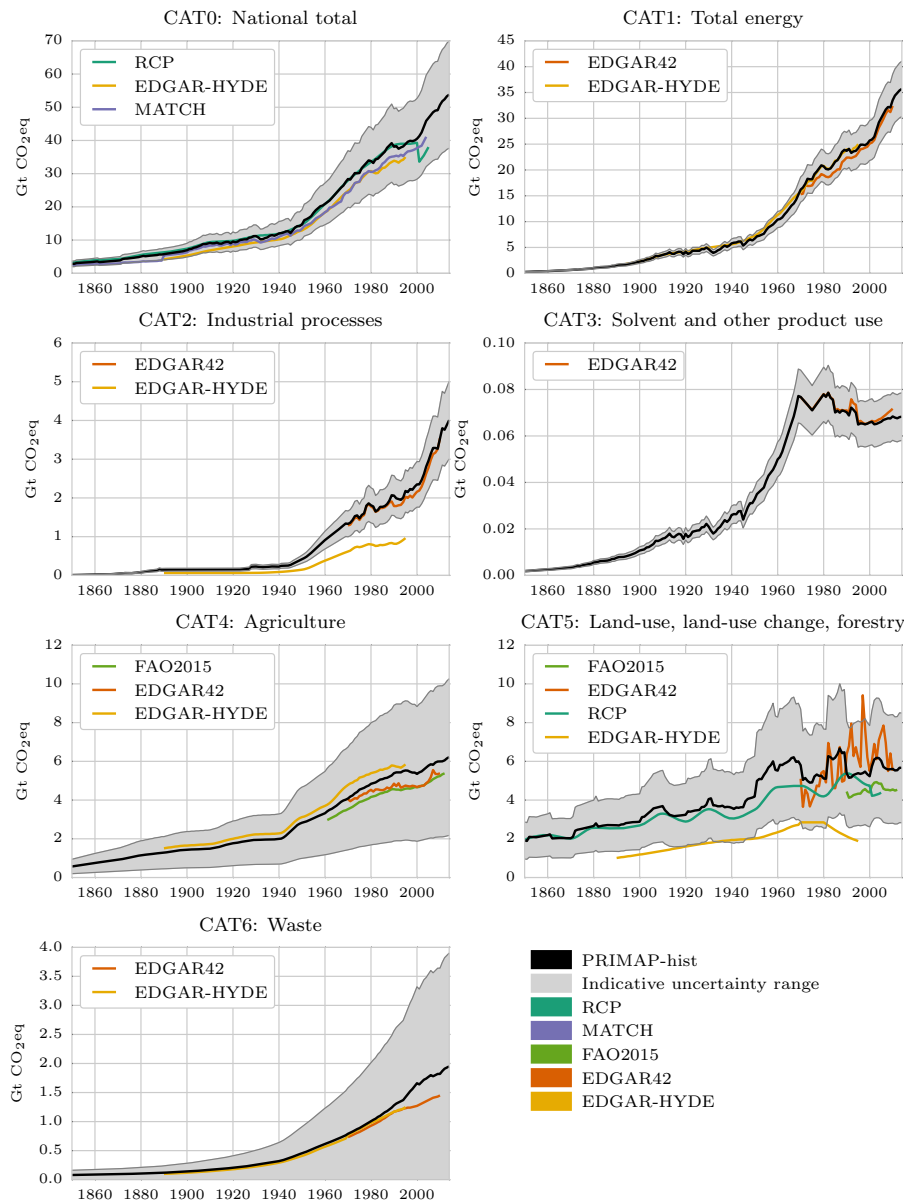


Figure 7. Comparison of the PRIMAP-hist dataset with both individual sources and composite datasets for aggregate Kyoto gases and the main IPCC 1996 categories. Grey shaded areas show the indicative uncertainty range from Table 10 applied to the PRIMAP-hist dataset. Where different uncertainty estimates exist, the average value is used. International shipping and aviation emissions are not included. The figure is discussed in Sect. 7.2.

is the standard deviation of the scaling factor, which depends on s_B and s_A in a manner determined by the exact matching algorithm (A denotes the time series which B is adjusted to).

8 Limitations of the method and use of the dataset

When combining time series from different data sources, one has to be careful because of the differences in methodology, assumptions, and data underlying the individual sources. The composite source generator of the PRIMAP emissions

module was built for this purpose and addresses those problems but some fundamental uncertainties and limitations of the method itself remain. In the following, we explain the sources of data discrepancies and the rationale behind our approach to the generation of a composite source as well as its limitations.

We begin with key sources for uncertainties and differences between datasets.

- Different methodologies for estimating emissions: some datasets are based on end-of-pipe measures, while oth-

