OECD Environment Working Papers No. 252

National greenhouse gas emission inventories, filling the gaps in official data

Miguel Cárdenas Rodríguez,

Emma Schwentner,

Edoardo Falchi,

Andrzej Suchodolski,

Rodrigo Pizarro

https://dx.doi.org/10.1787/f19f1663-en

OECD

Environment Working Paper No. 252

National greenhouse gas emission inventories: filling the gaps in official data

C82, E01, Q54, Q56

Authorised for publication by Jo Tyndall, Director, Environment Directorate.

Miguel Cárdenas Rodríguez (OECD), Emma Schwentner (OECD), Edoardo Falchi (OECD), Andrzej Suchodolski (OECD), Rodrigo Pizarro (OECD)

OECD ENVIRONMENT WORKING PAPERS

Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s). Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works.

This series is designed to make available to a wider readership selected studies on environmental issues prepared for use within the OECD. Authorship is usually collective, but principal author(s) are named. The papers are generally available only in their original language – English or French – with a summary in the other language. Comments on Working Papers are welcomed, and may be sent to:

OECD Environment Directorate
2 rue André-Pascal, 75775 Paris Cedex 16, France
or by email: *env.contact@oecd.org*

OECD Environment Working Papers are published on
www.oecd.org/environment/workingpapers.htm

Abstract

This paper addresses the need for a comprehensive global dataset on national greenhouse gas (GHG) emission inventories to support policy analysis and track progress towards climate change mitigation goals. While numerous datasets exist, gaps in official data, particularly from developing countries, hinder their utility. The paper develops a methodology to compile a complete dataset, prioritizing official data reported to the United Nations Framework Convention on Climate Change (UNFCCC) and the OECD. To fill gaps, estimates based on unofficial datasets, such as PRIMAP, will be used, as they demonstrate the greatest alignment with official data in terms of emission levels, trends, and categories. As official data become available under the Enhanced Transparency Framework of the Paris Agreement, estimated values would be replaced and additional data collection would not be necessary. The dataset disaggregates emissions by IPCC categories, gases, countries, and years, distinguishing between official data points and estimates. This harmonized dataset enhances transparency, improves data collection efficiencies, and enables more robust evidence-based policy analysis. Moving forward, the paper recommends refining the dataset with sector-specific data and fostering collaboration with international organizations to improve the reliability and consistency of global GHG emissions data and indicators.

Keywords: Greenhouse gas emissions, official data, climate change mitigation, emission sources, national inventories.

JEL Classification: C82, E01, Q54, Q56

Résumé

Ce document répond à la nécessité de disposer d'un ensemble complet de données mondiales sur les inventaires nationaux d'émissions de gaz à effet de serre (GES) afin de soutenir l'analyse des politiques et de suivre les progrès accomplis dans la réalisation des objectifs d'atténuation du changement climatique. Bien qu'il existe de nombreux ensembles de données, les valeurs manquantes dans les données officielles, en particulier dans les pays en développement, entravent leur utilité. Ce document développe une méthodologie pour compiler un ensemble complet de données, en donnant la priorité aux données officielles communiquées à la Convention-cadre des Nations Unies sur les Changements Climatiques (CCNUCC) et à l'OCDE. Pour combler les valeurs manquantes, des estimations basées sur des ensembles de données non officielles, telles que PRIMAP, sont utilisées, car elles présentent la plus grande concordance avec les données officielles en termes de niveaux, de tendances et de catégories d'émissions. Lorsque les données officielles seront disponibles dans le cadre de transparence renforcé de l'Accord de Paris, les valeurs estimées seront remplacées et il ne sera pas nécessaire de collecter des données supplémentaires. L'ensemble de données ventile les émissions par catégories du GIEC, gaz, pays et années, en faisant la distinction entre les points de données officiels et les estimations. Cet ensemble de données harmonisées renforce la transparence, améliore l'efficacité de la collecte des données et permet une analyse politique plus solide, fondée sur des données probantes. Pour aller de l'avant, le document recommande d'affiner l'ensemble de données avec des données sectorielles et de favoriser la collaboration avec les organisations internationales afin d'améliorer la fiabilité et la cohérence des données et des indicateurs sur les émissions de GES au niveau mondial.

Mots-clés : Émissions de gaz à effet de serre, données officielles, atténuation du changement climatique, sources d'émissions, inventaires nationaux.

Classification JEL : C82, E01, Q54, Q56

Acknowledgements

The authors would like to thank the Working Party on Environmental Information (WPEI) for their detailed review of the paper. The paper also benefitted from discussions at the International Programme for Action on Climate (IPAC) Technical Expert Group, and technical exchanges with the International Energy Agency (IEA) Energy Data Centre, the PRIMAP-hist team, and the European Commission Joint Research Centre. This paper is part of the OECD workstream on tracking countries' progress towards the Paris Agreement goals including the Environment Directorate, Statistics and Data Directorate, and the Economics Department.

This paper has been drafted by Miguel Cárdenas Rodríguez, Emma Schwentner, Edoardo Falchi, Andrzej Suchodolski, and Rodrigo Pizarro from the OECD Environment Directorate. The paper is under the supervision of Nathalie Girouard, Head of the Environmental Performance and Information Division in the OECD Environment Directorate.

The paper benefited from comments provided by Ivan Haščič, Nathalie Girouard, Daniel Nachtigall, and Sarah Miet from the Environment Directorate; Santaro Sakata from the Statistics and Data Directorate; Noura Takrouri Jolly from the Trade and Agriculture Directorate; and Nick Johnstone, Roberta Quadrelli, Pouya Taghavi-Moharamli, and Arnaud Risquez Martin from the IEA Energy Data Center.

Table of contents

FIGURES

TABLES

Executive summary

Anthropogenic emissions of greenhouse gases (GHG) have continued to rise since the Industrial Revolution, leading to significant alterations in the atmosphere, oceans, cryosphere, and biosphere. Tracking GHG emissions in a harmonised manner, including the identification of the sources of emissions, is essential to provide informed assessments and policy advise. Expanding the official information basis increases transparency, improves the robustness of policy analysis, and fosters our collective responsibility to reduce GHG emissions.

GHG inventory data feed into multiple OECD and IEA analytical products and workstreams. These include country reviews (Environmental Performance Reviews, Economic Surveys, Energy Reviews, etc.) and empirical analysis of policy effectiveness. GHG emission inventories are also the primary data input for estimating Air Emission Accounts (AEA) in line with the System of Environmental-Economic Accounting (SEEA), estimating demand-based (or footprint) emission indicators, and environment-economic modelling. Complete and timely GHG data will inform and drive forward efficient policies for the green transition in all countries.