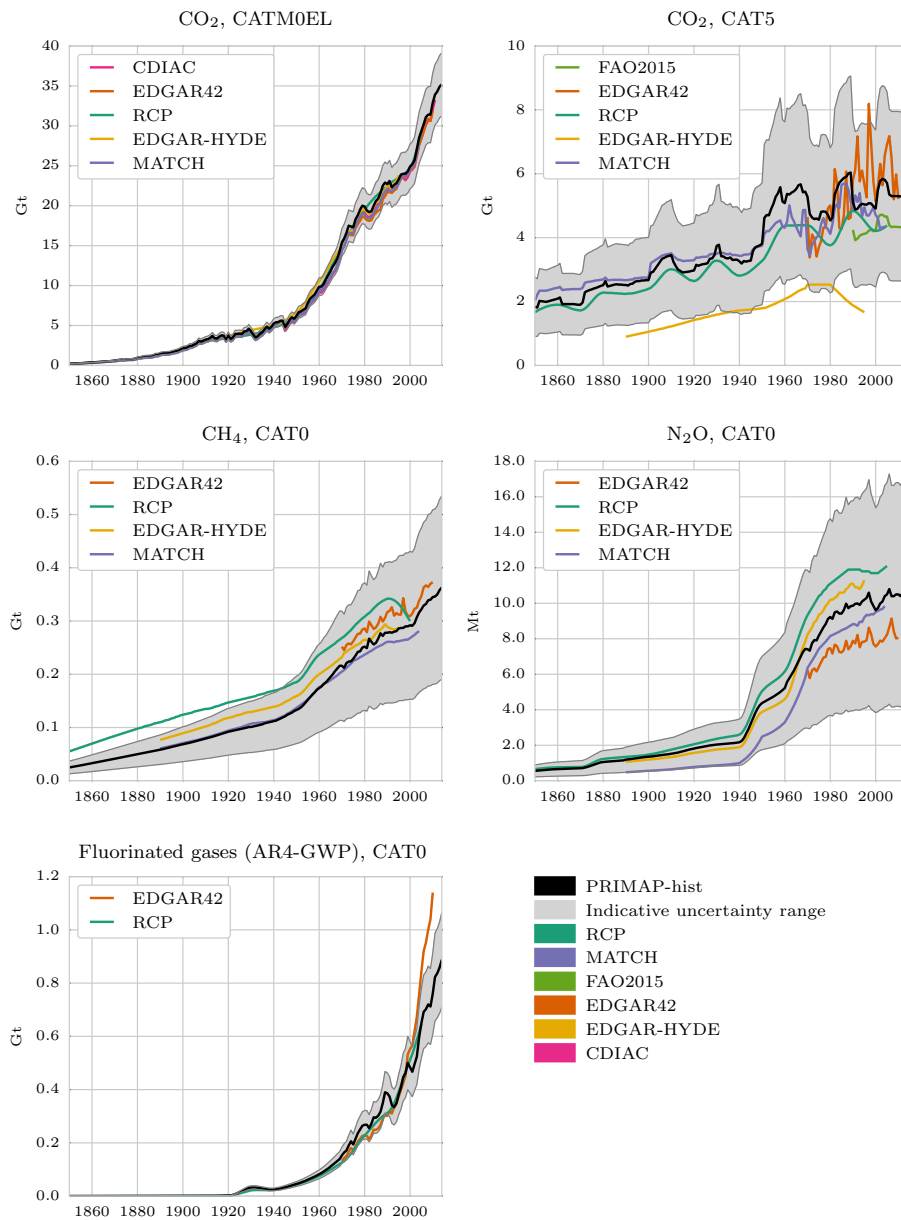


Figure 8. Comparison of the PRIMAP-hist dataset with both individual sources and composite datasets for different gases. Grey shaded areas show the indicative uncertainty range from Table 10 applied to the PRIMAP-hist dataset. Where different uncertainty estimates exist, the average value is used. International shipping and aviation emissions are not included. The figure is discussed in Sect. 7.2.

ers are based on economic activity data and assumed emission factors. Global emissions datasets can also be based on inverse emissions estimates from atmospheric concentration measurements.

- Different underlying data: two datasets using the same methodology would have different results when, for example, the data for the electricity production of individual power plants differ. Similarly, the data on the exact fuel type used and the emission factors used influence the resulting emissions.
- Differences in the detailed definitions of sectors: there are different ways to categorize emissions by economic sectors and not all data sources use the same categories. Categories from different sources can differ in their exact content despite having broadly the same definition.
- Different assumptions made for variables without data: the uncertainties are especially high for countries without a strong statistical record and sectors and gases, which need several assumptions for the calculation of emissions. Power sector CO₂ emissions have relatively

low uncertainty if a good record for power plant technology, the fuels used, and their electricity production exists. Agricultural emissions, on the other hand, have a high uncertainty as the emissions are based on natural processes, which depend on locally and seasonally fluctuating variables like soil moisture (see, e.g., Luo et al., 2013). See also Fig. 7.

An overview of the relative uncertainties for the different sources, countries, gases, and sectors is presented in Sect. 7.

To create a composite dataset we first prioritize the different data sources according to our judgment of their reliability and completeness. More complete sources at the top levels in the hierarchy will create a more consistent dataset than sources that cover only a few sectors or gases. However, if the top-level sources are unreliable, the resulting dataset will be unreliable, and it is beneficial to prioritize more reliable but less complete sources. Completeness has different dimensions, which we can often not optimize at the same time. Some datasets are very extensive in time and country coverage but only cover a few gases and sectors (e.g., CDIAC), while other sources cover only a fraction of the countries and years but with almost perfect sectoral and gas resolution (e.g., CRF, UNFCCC, BUR).

The first priority source is used as an anchor point for the other sources, which are used to extend the time series and to fill gaps. There are different options for the harmonization needed when extending one source with data from another source. We present some options below; a more detailed discussion is available in Rogelj et al. (2011):

1. No scaling: this does not alter data, but it also does not use information from the first priority source to improve data from the lower-priority sources.
2. Full scaling: here we scale the lower-priority sources such that they match the higher-priority sources at the borders. Effectively, we are using the growth rates of the lower-priority sources to extend the higher-priority source. If, for example, an emission factor is different for the two sources leading to a large difference in absolute emissions, the growth rates would still be the same and the extension with scaling would effectively use the emissions factor of the first source also for the second source. Of course, not all differences come from multiplicative errors like different emission factors. There could also be offsets from missing subsectors or incomplete data on individual emitters, which would not be corrected by using growth rates (unless one assumes the same growth rates for the missing subsectors as for the existing sectors).
3. Shifting using an offset: the lower-priority time series is harmonized by shifting the complete time series by a constant. This method implicitly assumes a constant error over time, which is not realistic if the emissions

time series is not constant. For extrapolation to the past it will likely overestimate emissions, while it will likely underestimate emissions for extrapolations to the future (assuming rising emissions).

We use a combination of approaches 1 and 2. We use scaling but limit the scaling to a factor of 1.5 to avoid introducing additional errors in the case of extremely different emissions data.

When combining the different sources, we cannot take into account all of their methodological differences. Often the exact assumptions and underlying data are not published with the datasets and an assessment of the uncertainty of the individual datasets is difficult because useful analysis is scarce (see also previous section). Thus, sometimes a time series using a slightly different sector definition is used to extend another time series. This introduces inconsistencies into the final dataset.

In Sect. 7 we presented uncertainties of the individual sources, sectors, and gases, which can reach over 100 % for some gases and sectors. We have to keep that in mind when designing and judging our methods. A very fine tuned and subsector-resolved method for the combination of datasets is still bound to the limitations of the input data and their uncertainties. While it is always possible to improve methods to reduce their uncertainty, it is not always sensible to invest more time if the major source of uncertainty is the input data and not the processing. Before adding further detail to future versions of the PRIMAP-hist dataset, it has to be assessed whether it adds real value to the data or whether the effects are overshadowed by uncertainties in the input data.

When using emissions data, one has to respect the uncertainties and limitations of the data. When making a statement about emissions intensities in different countries, the differences have to be seen in relation to the uncertainties before deducing anything from the calculated values. Individual country uncertainties can be much higher than the global uncertainties presented in Table 10. One of the purposes of this dataset is the calculation of countries contributions to climate change. Again we have to keep uncertainties in mind. This dataset can be used to study general effects, such as the impact of pre-1950 emissions on 2100 warming, but not the exact emissions targets for all countries according to a given equity principle (unless one accepts and communicates the uncertainties of the resulting emissions targets).

The land use downscaling methodology could be improved by a more detailed treatment of the different plant function types and the inclusion of savannas. Furthermore, the HYDE data do not account for deforestation for firewood, which influences the estimates of deforested areas, and the SAGE potential vegetation dataset also removes the human influence on the climate from the simulation. Climate is influenced globally, and thus some of the discrepancy between

potential and actual vegetation is caused by global climate change and not by local deforestation.²³

Finally, we have to note that the last years are obtained using extrapolations for most sectors and gases. Therefore, these data cannot be used to make statements about short-term emissions trends. We provide a version of this dataset that does not use numerical extrapolation to the future that can be used for this purpose. Where regional data are used for extrapolation to the past, individual country developments are not taken into account and cannot be deduced from the data. Short-term trends can also be influenced by the combination of different sources; thus, the consultation of original sources is advised before making statements about such trends.

This dataset is a combination of data from several models, measurements, and assumptions, including their shortcomings and uncertainties. It combines models and assumptions with new simplifications and uncertainties. However, it gives a more complete picture of the history of countries' greenhouse gas emissions than any of the individual sources can. From this perspective, our aggregate dataset is very useful.

²³Other causes of deforestation are also global (e.g., through demand for agricultural products) but under the UNFCCC emissions are attributed to the state they originate from. Neither where the products are consumed nor where the profits are made is considered.

Appendix A: Details of methodology used

In this section we explain technical details of the methodology used to create this dataset.

A1 Preprocessing

We use the same methods of preprocessing for all sources, though not all steps are used for all sources. Source-specific information is provided in Appendix B.

A1.1 Zero data and implausible data

We remove all time series that contain only zero values to ensure that zero values in higher-priority sources do not prevent the use of nonzero data from lower-priority sources. If negative data occur in time series that physically have to be positive, we replace the negative data by zero.

A1.2 Gas and category aggregation

Where necessary, we aggregate gases to gas baskets (e.g., individual HFCs to the HFC basket). If data are available at a more detailed sectoral level, we aggregate the categories to obtain time series at the sectoral resolution needed for the PRIMAP-hist dataset. In the process of aggregation we fill gaps in individual time series and extrapolate individual time series such that all gases or subsectors cover the same time period. Details of the extrapolation methods are discussed in Appendix A5.2 below. The same aggregation routine is also used in postprocessing to aggregate higher categories and the Kyoto GHG basket.

A2 Accounting for territorial changes

Where necessary, countries are summed or split to match our territorial definitions. Where only aggregate information is available, we use downscaling to obtain country-level information. In the case that we have to downscale emissions of formerly existing larger countries to the current individual countries, we downscale the larger countries' emissions using constant shares defined by the average of the first 5 years with data for the individual countries. This is used, for example, for countries of the former USSR. If no data for individual countries are available, we use an external downscaling key, e.g., emissions from a different source or GDP. When countries merge we sum the individual countries' time series. This is used for Germany, for example.

A3 Downscaling

We downscale regional data using country shares calculated from a different source, the key. Before downscaling, the key is preprocessed such that time series for all countries present cover the whole period to be downscaled. Extrapolation of

country pathways is done using the growth rates of all countries present in the region. This implies that the shares in regional emissions of countries with missing data stay constant from the last year with data (both for extrapolation to the future and to the past). If no data are present for any country in a region for a certain year it is extrapolated using constant emissions implying constant shares for the downscaling. Once the key time series is complete, the downscaling itself is done by multiplication of the country shares with the regional data.

A4 Combination of sources

The main features and functionality of the composite source generator (CSG) are described in Sect. 4. Here we add the missing details. To calculate the harmonization factor to be used for a lower-priority source, we use the linear trend of the last 6 years of the higher-priority source to calculate a year $n + 1$ value (or $n - 1$ when extending a time series to the past). The lower-priority source is then scaled such that it matches the extrapolated value in the given year. The scaling is confined to the interval $[0.67, 1.5]$ to avoid introducing large changes in emissions time series due to scaling.

In the case of land use emissions, we do not use scaling but rather fill gaps with unchanged data from lower-priority sources. The high fluctuations of land use data including different signs for data from different sources for the same year introduce high uncertainty in the scaling and render it meaningless in some cases, e.g., when one dataset shows removals while the other shows emissions for the period of overlap.

A5 Extrapolation

A5.1 Extrapolation with regional growth rates

For each region in the extrapolation source we loop over all countries contained in the region. We identify whether there are years within the given span where the extrapolation source contains data that could extend the country data. If this is the case, we compute the value for the last year without data for the country (the matching year) given by a linear trend. We compute the trend from opposite sides – i.e., for extrapolation to the past from 1850 to 1890, we compute the 1890 value of the country data from a linear trend through 1891 to 1905 and the 1890 value for the regional data from a linear trend through 1876 to 1890. The regional time series is then scaled such that they are identical in the matching year, and we extend the country data with the resulting time series. Unless stated otherwise we use 15-year trends.

A5.2 Numerical extrapolation

In this paper we use numerical extrapolation for extension of time series to the past on the scale of decades where historical data are not available, e.g., for land use N_2O and CH_4

emissions. It is also used before the gas and category aggregation process to extrapolate those time series for individual countries, gases, and categories which do not have data for the latest years to 2014.

Our framework for numerical extrapolation consists of different methods for extrapolation and a wrapper that controls the results and uses a fallback option if necessary. The following options are available:

Constant Data are extrapolated with a constant value, which is computed as the mean of the n last values before the extrapolation. Constant extrapolation has no fallback option.