Despite the existence of numerous datasets on GHG emissions, the availability of official (i.e., country-reported or -validated) national data remains mostly limited to developed countries. The most extensive datasets using data reported by countries, or estimates validated by countries, are the national GHG emission inventories reported to the United Nations Framework Convention on Climate Change (UNFCCC) and the GHG Emission by Source dataset collected from countries by the OECD. At this juncture, official national GHG inventories reported to UNFCCC provide complete coverage for Annex-I parties to the UNFCCC, whereas many non-Annex I countries present important data gaps[1]. OECD data collection fills the gaps for a selection of non-Annex I countries. Still, sizeable economies do not report any additional information to the OECD. These data gaps hamper the use of official information for policy analysis.

The aim of this paper is to compile a complete and comprehensive dataset on national GHG emission inventories, filling gaps by using available unofficial datasets closely aligned with UNFCCC and OECD data. This global dataset includes all officially reported data available. Data gaps are filled with values estimated based on the strategy developed in this paper. The dataset includes a breakdown of GHG emissions by IPCC emission categories, gases, countries, and years. Finally, the dataset includes additional data attributes to differentiate official data points from estimates and indicate the type of estimation performed.

To develop the dataset, unofficial datasets publishing national GHG emission inventories are examined and compared against official data. Three datasets have been selected for analysis based on their emissions scope, statistical principles, data availability, timeliness, and their applicability in policy analysis. These datasets are 1) the PRIMAP-hist national historical emission time series (PRIMAP); 2) the Emission Database for Global Atmospheric Research (EDGAR), and 3) the Climate Watch GHG Emission database

[1] Note that the Paris Agreement distinguishes between ‘developed’ and ‘developing’ parties to the UNFCCC instead of ‘Annex-I’ and ‘non-Annex I’. This paper uses the latter terminology as the analysis focuses on historical data predating the reporting expected under the Paris Agreement.

(CLIMWATCH).[2] The comparison against official data is based on emission levels, trends, and the relative composition of emission categories and subcategories.

The examination of the three unofficial datasets reveals a stronger alignment of PRIMAP with UNFCCC compared to EDGAR and CLIMWATCH, for both Annex-I and Non-Annex-I countries. PRIMAP demonstrates significant consistency with official sources is terms of total emission levels, emission trends and category composition. The alignment of PRIMAP to official data is observed only for aggregated categories of emissions. Detailed IPCC emission categories could not be examined, as none of the unofficial datasets present comparable data.

Based on this analysis, the paper develops a strategy for compiling a global dataset on national GHG emission inventories, anchored on official data reported to the OECD Environment Directorate and UNFCCC. The gaps in data are filled with estimates based on PRIMAP trends, maintaining the alignment with officially reported emission levels. Future updates of this global inventory compilation will prioritise the integration of all official information accessible through UNFCCC's data reporting mechanisms. As official data become available, including historical emissions, estimated values would be replaced, and additional OECD data collection would not be necessary. In this regard, the dataset aims to increase efficiencies in data collection and support the transition to the improved reporting under the Enhanced Transparency Framework of the Paris Agreement. This comprehensive dataset enables aggregate calculations, establishes a harmonised dataset for OECD activities, facilitates the development of indicators to track GHG emissions reduction progress, and supports more robust evidence-based policy analysis.

Moving forward, the dataset would need to be refined, in terms of the breakdown of emission source categories, to meet data demands for more tailored policy analysis. Such disaggregation could be explored with the use of sector-specific datasets produced by specialised agencies such as the GHG Emissions from Energy dataset by the International Energy Agency (IEA), and the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT) Emissions Totals dataset on GHG emissions from Agriculture and LULUCF. Throughout this workstream, technical collaboration with institutions producing the datasets discussed in this paper is sought to improve the reliability of data disaggregation. Importantly, close coordination with other international organisations using GHG emission estimates in their analytical work, including the IMF, the IEA, EUROSTAT, UNSD will be pursued to increase the consistency and comprehensiveness of statistics and indicators produced globally.

[2] It is important to note that the three datasets use either IEA data on energy balances or IEA estimates of GHGs from fuel combustion as inputs in their estimation.

1 Introduction

Anthropogenic emissions of greenhouse gases (GHG) have continued to increase since the Industrial Revolution, causing widespread and rapid changes in the atmosphere, ocean, cryosphere, and biosphere (IPCC, 2022[1]). Tracking GHG emissions systematically, and in a harmonised manner, is essential to understand the policy levers to mitigate climate change. While numerous GHG emission estimates have been produced by international organisations, academia, or the private sector; country-validated (i.e., official) data are still limited. Complete and timely official GHG emission data are needed to inform and drive forward efficient policies for the green transition in all countries.

Official GHG emissions data feed into multiple analytical products and workstreams of the OECD and IEA. These include datasets (OECD, 2023[2]), indicator reports (Environment at a Glance (OECD, 2023[3])), country reviews (Environmental Performance Reviews, Economic Surveys, Energy Reviews, etc.), dashboards (GHG Emission Trends and Trajectories by IPAC (Pizarro et al., 2024[4])), the mapping of climate policies to GHG emissions by the Inclusive Forum on Carbon Mitigation Approaches (IFCMA)[3], empirical analysis of policy effectiveness (D'Arcangelo, Kruse and Pisu, 2023[5]), and environment-economic modelling (e.g., the OECD Environmental Outlook (OECD, 2012[6])). Importantly, national GHG emission inventories are also the main data input for the construction of complementary datasets and indicators on GHG emissions, such as Air Emission Accounts (AEA) in line with the System of Environmental-Economic Accounting (SEEA) ((Flachenecker, Guidetti and Pionnier, 2018[7]), (OECD, 2023[8])) and demand-based (or footprint) emission indicators ((Yamano and Guilhoto, 2020[9]), (OECD, 2021[10])).