Linear A linear trend is computed from the last n years before extrapolation. This trend is continued for the period of extrapolation. To control the extrapolated pathway, a check is made to see whether it crosses zero (negative emissions are currently impossible for most gases and sectors and have to be excluded). If crossing is not allowed, the fallback option for this case is used. The default option is to replace all values after the crossing point by zero. If emissions are extrapolated to the past and a trend is computed which has higher emissions in the past, a fallback option is triggered as well. The default is linear to zero extrapolation.

Linear to zero A linear pathway is constructed from a starting value to zero in the last year of the extrapolation. The starting value is computed from the linear trend of the last n values. If the calculated value is below zero despite all n values being positive, we use the last value instead of the value calculated from the linear trend. There is no fallback option.

Exponential The last n years are used to fit an exponential function, which extrapolates the data. The exponential function is of the form $E(y) = a \cdot e^{by}$, where a and b are determined by the fit. A fallback option is used if exponential fitting is not possible (e.g., when the n years contain positive as well as negative values), if too few of the n years have data available, or if during extrapolation to the past we obtain a negative exponent (i.e., emissions in the past higher than in the future). The default fallback option is linear to zero.

Options for all methods are the number n of years to use for the fit (default 15) and the minimal number of these years that have to contain data (default 8). Fitting can be controlled independently for extrapolations to the past and the future.

Appendix B: Details on data source preprocessing

Here we briefly describe the preprocessing steps carried out for each of the sources used. We only describe the steps for the time series needed for this paper. Aggregation of additional sectors, gas baskets, and regional groups is omitted

because, for the PRIMAP-hist dataset, it is done using the final time series.

BP2015 BP resolves only some states, while other states are summed into five regional groups. We downscale these groups using shares of CDIAC2015 CAT1A emissions. After downscaling, countries are summed to the territorial definitions used in this paper.

BUR2015 We remove all time series which contain less than three data points or cover less than 11 years. We build the HFC and PFC baskets for both SAR and AR4 global warming potentials using the gas and category aggregation functionality of the emissions module (Appendix A1.2). Category aggregation is not necessary as we directly read the data into the PRIMAP emissions database in the needed categorical detail.

CDIAC2015 CDIAC makes country unification and splitting explicit by issuing different country codes. We sum and downscale countries where needed to obtain current countries and territories for all years. Where downscaling is needed, we use the first 5 years with data for the individual countries as a downscaling key and downscale with constant shares. Where no data for the individual countries are available, we use CRF2014 data for the same category as downscaling key. This affects downscaling of France and Monaco as well as Switzerland and Liechtenstein. Where CRF data are not available (Italy and San Marino), we use the GDP data from the World Bank (2015) as the downscaling key. Finally, we sum countries to the territorial definitions used in this paper.

The emissions categories covered are fossil fuel burning, which corresponds to IPCC category 1A; gas flaring, which corresponds to IPCC 1996 category 1B2C22, which we use as a proxy for category 1B2; and cement production, which corresponds to IPCC 1996 category 2A1, which we use as a proxy for category 2A.

CRF2014 and CRF2013 CRF data only need minimal preprocessing. We build the HFC and PFC baskets for both SAR and AR4 global warming potentials using the gas and category aggregation functionality of the emissions module (Appendix A1.2). Actual emissions are used for the PRIMAP-hist dataset (in contrast to potential emissions also available from CRF data).

EDGAR42 First, EDGAR v4.2 and EDGAR v4.2 FT 2010 are independently aggregated to the categorical resolution needed. We retain any existing aggregate time series, as in some cases (at least in EDGAR v4.2 FT2010) not all subsectors are present as individual time series and re-aggregation would lose emissions from the sectors not available individually. Then the two sources are combined using the composite source generator with

EDGAR 4.2 FT2010 as the first priority source. The harmonization in the CSG does not use linear trends here. Subsequently, HFC and PFC gas baskets are aggregated including extrapolation of individual gases such that all gases of a basket cover the same time span. Finally, we calculate emissions for some small countries where emissions are included in time series of larger countries. In detail, these are downscaling of Serbia and Montenegro as a region to individual countries, downscaling of Monaco from France, downscaling of Liechtenstein from Switzerland, and downscaling of the Vatican City and San Marino from Italy. The downscaling key used is population data from the UN Population Division (2015).

EDGAR-HYDE14 EDGAR-HYDE data uses the EDGAR v2.0 categorization, which differs from the IPCC 1996 categorization used here. The IPCC 1996 categories we identify with the EDGAR42 categories are shown in Table B1.

The summation of subcategories is done using the emissions module's aggregation framework. We do not use international bunker fuel emissions (EXX) as we do not include bunker fuels in this analysis. Data are interpolated using Matlab's "pchip" function.

FAO2015 Like CDIAC, FAO data explicitly model division and unification of countries. Our first step is to sum and split these countries to obtain time series for the current countries and the territorial definitions used here (see Sect. 4.3). FAO uses different subcategories for agriculture and land use than IPCC 1996, which need to be translated to IPCC 1996 categories. For this paper the details are not relevant as we operate on aggregate agricultural and land use data.

HOUGHTON2008 The downscaling is described in Sect. 4.2.2. Here we add some further details. The downscaling uses regional shares in cumulative deforested areas to split the regional emissions pathway to countries. In some regions there are countries with both net deforestation and net afforestation, so some countries have negative shares, which cannot be used directly for downscaling in a meaningful way. Instead, we first calculate shares from only deforestation and multiply those by the regional pathway to obtain preliminary emissions pathways. These pathways are then shifted such that the cumulative net emissions (or removals) equal the cumulative net emissions (or removals) calculated directly from the net deforestation shares. This approach avoids inverted growth rates for countries with net afforestation in a region with net deforestation.

Countries missing in the Houghton source are added using the regional growth rates and shares computed by

Table B1. Category matching for EDGAR-HYDE and IPCC 1996 categories.

EDGAR-HYDE	IPCC1996
FNN	CAT1A
FPP	CAT1B
I00	CAT2
LGG + LNN + L42 + L43 + L70	CAT4
L41	CAT5
W10 + WNN	CAT6

the relative deforestation compared to a Houghton region with similar climate.

HYDE No preprocessing is needed.

RCP Data are first interpolated using MATLAB's "pchip" function. For CH₄ we aggregate time series to the necessary regional level. HFC and PFC baskets are created. For CH₄ from categories 1, 2, and 4, the years 1860–1880 are removed before interpolation. They show a steep decline to almost zero emissions from 1850 to 1860, which rise again to much higher values in 1890. This cannot be observed in the data presented in Lamarque et al. (2010), which is the original source of the data according to the RCP website (Meinshausen, 2011). We judge this to be an error and thus replace the values by interpolation. RCP data are published in IPCC 1996 categories and thus no mapping is needed.