Data on national GHG emissions reported to the United Nations Framework Convention on Climate Change (UNFCCC) and the dataset GHG Emissions by Source of the Organisation for Economic Co-operation and Development (OECD) are country reported and therefore considered to be official data (OECD, 2023[2]). The reporting of GHG emission data to UNFCCC (hereafter UNFCCC data) and OECD follows the reporting Guidelines of the Intergovernmental Panel on Climate Change (IPCC). GHG emissions are reported for six IPCC emission source categories. Complete time series are primarily available for Annex I parties to the UNFCCC. Due to different reporting requirements dating back to the Kyoto protocol[4], most non-Annex I Parties to the UNFCCC do not submit complete inventories systematically or in a standardised manner, resulting in missing data points on their historical national GHG emission inventories (Sakata, Aklilu and Pizarro, 2024[11]). The OECD, through its State of the Environment (SoE) questionnaire[5], complements the data collection of the UNFCCC for non-Annex I Parties which are

[3] See https://www.oecd.org/climate-change/inclusive-forum-on-carbon-mitigation-approaches/

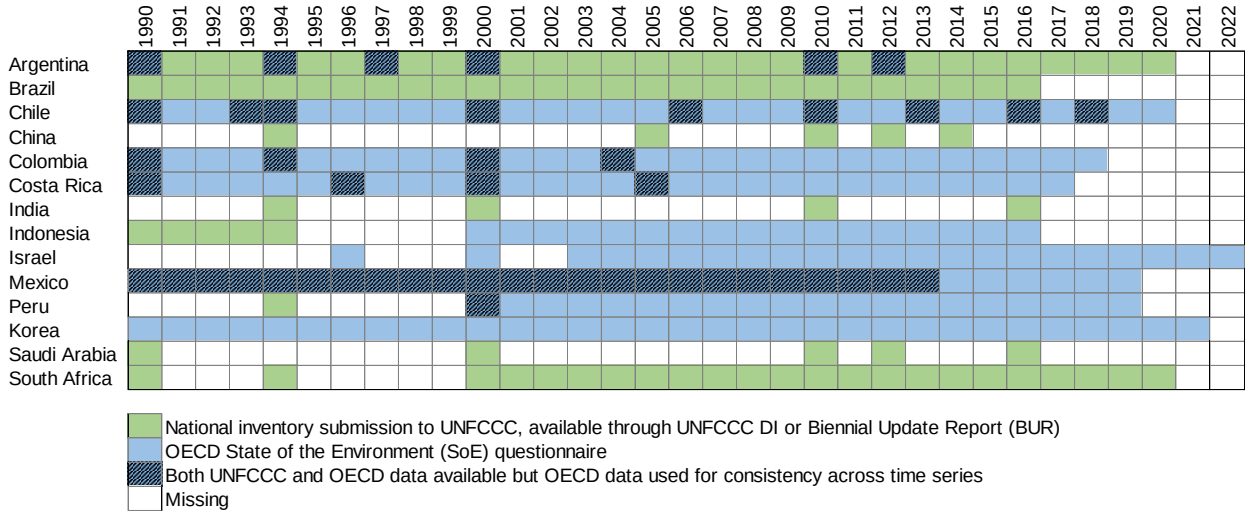
[4] Decision 15/CMP.1 from the first session of the Conference of the Parties serving as the meeting of the Parties to the Kyoto Protocol, held at Montreal from 28 November to 10 December 2005.

[5] The SOE questionnaire was developed in 1980 and has been used jointly with Eurostat since 1988. The questionnaire responds to multiple OECD Council recommendations: 1979 OECD Council recommendation on reporting on the state of the environment; 1991 OECD Council recommendation on environmental indicators and information; 1998 OECD Council recommendation on environmental information; 2022 OECD council recommendation on Environmental Information and Reporting. The SOE questionnaire collects emission levels for the main IPCC categories and includes additional breakdown for gases and subcategories for agriculture emissions. Data

OECD member countries, OECD Accession Candidates or OECD Key Partners[6]. Still, despite decades of efforts to improve the information on GHG emissions, major economies have important data gaps in official inventories (Figure).

Figure 1. There are important data gaps in GHG emission inventories for selected UNFCCC non-Annex-I countries.

Availability of emissions data for total GHGs and total categories per year.



Note: Each cell is coloured if the value for the year is available in the source, blanks indicate if the value is missing. Note that disaggregated information may exhibit more data gaps than for the totals shown in this table.
Source: Authors' calculation based on information available as of July 2024.

The improved reporting requirements set out in the Enhanced Transparency Framework (ETF) of the Paris Agreement (Article 13) replace earlier requirements for the disclosure of information under the UNFCCC. The ETF is set out to receive the first Biennial Transparency Reports (BTR) by end-2024. Under the ETF[7], each UNFCCC party shall regularly provide: 1) national GHG inventories using IPCC guidelines and 2) other information necessary to track progress made in implementing Nationally Determined Contributions (NDCs). As such, the availability of GHG emission inventories is expected to increase substantially over time. Nevertheless, the ETF provides flexibility on the implementation of such reporting provisions to those developing country parties that need it in the light of their capacities[8]. In addition, under the ETF, countries are not required to submit complete historical emission inventories, only previously reported years and the reference points or base years for NDCs.

on emissions from agriculture are collected and validated in collaboration with the OECD Trade and Agriculture Directorate.

[6] UNFCCC non-Annex I countries under OECD membership include Chile, Colombia, Costa Rica, Israel, Korea, Mexico. OECD accession candidates which are non-Annex I include Argentina, Brazil, Indonesia, Peru, and Thailand (other OECD accession candidates such as Bulgaria, Croatia, and Romania are Annex-I), OECD Key Partners which are non-Annex I Parties to the UNFCCC include: Brazil, People's Republic of China (hereafter 'China'), India, Indonesia, South Africa.

[7] Article 13 paragraph 7.

[8] Article 13 paragraph 2.

This paper examines several sources of unofficial GHG emission data to fill gaps in official GHG emission inventories. Three datasets are selected for the analysis: 1) the PRIMAP-hist national historical emission time series (PRIMAP); 2) the Emission Database for Global Atmospheric Research (EDGAR), and 3) the Climate Watch GHG Emission database (CLIMWATCH). The analysis of the three unofficial datasets reveals a stronger alignment of PRIMAP with UNFCCC compared to EDGAR and CLIMWATCH, for both Annex-I and Non-Annex-I countries. PRIMAP demonstrates significant consistency with official sources is terms of total emission levels, emission trends and category composition. Based on these results, the paper develops a data-gap filling strategy for compiling a global dataset on GHG emission inventories, anchored on official data reported to the OECD Environment Directorate and UNFCCC.

This paper is organised as follows. The next section identifies and describes the characteristics of unofficial datasets on GHG emissions. Section three compares the data against UNFCCC data over multiple dimensions. Based on these comparisons, section four develops a strategy to fill data gaps in official inventories, defining a list of priority datasets and imputation methods. In the section of next steps, the paper discusses potential improvements to the OECD data collection exercises and proposes several options to advance the disaggregation of the dataset. Additional comparisons across datasets are presented in Annex A, while Annex B presents initial estimates of the dataset for country groupings.

2 National GHG emission inventory datasets

GHG data can be compiled according to different statistical principles, classification systems, and applied to different scopes of emissions. While estimates of GHG emissions with ample geographic and time coverage are generally available, these estimates tend to vary widely among datasets due to differences in compilation methods and scope. Box summarises the main statistical principles used for compiling GHG emissions data and official national GHG emission inventories reported to the UNFCCC and the OECD (Sakata, Aklilu and Pizarro, 2024[11]).

Box 1. Compilation principles and classifications of GHG emission data

Emissions may be attributed to the producer or consumer of the good or service which generated the GHGs. The national GHG emissions inventory data by OECD and UNFCCC measure emissions from the **production perspective**. The OECD dataset on CO_2 emissions embodied in international trade, estimates CO_2 emissions based on the consumption (demand) perspective.

Emissions may be allocated to the territory where GHGs are physically emitted, or to the jurisdiction where the economic units are resident irrespective of the geographic location of the emissions. Inventory data by OECD and UNFCCC are compiled according to the **territory principle**. The SEEA Air Emission Accounts (AEA) record air emissions according to the residence principle.

Emissions may be classified according to the (physical) source of emissions or the economic activity (i.e., industry) generating the emissions. Inventories are classified according to **IPCC emission source categories**. The SEEA AEA record air emissions according to economic activities based on the International Standard Industrial Classification (ISIC).

The sectoral **scope of emissions** in national GHG emission inventories may vary according to the emission categories, activities, gases, and production processes considered. National inventories include categories on Energy, Industrial Processes and Product Use, Agriculture, Land-use, Land-use change, and Forestry, and Waste. Other datasets may focus on a specific activity, e.g., The most detailed GHG emission dataset by the International Energy Agency (IEA) dataset on CO_2 emissions, focuses on CO_2 emissions from fuel combustion.

Source: Based on (Sakata, Aklilu and Pizarro, 2024[11])

The criteria to assess the suitability of GHG emission datasets to fill the gaps of national official inventories encompass: the scope of emissions, the statistical principles, data coverage and availability, timeliness, and the use of the dataset in policy analysis. Datasets must follow the statistical principles outlined by the IPCC guidelines (IPCC, 2006[12]), this would ensure that the scope of emissions is aligned with the official UNFCCC inventories, and the structure of emission sources or categories is respected (UNFCCC, 2023[13]). Next, the dataset must provide a breakdown by gas and emission category, for long time series.

The datasets must be free of access and updated regularly. Finally, the dataset must be widely used in policy documents and scientific research.

Following these criteria, three unofficial inventory datasets have been selected[9]:

- The PRIMAP-hist national historical emission time series (PRIMAP)
- The Emission Database for Global Atmospheric Research (EDGAR)
- The Climate Watch GHG Emission/Climate Analysis Indicators Tool database (CLIMWATCH).

These three datasets rely on two conceptually different compilation methodologies. On the one hand, PRIMAP and CLIMWATCH use multiple datasets and imputation rules to estimate complete emission inventories. EDGAR, on the other hand, estimates emissions based on a modelling approach, using emission factors and economic activity data. Table presents a summary of the main characteristics of each dataset. The remainder of this section summarises their methods, scope, and underlying sources used for their compilation.

Table 1. Three selected unofficial datasets that provide global coverage and high timeliness.

	Official		Unofficial		
Dataset	**UNFCCC inventories**	**OECD greenhouse gas emissions by source**	**PRIMAP**-hist, Potsdam Institute for Climate Impact Research	**EDGAR,** Joint Research Centre, European Commission	**Climate Watch,** World Resources Institute
Sources	Country reported	Country reported	CRF, BUR, NIR, NC, UNFCCC DI, CDIAC, BP, Andrew, FAOTSTAT, EDGAR, IEA	Model-based, aligned with IEA GHG emissions from energy.	IEA, FAOSTAT, EPA, Andrew.
Geographic coverage	Annex I, Incomplete for non-Annex I	Non-Annex-I OECD member countries and accession candidates	World	World	World
Time coverage	1990-2021	1990-2021	1750-2022	1990-2022	1990-2020
Timeliness	t – 2 or 3	t – 2 or 3	t – 1 or 2	t – 1 or 2	t – 3 or 4
GHGs	All	All	All	All	CO_2, CH_4, N_2O, F-gases
Category disaggregation	Maximum level	Only for energy	Only for energy, ind. processes and agriculture for CO_2, CH_4, N_2O	High level of detail, especially for energy	Only for energy

Note: "All" GHGs in this table refers to CO_2, CH_4, N_2O, and F-Gases disaggregated into HFCs, PFCs, SF_6, and NF_3. Table includes more details on the coverage of emission categories.
Source: EDGAR v8 from October 2023, PRIMAP v2.5 from October 2023, CLIMWATCH vDec2023, OECD data extracted December 2023.

It is important to note that these three unofficial datasets use either IEA data on energy balances and/or IEA estimates of GHGs from fuel combustion to increase the data availability and disaggregation detail on

[9] Due to these criteria, widely used datasets covering specific sectors or gases (e.g., GHG emissions from fossil combustion by the IEA, CDIAC for CO_2 emissions) are not used for the comparison. Yet, these specialised datasets are compared against official data somewhat indirectly, since they are used in the creation of EDGAR and CLIMWATCH.

energy related emissions. IEA collects energy balances data via standardised fuel-specific questionnaires distributed to over 50 IEA member and non-member countries; and works with counterparts from other countries to compile comparable energy statistics based on existing official sources wherever possible.[10] With respect to GHG emissions data, the IEA estimates emissions using energy balances and default fuel-specific emission factors (Tier 1 methodology[11]) provided by IPCC. The IEA also conducts a comparative analysis of IEA data and the energy data underlying the national GHG inventory submissions.[12] IEA expertise would be fundamental for developing a data gap-filling strategy specific to energy emission inventories.

2.1 The PRIMAP-hist national historical emission time series (PRIMAP)

The PRIMAP dataset is estimated by imputing data based on a hierarchy of other datasets and sources. The hierarchy of imputation distinguishes between two types of data sources: country-reported or country-validated data, and third-party data (Gütschow and Pflüger, 2023[14]). Two time series are produced: HISTCR where country-reported data are prioritised over third-party data; and HISTTP, where third-party data are prioritised over country-reported data. The time series HISTCR, where country-reported data are prioritised, are the series used for the analysis in this paper.