SAGE No preprocessing is needed.

UNFCCC2015 See BUR2015.

Appendix C: Data sources not used

In this section we describe data sources that were considered but not used in the final composite source and give the reasons why the data were not used.

C1 Biennial reports

Biennial reports are submitted to the UNFCCC by Annex I parties. The UNFCCC biennial reporting guidelines for developed-country parties (Decision 2.CP17, Annex I) state that "the information provided in the biennial report should be consistent with that provided in the most recent annual inventory submission, and any differences should be fully explained". It is therefore safe to assume that data submitted with the biennial reports are consistent with CRF data (Sect. 2.2.3).

C2 National communications by developed countries

National communications by developed-country parties UNFCCC (2014b) serve the purpose of giving information on the commitments that parties are undertaking to limit their greenhouse gas emissions and the policies implemented and planned to reach the commitments. They contain some greenhouse gas data, but historical data do not add to CRF data, so national communications by developed-country parties are not used here.

C3 CAIT 2.0

The Climate Analysis Indicators Tool (CAIT) dataset is published by the World Resources Institute (WRI) (World Resources Institute, 2016). It contains data for several countries until 2011 (some countries have less coverage). Emissions time series are available either on an aggregate Kyoto GHG level or with details for either sectors or gases. Unfortunately, there are no data with details for sector and gas at the same time. For fluorinated gases, only aggregate data are available without the details on HFCs, PFCs, and SF₆ needed for this project.

Similar to our work, CAIT 2.0 emissions time series are assembled from different sources. Data from the International Energy Agency (IEA) (see Appendix C6), the US Energy Information Administration (EIA) (see Appendix C5), and CDIAC (see Sect. 2.3.2) are used for fossil CO₂. Non-CO₂ emissions are taken from the US EPA source (see Appendix C7). LULUCF data are taken from FAO (see Sect. 2.3.4).

All sources are either included in our dataset individually (CDIAC, FAOSTAT), not publicly accessible (IEA), or only contain emissions already covered from other sources (EIA, US EPA). We do not use CAIT data, as the results are more transparent when using the original data sources directly.

C4 CDIAC CH₄

This dataset has been described in Stern and Kaufmann (1995, 1996, 1998) and covers global CH₄ emissions for a period from 1860 to 1994. It is created using correlations of methane emissions to socioeconomic variables or emissions of other gases for which time series are available. It is tested against emissions estimates from measurements of atmospheric methane concentrations. Due to its lack of country or regional data, it could only be used for extrapolation. However, we have RCP data that covers the same period and sectoral detail but has a regional resolution. We therefore do not use the CDIAC CH₄ data.

C5 EIA energy CO₂

The US Energy Information Administration's (EIA) publishes CO₂ emissions from energy consumption for most of the world countries. The period from 1980 to 2012 is covered. The covered sectors are consumption of coal, petroleum, and natural gas (together these correspond to IPCC 1996 category 1A) and flaring of natural gas (IPCC 1996 category 1B2C22).

We do not use the dataset because the sectors and time frame are covered by CDIAC2015.

C6 IEA energy CO₂

The International Energy Agency offers CO₂ emissions from fuel combustion for purchase. The dataset covers 34 OECD countries and 100 non-OECD countries. As it is not publicly available, we do not include it in our dataset.

C7 US EPA

The United States Environmental Protection Agency (EPA) published data for non-CO₂ emissions (US Environmental Protection Agency, 2012). The dataset covers many countries and the years 1990 to 2005. It is a composite of different data sources where publicly available country-prepared reports are prioritized. A main source for the historical data is the UNFCCC flexible query system. Annex I countries therefore use CRF data, while non-Annex I countries use data from the national communications and national inventory reports. However, each time series has only a few data points. We already include the individual sources used in this dataset and only a small amount of information is added. Thus, we do not use the US EPA data.

Author contributions. All authors contributed to checking the results and writing the manuscript. Johannes Gütschow conceptualized the study, programmed the necessary addition to the PRIMAP emissions module, created the composite source, and prepared some of the input datasets. M. Louise Jeffery conceptualized and carried out the calculations to obtain deforestation estimates from the SAGE and HYDE datasets. Robert Gieseke created the accompanying website and most of the figures in this paper, in addition to preparing data. Ronja Gebel prepared input data. David Stevens collected and analyzed the uncertainty data.

Acknowledgements. The authors acknowledge and appreciate funding by the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (11_II_093_Global_A_SIDS_and_LDCs) and the Economic Commission for Latin America and the Caribbean (project “Development of a reference methodology on historical responsibility for CO₂ emissions”).