PRIMAP's priority of datasets is determined based on their completeness and reliability, and it differs across sectors and gases. For example, third-party datasets produced by specialised energy agencies (e.g., CDIAC, IEA) are prioritised for fossil and industrial CO_2 emissions, while FAO datasets are prioritised for CH_4 emissions. A description of the prioritisation exercise is available in (Gütschow et al., 2016[15]). Complete time series are calculated based on the highest-priority dataset imputing data gaps using emission trends from lower-priority datasets. Regional growth rates or growth rates from subcategories are used to expand the geographic coverage. Numerical extrapolation is used only when no national or regional data is available. The time series is made available with and without numerical extrapolation, and the series with extrapolated data until 2022 is used for the present analysis. PRIMAP is maintained and updated at least once a year[13], and makes available a comprehensive description of the compilation methodology and metadata. The metadata documentation includes detailed and clear information about special treatment of datapoints[14] or important changes between dataset versions[15].

[10] IEA World Energy Balances and Statistics are collected following the methodologies set by the International Recommendations on Energy Statistics (IRES), adopted by the UN Statistical Commission, ensuring harmonized methodologies and definitions across all countries.

[11] The IPCC guidelines provide three *tiers* of emission estimation methods, representing varying levels of methodological complexity. Tier 1 is the basic method, it makes use of default emission factors and simple country specific activity. Tier 2 methods apply emission and stock change factors based on county- or region-specific data for key categories. Tier 3 methods are the most complex, they are based on models and inventory measurement systems, repeated over time, using detailed spatial and temporal activity data (IPCC, (2019[25]).

[12] The IEA is co-author of the 1996, 2006 and 2019 editions of the IPCC Guidelines for GHG inventories. In addition, the IEA and the UNFCCC initiated a collaborative programme in 2023 to improve the technical capacity of selected countries on energy statistics to reinforce transparency in reporting national GHG inventories and mitigation actions, in accordance with the requirements under the Enhanced Transparency Framework of the Paris Agreement.

[13] Johannes Gütschow, personal communication, May 16, 2024

[14] For example, series are truncated at zero when extrapolations of low values result in negative emissions.

[15] Comprehensive changelogs and metadata are available at their website, detailing adjustments done for specific countries, gases, or years. For example, see changelog of version 2.5 at https://doi.org/10.5281/zenodo.10006301.

Country-reported data used in the compilation of PRIMAP include, in order of priority, the Common Reporting Format (CRF)[16], Biannual Update Reports (BUR)[17], National Inventory Reports (NIR)[18], National Communications (NC)[19], and the UNFCCC Digital Interface (UNFCC DI)[20]. Third-party data include, in order of priority, the Carbon Dioxide Information and Analysis Center (CDIAC)[21] of the United States Department of Energy, Statistical Review of World Energy by British Petroleum (BP), Global CO_2 emissions from cement production data – also known as "Andrew" data (Andrew, 2023[16]), the Food and Agriculture Organization data on agricultural emissions, and the Emission Database for Global Atmospheric Research (EDGAR) of the Joint Research Centre (JRC) of the European Commission[22]. Note that third-party datasets, such IEA data on CO_2 emissions from fuel combustion, are incorporated in EDGAR and are not mentioned explicitly as a source in PRIMAP's metadata. Finally, note that PRIMAP does not currently use OECD data in the data gathering processes.

2.2 Emission Database for Global Atmospheric Research (EDGAR)

EDGAR is a global dataset of anthropogenic emissions of greenhouse gases and air pollutants (Joint Research Centre (JRC), 2023[17]). EDGAR emission estimates are modelled using economic activity data taken from international statistics and emission factors, i.e., a tier 1-2 approach of the IPCC 2006 Guidelines. Details on the economic activity data and the emission factors used for different sectors are specified in (Janssens-Maenhout, 2019[18]). The model and data sources vary by emission category. For example, EDGAR uses IEA data on CO_2 emissions from fossil fuel combustion for the energy emission category (IEA, 2023[19]). For other IPCC categories, emissions in EDGAR are calculated as a function of economic activity data, the technology mix of the economic activity, abatement percentage by technology, and country-sector specific emission factors. For instance, agriculture emissions from livestock are estimated using livestock data from FAOSTAT combined with approximations for animal waste per head.

In addition to estimates of annual GHG emissions by countries and emission categories, EDGAR makes available data products with various geographical and temporal disaggregation. For example, EDGAR makes available gridded (spatially allocated) data using the downscaling factors related to the location of energy facilities, road networks, shipping routes, population density and agricultural land use. Monthly estimates are calculated by applying sector- and region-specific monthly profiles. Data on emissions from

[16] CRF tables are a series of standardized data tables containing mainly quantitative information. CRFs are part of the standardized reporting requirements for Annex-I Parties.

[17] BURs are reports to be submitted by non-Annex I Parties, containing updates of national GHG inventories, including a national inventory report and information on mitigation actions, needs and support received.

[18] NIRs contain detailed information on the inventory. They include descriptions of the methodologies used in the estimations (including references and sources of information), the data sources, the institutional arrangements for the preparation of the inventory (including quality assurance and control procedures), and recalculations and changes compared with the previous inventory. NIRs are part of the standardized reporting requirements for Annex-I Parties.

[19] NC is a report including descriptions on the country's national circumstances, GHG inventories and the steps taken to mitigate and adapt to climate change. Since 2010, non-Annex-I parties are expected to submit NCs every four years.

[20] The UNFCCC DI compiles GHG emissions data submitted to UNFCCC in a standardised format. It may not include emissions data reported as text documents.

[21] The time series was originally developed by the U.S. Department of Energy. CDIAC fossil fuel CO_2 emissions series are now maintained by the Appalachian State University. The new version, referred to as CDIAC-FF can be found at https://energy.appstate.edu/cdiac-appstate/data-products.

[22] Historical emissions pre-1960 are based on estimates developed with the Community Emissions Data Systems (CEDS).

biological processes are reported separately to ensure that the scope of emissions remain aligned with IPCC reporting guidelines. EDGAR is frequently used for the calculation indicators at the global level which require comprehensive and complete data. For example, EDGAR is used by the Global Trade Analysis Project (GTAP) as the main source of GHG emissions data (Chepeliev, 2024[20]), and by Eurostat for the estimation of CO_2 footprints in the framework of the FIGARO (Full International and Global Accounts for Research in Input-Output) project (European Commission, 2024[21]).