Edited by: D. Carlson

Reviewed by: two anonymous referees

References

- Andres, R. J., Marland, G., Fung, I., and Matthews, E.: A $1^\circ \times 1^\circ$ distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990, *Global Biogeochem. Cy.*, 10, 419–429, doi:10.1029/96GB01523, 1996.
- Andres, R. J., Fielding, D. J., Marland, G., Boden, T. A., Kumar, N., and Kearney, A. T.: Carbon dioxide emissions from fossil-fuel use, 1751–1950, *Tellus B*, 51, 759–765, doi:10.1034/j.1600-0889.1999.t01-3-00002.x, 1999.
- Andres, R. J., Boden, T. A., and Higdon, D.: A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission, *Tellus B*, 66, 1–15, doi:10.3402/tellusb.v66.23616, 2014.
- Baer, P., Fieldman, G., Athanasiou, T., and Kartha, S.: Greenhouse Development Rights: towards an equitable framework for global climate policy, *Camb. Rev. Int. Aff.*, 21, 649–669, doi:10.1080/09557570802453050, 2008.
- Bode, S.: Equal emissions per capita over time – A proposal to combine responsibility and equity of rights for post-2012 GHG emission entitlement allocation, *European Environment*, 14, 300–316, doi:10.1002/eet.359, 2004.
- Boden, T. A., Marland, G., and Andres, R. J.: Global, Regional, and National Fossil-Fuel CO₂ Emissions, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., USA, doi:10.3334/CDIAC/00001_V2015, 2015.
- British Petroleum: BP Statistical Review of World Energy 2015, Tech. Rep. June, available at: <http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html> (last access: 15 July 2015), 2015.
- Davidson, E. A.: The contribution of manure and fertilizer nitrogen to atmospheric nitrous oxide since 1860, *Nat. Geosci.*, 2, 659–662, doi:10.1038/ngeo608, 2009.
- den Elzen, M. G. J., Olivier, J. G. J., Höhne, N., and Janssens-Maenhout, G.: Countries’ contributions to climate change: effect of accounting for all greenhouse gases, recent trends, basic needs and technological progress, *Climatic Change*, 121, 397–412, doi:10.1007/s10584-013-0865-6, 2013.
- Federici, S., Tubiello, F. N., Salvatore, M., Jacobs, H., and Schmidhuber, J.: New estimates of CO₂ forest emissions and removals: 1990–2015, *Forest Ecol. Manage.*, 352, 89–98, doi:10.1016/j.foreco.2015.04.022, 2015.
- Federico, G.: Feeding the World: An Economic History of Agriculture, 1800–2000, Princeton University Press, Princeton, available at: <http://press.princeton.edu/titles/8057.html> (last access: 3 November 2016), 2008.
- Food and Agriculture Organization of the United Nations: FAO-STAT database, available at: <http://faostat3.fao.org/home/E> (last access: 3 June 2015), 2015a.
- Food and Agriculture Organization of the United Nations: FAO-STAT Methodology, available at: http://faostat3.fao.org/mes/methodology_list/E (last access: 8 November 2016), 2015b.
- Food and Agriculture Organization of the United Nations: FAO GEONETWORK. Country boundaries of the world (vector version) (FGGD) (GeoLayer), latest update: 4 June 2015, available at: <http://ref.data.fao.org/map?entryId=18329470-472d-11db-88e0-000d939bc5d8> (last access: 13 July 2015), 2015c.
- Food and Agriculture Organization of the United Nations: FAO-STAT cropland metadata, available at: http://faostat3.fao.org/download/GC/*E (last access: 9 September 2016), 2016.
- Goldeewijk, C. K. and Battjes, J.: A hundred year (1890–1990) database for integrated environmental assessments (HYDE, version 1.1), Tech. rep., Bilthoven, available at: http://www.rivm.nl/en/Documents_and_publications/Scientific/Reports/1997/februari/A_Hundred_Year_1890_1990_Database_for_Integrated_Environmental_Assessments_HYDE_version_1_1 (last access: 3 November 2016), 1997.
- Gütschow, J., Jeffery, M. L., Gieseke, R., Gebel, R., Stevens, D., Krapp, M., and Rocha, M.: The PRIMAP-hist national historical emissions time series (1850–2014), GFZ Data Services, doi:10.5880/PIK.2016.003, 2016.
- Höhne, N., Blum, H., Fuglestvedt, J., Skeie, R. B., Kurosawa, A., Hu, G., Lowe, J., Gohar, L., Matthews, B., Nioac de Salles, A. C., and Ellermann, C.: Contributions of individual countries’ emissions to climate change and their uncertainty, *Climatic Change*, 106, 359–391, doi:10.1007/s10584-010-9930-6, 2010.
- Houghton, R. A.: The annual net flux of carbon to the atmosphere from changes in land use 1850–1990, *Tellus B*, 51, 298–313, doi:10.1034/j.1600-0889.1999.00013.x, 1999.
- Houghton, R. A.: Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000, *Tellus B*, 55, 378–390, doi:10.1034/j.1600-0889.2003.01450.x, 2003.
- Houghton, R. A.: Carbon Flux to the Atmosphere from Land-Use Changes: 1850–2005, in: *TRENDS: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., USA, available at: <http://cdiac.ornl.gov/trends/landuse/houghton/houghton.html> (last access: 3 November 2016), 2008.