2.3 Climate Watch GHG Emission database (CLIMWATCH)

Like PRIMAP, the CLIMWATCH series is compiled from multiple datasets. Unlike PRIMAP, CLIMWATCH does not use any official data in their compilation methodology. CLIMWATCH estimates a complete dataset imputing missing values based exclusively on third-party sources. The prioritisation includes: the CO_2 Emissions from Fuel Combustion dataset of the IEA, the United Nations Food and Agriculture Organization (FAO), U.S. Environmental Protection Agency (EPA), Global Carbon Project, Global CO_2 emissions from cement production data (Andrew 2021) (Climate Watch, 2022[22]).

3 Comparison of datasets

This section compares UNFCCC and OECD data with PRIMAP, EDGAR and CLIMWATCH. First, the level of data disaggregation in these datasets is examined. The IPCC classification can reach up to five levels of disaggregation depending on the category[23], and the availability of such level of detail varies substantially across official and unofficial datasets. The comparison in 3.1 Availability of disaggregated data on emission categories identifies gaps and differences in disaggregation levels, and highlights the suitability of each dataset for the compilation of inventories by detailed categories of importance for OECD analytical products.

Thereafter, the alignment of unofficial data with official data is evaluated for emission levels, trends, and the relative composition of emissions within IPCC categories (3.2 Alignment of emissions data). Comparisons are done separately for Annex-I and non-Annex-I parties to the UNFCCC, due to the substantial differences in data availability. The analysis is conducted for the three main GHGs (CO_2, CH_4, N_2O), as they represent 98% of estimated emissions[24] from all GHGs over 1990 to 2022. Section 3.3 Discussion summarises the findings and examines the options to compile a global dataset.

3.1 Availability of disaggregated data on emission categories

GHG emission data by the main IPCC source categories is fully sufficient for the analysis of overall countries' emission profiles. Yet, highly disaggregated data for subcategories is a key data input for analytical products such as the estimation of air emission accounts, modelling future emission scenarios, and the estimation of demand-based emission (footprint) indicators. Note that none of the unofficial datasets analysed estimate emissions over the fully disaggregated classification of IPCC categories. The suitability of unofficial sources to fill data gaps depends not only on how closely their values align with official data, but also on whether they provide disaggregated categories and subcategories that match totals.

The inventory datasets analysed have substantially different data availability by classification levels (Table). The availability of the highest levels of disaggregation is highly limited across unofficial datasets. The case of 1A Energy – Fuel Combustion is highlighted below because, globally, this category represents 55% of total emissions[25]. Disaggregated data is available for the 4th level subcategories of Energy only for EDGAR, these energy subcategories include important sources of emissions such as electricity and heat production, petroleum refining or manufacture of solid fuels and other energy industries. Figure A.3 and Figure A.4 in Annex A show more details on data availability by dataset, group of countries, and emission category.

[23] In IPCC guidelines, the first level of the classification of emission sources are in fact labelled "sectors", while lower levels are labelled "categories". This paper refers to "categories" for all levels of the IPCC classification, to prevent a possible confusion related to other uses of the term "sector" (e.g., economic sectors, institutional sectors).

[24] Based on PRIMAP emission estimates.

[25] Based on PRIMAP emissions estimates over the 1990 – 2022 period

Table 2. The breakdown by emission categories varies substantially across datasets.

Data available by emission source categories across datasets

UNFCCC source categories	OECD ENV	PRIMAP	EDGAR	CLIMWATCH
Total excluding LULUCF	Yes	Yes	Yes	Yes
1. Energy	Yes	Yes	Yes	Yes
A. Fuel combustion	Yes	Yes	Yes	Yes
3rd level (e.g., transport)	Yes	No	Partial	Yes
4th level (e.g., road transport)	No	No	Partial	No
B. Fugitive emissions from fuels	Yes	Yes	Yes	Yes
3rd level (e.g., oil and gas fugitive emissions)	No	Yes	Yes	No
C. CO_2 transport and storage	Yes	Yes	No	No
2. Industrial processes and product use	Yes	Yes	Yes	Yes
A. Mineral industry	No	Yes	Yes	No
3rd level (e.g., cement production)	No	No	Yes	No
B. Chemical industry	No	Yes	Yes	No
C. Metal industry	No	Yes	Yes	No
D. Non-energy products from fuels and solvent use	No	Yes	Yes	No
E. Electronic Industry	No	Yes	No	No
F. Product uses as ODS substitutes	No	Yes	No	No
G. Other product manufacture and use	No	Yes	Yes	No
H. Other	No	Yes	No	No
3. Agriculture	Yes	Yes	Yes	Yes
2nd level (e.g., Manure management)	No	No	Partial	No
4. Land use, land-use change and forestry (LULUCF)	Yes	Yes	No	Yes
2nd level (e.g., forest land)	No	No	No	No
5. Waste	Yes	Yes	Yes	Yes
2nd level (e.g., solid waste disposal)	No	No	Partial	No
6. Other	Yes	Yes	Yes	Yes
Memo item: International bunkers	Yes	No	No	Yes

Note: This table shows the coverage of disaggregated categories by dataset. The column for data reported to UNFCCC is not included as it corresponds to the highest level possible of the category classification. Partial coverage indicates that only some individual subcategories are published. UNFCCC source categories follow the 2006 IPCC guidelines with the exception that guidelines group categories three and four into an AFOLU category. Please note that the first-level emission source categories are labelled as "sectors" in 2006 IPCC guidelines, while lower levels are labelled "categories".
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023.

The availability of disaggregated data has an impact on the suitability of unofficial sources to complement official data. In the absence of granular data, ad-hoc estimations done to complete input data used in the calculation of indicators are likely to generate issues of comparability and consistency. In addition to limited disaggregation, unofficial datasets which publish granular subcategories also show important misalignments relative to official sources. Details on these comparisons are presented and discussed in 3.2.3 Emissions by category and in Annex A.

3.2 Alignment of emissions data

A graphical analysis of the differences of emissions of the three main GHGs (CO_2, CH_4, N_2O) with respect to UNFCCC is provided below. The analysis is conducted over the 1990-2022 period for 43 Annex- and

155 non-Annex-I countries and territories[26]. All series are compared using the Global Warming Potential of the IPCC Fourth Assessment Report (IPCC, 2019[23]) Annex A shows additional comparisons across datasets by gases, categories and subcategories.

3.2.1 Emission levels

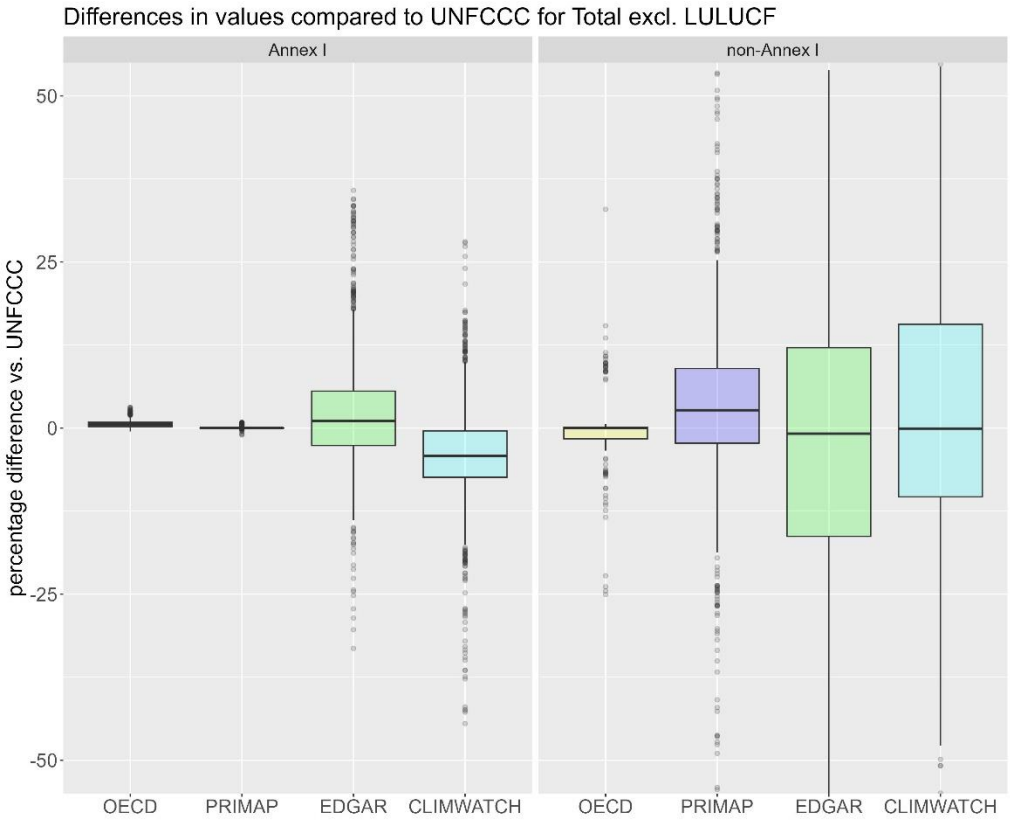
As expected, the alignment of unofficial datasets to UNFCCC is higher for Annex-I than for non-Annex-I countries (Figure)[27]. For both groups of countries, the comparison of datasets to official information shows relatively higher emissions levels in EDGAR and lower emission levels in CLIMWATCH. For total emissions and the first level disaggregation of IPCC categories, PRIMAP levels are better aligned with official UNFCCC data than EDGAR and CLIMWATCH. Interestingly, OECD data also shows some discrepancies with respect to UNFCCC data for some non-Annex I countries, which could indicate that countries might report or validate different versions of inventories.

[26] See a list of Annex-I and non-Annex I parties to the UNFCCC at https://unfccc.int/process/parties-non-party-stakeholders/parties-convention-and-observer-states?field_national_communications_target_id%5B514%5D=514

[27] This figure shows the relative differences between UNFCCC data and other datasets for the sum of the main three GHGs for all categories excluding LULUCF. The relative difference is expressed in percentage terms and is calculated between the total emissions level of each dataset and the levels reported to UNFCCC, for each country and year. i.e., $Diffsource_{ijt} = 100 * \left\{\left(\frac{Source_{ijt}}{UNFCC_{ijt}}\right) - 1\right\}$

Figure 2. PRIMAP emission levels align best with UNFCCC data among unofficial inventories.

Distribution of the percentage difference of each dataset with respect to UNFCCC. CO_2, CH_4 and N_2O emissions for all IPCC emission categories except LULUCF, 1990-2021.