- IPCC: Climate Change 2014: Mitigation of Climate Change: Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on, Cambridge University Press, Cambridge, UK and New York, NY, USA, 2014.
- JRC and PBL: Emission Database for Global Atmospheric Research (EDGAR), release version 4.2, doi:10.2904/EDGARv4.2, available at: <http://edgar.jrc.ec.europa.eu/overview.php?v=42> (last access: 3 November 2016), 2011.
- JRC and PBL: Emission Database for Global Atmospheric Research (EDGAR), release version 4.2 FT2010, available at: <http://edgar.jrc.ec.europa.eu/overview.php?v=42FT2012> (last access: 3 November 2016), 2013.
- JRC and PBL: Emission Database for Global Atmospheric Research (EDGAR), release version 4.2 FT2012, available at: <http://edgar.jrc.ec.europa.eu/overview.php?v=42FT2012> (last access: 3 November 2016), 2014.
- Klein Goldewijk, K., Beusen, A., and Janssen, P.: Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1, *Holocene*, 20, 565–573, doi:10.1177/0959683609356587, 2010.
- Klein Goldewijk, K., Beusen, A., Van Drecht, G., and De Vos, M.: The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years, *Global Ecol. Biogeogr.*, 20, 73–86, doi:10.1111/j.1466-8238.2010.00587.x, 2011.
- Korsbakken, J. I., Peters, G. P., and Andrew, R. M.: Uncertainties around reductions in China's coal use and CO₂ emissions, *Nature Climate Change*, 6, 687–690, doi:10.1038/nclimate2963, 2016.
- Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, *Atmos. Chem. Phys.*, 10, 7017–7039, doi:10.5194/acp-10-7017-2010, 2010.
- Le Quéré, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Friedlingstein, P., Peters, G. P., Andres, R. J., Boden, T. A., Houghton, R. A., House, J. I., Keeling, R. F., Tans, P., Arneeth, A., Bakker, D. C. E., Barbero, L., Bopp, L., Chang, J., Chevallier, F., Chini, L. P., Ciais, P., Fader, M., Feely, R. A., Gkritzalis, T., Harris, I., Hauck, J., Ilyina, T., Jain, A. K., Kato, E., Kitidis, V., Klein Goldewijk, K., Koven, C., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lima, I. D., Metzl, N., Millero, F., Munro, D. R., Murata, A., Nabel, J. E. M. S., Nakaoka, S., Nojiri, Y., O'Brien, K., Olsen, A., Ono, T., Pérez, F. F., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Rödenbeck, C., Saito, S., Schuster, U., Schwinger, J., Séférian, R., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Vandemark, D., Viovy, N., Wiltshire, A., Zaehle, S., and Zeng, N.: Global Carbon Budget 2015, *Earth Syst. Sci. Data*, 7, 349–396, doi:10.5194/essd-7-349-2015, 2015.
- Levin, I., Naegler, T., Heinz, R., Osusko, D., Cuevas, E., Engel, A., Ilmberger, J., Langenfelds, R. L., Neining, B., Rohden, C. v., Steele, L. P., Weller, R., Worthy, D. E., and Zimov, S. A.: The global SF₆ source inferred from long-term high precision atmospheric measurements and its comparison with emission inventories, *Atmos. Chem. Phys.*, 10, 2655–2662, doi:10.5194/acp-10-2655-2010, 2010.
- Liu, Y. Y., van Dijk, A. I. J. M., de Jeu, R. A. M., Canadell, J. G., McCabe, M. F., Evans, J. P., and Wang, G.: Recent reversal in loss of global terrestrial biomass, *Nature Climate Change*, 5, 1–5, doi:10.1038/nclimate2581, 2015.
- Luo, G. J., Kiese, R., Wolf, B., and Butterbach-Bahl, K.: Effects of soil temperature and moisture on methane uptake and nitrous oxide emissions across three different ecosystem types, *Biogeochemistry*, 10, 3205–3219, doi:10.5194/bg-10-3205-2013, 2013.
- Marland, G. and Rotty, R. M.: Carbon dioxide emissions from fossil fuels: a procedure for estimation and results for 1950–1982, *Tellus B*, 36, 232–261, doi:10.1111/j.1600-0889.1984.tb00245.x, 1984.
- Mathews, H. D., Graham, T. L., Keverian, S., Lamontagne, C., Seto, D., and Smith, T. J.: National contributions to observed global warming, *Environmental Research Letters*, 9, 014010, doi:10.1088/1748-9326/9/1/014010, 2014.
- Meinshausen, M.: RCP Concentration Calculations and Data Final Version, background data, acknowledgements and further info, available at: <http://www.pik-potsdam.de/~mmalte/rcps/index.htm> (last access: 15 March 2016), 2011.
- Meinshausen, M. and Alexander, R.: INDC Factsheets, available at: <http://climatecollege.unimelb.edu.au/indc-factsheets> (last access: 9 September 2016), 2016.
- Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, *MAGICC6 – Part 1: Model description and calibration*, *Atmos. Chem. Phys.*, 11, 1417–1456, doi:10.5194/acp-11-1417-2011, 2011a.
- Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and Vuuren, D. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, *Climatic Change*, 109, 213–241, doi:10.1007/s10584-011-0156-z, 2011b.
- Meinshausen, M., Jeffery, L., Guetschow, J., Robiou du Pont, Y., Rogelj, J., Schaeffer, M., Höhne, N., den Elzen, M., Oberthür, S., and Meinshausen, N.: National post-2020 greenhouse gas targets and diversity-aware leadership, *Nature Climate Change*, 5, 1098–1106, doi:10.1038/nclimate2826, 2015.
- Metz, B., Davidson, O., de Coninck, H., Loos, M., and Meyer, L.: Special Report on Safeguarding the Ozone Layer and the Global Climate System, Tech. rep., IPCC, Cambridge, available at: https://www.ipcc.ch/publications_and_data/_safeguarding_the_ozone_layer.htm (last access: 3 November 2016), 2007.
- Miguez, J. D. G. and Filho, L. G. M.: Note on the time-dependent relationship between emissions of greenhouse gases and climate change, Tech. rep., Ministry of Science and Technology, Federative Republic of Brazil, available at: <http://unfccc.int/resource/brazil/documents/proposta.pdf> (last access: 3 November 2016), 2000.
- Nabel, J. E., Rogelj, J., Chen, C. M., Markmann, K., Gutzmann, D. J., and Meinshausen, M.: Decision support for international climate policy – The PRIMAP emission module, *Environ. Model. Softw.*, 26, 1419–1433, doi:10.1016/j.envsoft.2011.08.004, 2011.

- Olivier, J. and Berdowski, J.: Global Emissions sources and sinks, in: *The Climate System*, edited by: Berdowski, J., Guicherit, R., and Heij, B., 33–78, Balkema Publishers/Swets & Zeitlinger Publishers, the Netherlands, 2001.
- Olivier, J. and Janssens-Maenhout, G.: Part III: Greenhouse gas emissions, in: *CO₂ Emissions from Fuel Combustion – 2012 Edition*, *CO₂ Emissions from Fuel Combustion*, OECD Publishing, doi:10.1787/co2_fuel-2012-en, 2012.
- Olivier, J., Bouwman, A., Berdowski, J., Veldt, C., Bloos, J., Visschedijk, A., van der Maas, C., and Zandveld, P.: Sectoral emission inventories of greenhouse gases for 1990 on a per country basis as well as on $1^\circ \times 1^\circ$, *Environ. Sci. Pol.*, 2, 241–263, doi:10.1016/S1462-9011(99)00027-1, 1999.
- Olivier, J., Muntean, M., and Peters, J.: Trends in global CO₂ emissions: 2015 report, Tech. rep., PBL Netherlands Environmental Assessment Agency & European Commission's Joint Research Centre (JRC), available at: http://edgar.jrc.ec.europa.eu/news_docs/jrc-2015-trends-in-global-co2-emissions-2015-report-98184.pdf (last access: 3 November 2016), 2015.
- PBL: Emission Database for Global Atmospheric Research (EDGAR): Uncertainties, available at: <http://themasites.pbl.nl/tridion/en/themasites/edgar/documentation/uncertainties/index-2.html> (last access: 31 March 2016), 2010.
- PBL: HYDE 3.1, available at: ftp://ftp.pbl.nl/hyde/hyde3.1/hyde31_final/ (last access: 8 November 2016), 2015.
- Pongratz, J. and Caldeira, K.: Attribution of Atmospheric CO₂ and Temperature Increases to Regions: Importance of Preindustrial Land Use Change, *Environ. Res. Lett.*, 7, 034001, doi:10.1088/1748-9326/7/3/034001, 2012.
- Ramankutty, N. and Foley, J. A.: Estimating historical changes in global land cover: Croplands from 1700 to 1992, *Global Biogeochem. Cy.*, 13, 997–1027, doi:10.1029/1999GB900046, 1999.
- Ramankutty, N. and Foley, J. A.: SAGE Global Potential Vegetation Dataset, available at: <https://nelson.wisc.edu/sage/data-and-models/global-potential-vegetation/index.php> (last access: 9 February 2016), 2015.
- Rogelj, J., Hare, W., Chen, C., and Meinshausen, M.: Discrepancies in historical emissions point to a wider 2020 gap between 2 °C benchmarks and aggregated national mitigation pledges, *Environ. Res. Lett.*, 6, 024002, doi:10.1088/1748-9326/6/2/024002, 2011.
- Smith, P., Bustamante, M., Ahammad, H., Clark, H., Dong, H., El-siddig, E., Haberl, H., Harper, R., House, J., Jafari, M., Masera, O., Mbow, C., Ravindranath, N., Rice, C., Abad, C. R., Romanovskaya, A., Sperling, F., and Tubiello, F.: Agriculture, Forestry and Other Land Use (AFOLU), in: *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., and Minx, J., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2014.
- Stern, D. and Kaufmann, R.: Estimates of global anthropogenic methane emissions 1860–1993, Boston University, Center for energy and Environmental Studies, Working Paper Series, Number 9504, 1995.
- Stern, D. and Kaufmann, R.: Annual Estimates of Global Anthropogenic Methane Emissions: 1860–1994, in: *Trends Online: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., USA, doi:10.3334/CDIAC/tge.001, 1998.
- Stern, D. I. and Kaufmann, R. K.: Estimates of global anthropogenic methane emissions 1860–1993, *Chemosphere*, 33, 159–176, doi:10.1016/0045-6535(96)00157-9, 1996.
- Tubiello, F. N., Salvatore, M., Ferrara, A. F., House, J., Federici, S., Rossi, S., Biancalani, R., Condor Golec, R. D., Jacobs, H., Flammini, A., Proserpi, P., Cardenas-Galindo, P., Schmidhuber, J., Sanz Sanchez, M. J., Srivastava, N., and Smith, P.: The Contribution of Agriculture, Forestry and other Land Use activities to Global Warming, 1990–2012, *Glob. Change Biol.*, 21, 2655–2660, doi:10.1111/gcb.12865, 2015.
- UNEP: The Emissions Gap Report 2012, Tech. rep., United Nations Environment Programme, available at: <http://www.unep.org/pdf/2012gapreport.pdf> (last access: 3 November 2016), 2012.
- UNEP: The Emissions Gap Report 2014, UNEP, available at: <http://www.unep.org/publications/ebooks/emissionsgapreport2014/> (last access: 3 November 2016), 2014.
- UNEP: The Emissions Gap Report 2015, United Nations Environment Programme, available at: <http://web.unep.org/emissionsgapreport2015/> (last access: 3 November 2016), 2015.
- UNFCCC: National Inventory Submissions 2013, available at: http://unfccc.int/national_reports/annex_i_ghg_inventories/national_inventories_submissions/items/7383.php (last access: 3 November 2016), 2013.
- UNFCCC: National Inventory Submissions 2014, available at: http://unfccc.int/national_reports/annex_i_ghg_inventories/national_inventories_submissions/items/8108.php (last access: 3 November 2016), 2014a.
- UNFCCC: 6th National Communications to the UNFCCC by Annex I parties, available at: http://unfccc.int/national_reports/annex_i_natcom/submitted_natcom/items/7742.php (last access: 3 November 2016), 2014b.
- UNFCCC: National Communications to the UNFCCC by non-Annex I parties, available at: http://unfccc.int/national_reports/non-annex_i_natcom/submitted_natcom/items/653.php (last access: 3 November 2016), 2015a.
- UNFCCC: UNFCCC Decision 1/CP.21: Paris Agreement, Tech. rep., United Nations Framework Convention on Climate Change, available at: <http://unfccc.int/resource/docs/2015/cop21/eng/10a01.pdf> (last access: 3 November 2016), 2015b.
- UNFCCC: UNFCCC Detailed by Party data, available at: <http://unfccc.int/di/DetailedByParty.do> (last access: 3 November 2016), 2015c.
- UNFCCC: Submitted biennial update reports (BURs) from non-Annex I parties, available at: http://unfccc.int/national_reports/non-annex_i_natcom/reporting_on_climate_change/items/8722.php (last access: 3 November 2016), 2016.
- United Nations: 2013 Energy Statistics Yearbook, United Nations Department for Economic and Social Information and Policy Analysis, Statistics Division, New York, available at: <http://unstats.un.org/unsd/energy/yearbook/> (last access: 3 November 2016), 2016.

- UN Population Division: World Population Prospects 2015 revision, available at: <http://esa.un.org/unpd/wpp/> (last access: 3 November 2016), 2015.
- US Environmental Protection Agency: Global Anthropogenic Non-CO₂ Greenhouse Gas Emissions : 1990–2030, Tech. Rep. December, US EPA, 2012.
- Van Aardenne, J. A., Dentener, F. J., Olivier, J. G. J., Goldewijk, C. G. M. K., and Lelieveld, J.: A 1 × 1 deg resolution data set of historical anthropogenic trace gas emissions for the period 1890–1990, *Global Biogeochem. Cy.*, 15, 909–928, doi:10.1029/2000GB001265, 2001.
- Winkler, H., Jayaraman, T., Pan, J., de Oliveira, A. S., Zhang, Y., Sant, G., Miguez, J. D. G., Letete, T., Marquard, A., and Raubenheimer, S.: Equitable access to sustainable development A paper by experts from BASIC countries, Tech. rep., BASIC expert group, Beijing, Brasilia, Cape Town and Mumbai, 2011.
- World Bank: World Development Indicators 2015, doi:10.1596/978-0-8213-7386-6, 2015.
- World Resources Institute: CAIT Historical, available at: <http://cait.wri.org/historical> (last access: 7 October 2016), 2016.
- Yamin, F. and Depledge, J.: The International Climate Change Regime. A guide to Rules, Institutions and Procedures, Cambridge University Press, Cambridge, UK, available at: <http://www.cambridge.org/co/academic/subjects/law/environmental-law/international-climate-change-regime-guide-rules-institutions-and-procedures> (last access: 3 November 2016), 2005.