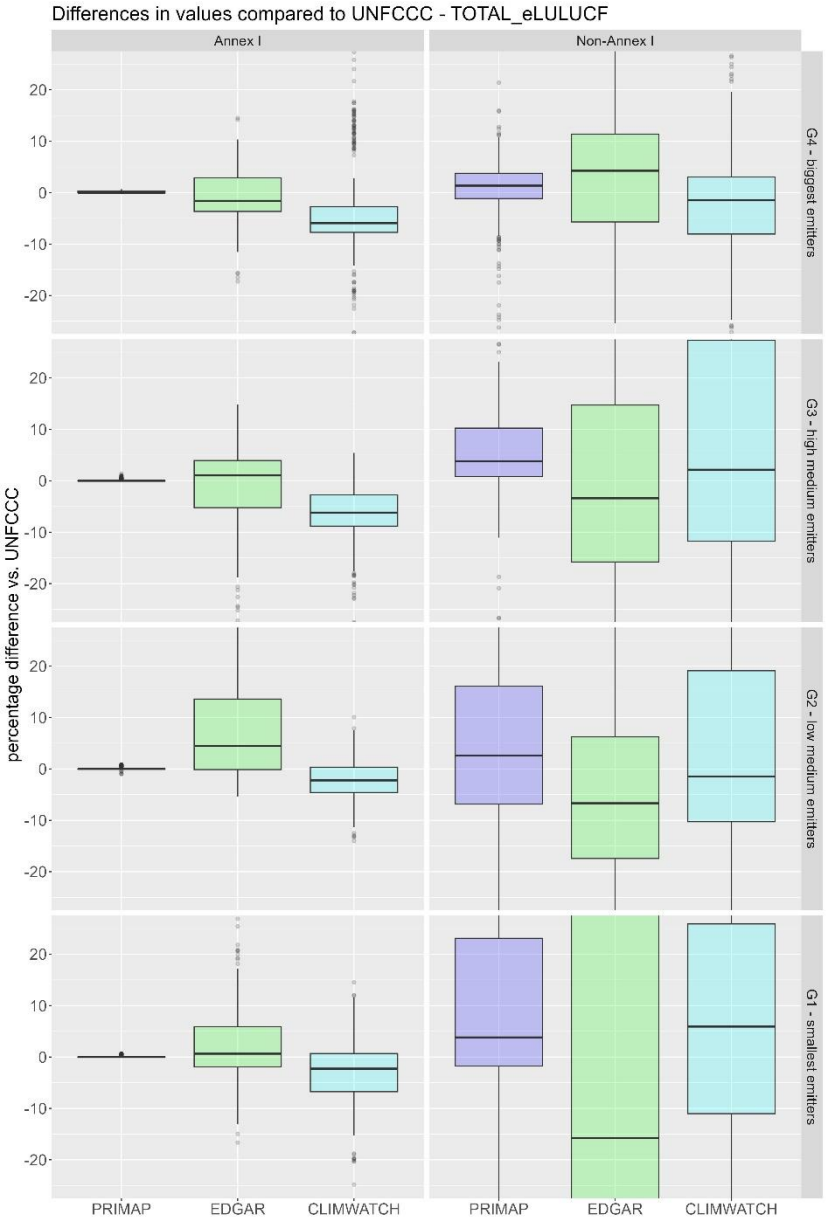
Note: Boxplots show the distribution of observed values. The box in the middle represents the central 50% of the data, while the box limits show the first and third quartile (i.e., the interquartile range). The horizontal line inside the box represents the median. The lines extending from the box show the data outside the quartiles extending 1.5 times the interquartile range. Outliers are plotted individually as points. The size of the box indicates the spread of the data. In the comparison of datasets, a smaller boxplot centred around zero indicates less spread of the distribution and better alignment to the official source.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Comparisons based on the sum of the three main GHGs may be subject to uncertainties arising from the aggregation. To explore this, additional comparisons are conducted for each gas in Figure A.2. Results confirm that PRIMAP data is best aligned to official data for each individual gas, and that EDGAR and CLIMWATCH exhibit similar differences with respect to UNFCCC official values for the three gases either individually or aggregated. This result confirms that comparisons based on the sum of the three gases can serve as a good proxy for comparisons by gas.

Next, countries are grouped into four baskets based on total emission levels, allowing to analyse the comparability of unofficial sources according to relative size of countries (Figure). The comparison shows that, both for Annex I and non-Annex I countries, PRIMAP is best aligned to official data for large emitters.

Figure 3. Emission levels of PRIMAP align best for big emitters.

Differences in emission levels, unofficial data relative to UNFCCC. Countries are grouped based on total emission levels excl. LULUCF.



Note: See note on Figure 3.1 for details on the interpretation of boxplots. Countries are grouped into four baskets according to total emissions excl. LULUCF, based on UNFCCC data. The 4th group (G4) includes the countries with the highest level of total emissions relative to the sample (i.e., countries above the third quartile of the distribution), while the 1st group (G1) includes the countries with the lowest level of emissions (i.e., countries below the first quartile of the distribution).
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

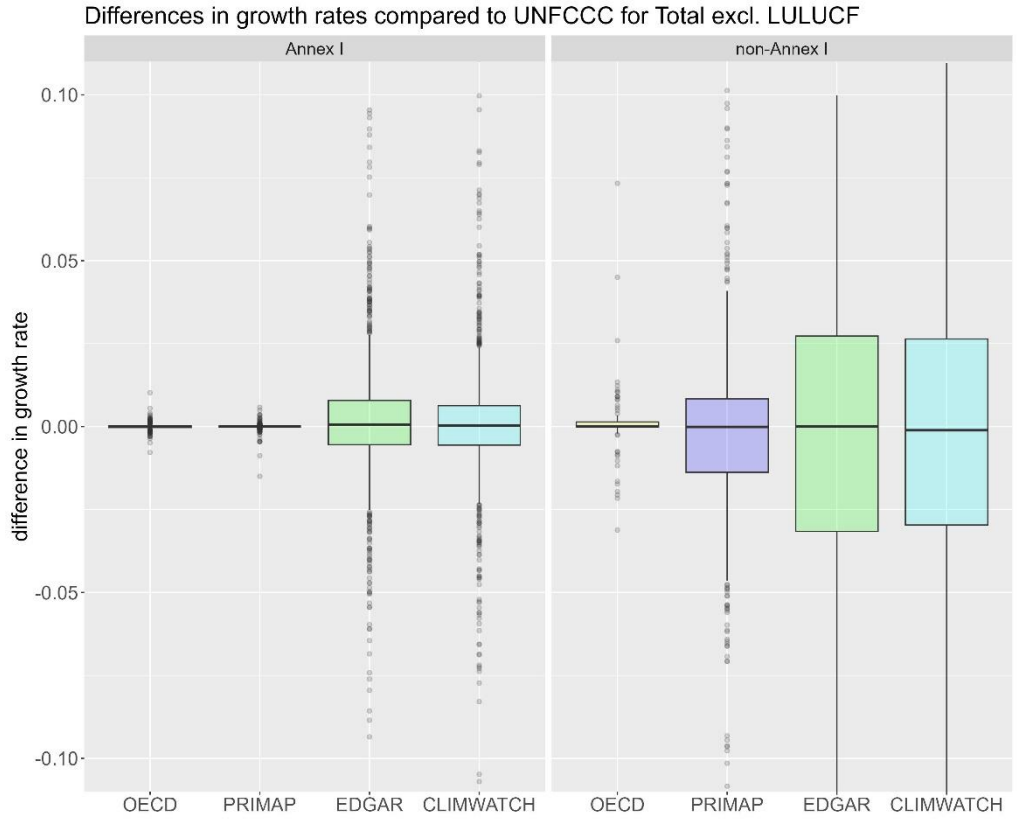
3.2.2 Trends of emissions

Next, the trends of emissions are compared across datasets. Figure shows for instance the difference in growth rates between each dataset and UNFCCC for the sum of the three main GHGs for all categories

excluding LULUCF[28]. As expected, overall data alignment of unofficial data with respect to UNFCC in terms of growth rates, is higher for Annex I countries than for non-Annex I. For total emissions and the first level disaggregation of IPCC categories, PRIMAP growth rates are better aligned with official data growth rates than EDGAR and CLIMWATCH.

Figure 4. PRIMAP emission trends align best with UNFCCC data.

Distribution of differences of growth rates, for total CO_2, CH_4, and N_2O emissions, all categories excluding LULUCF.



Note: See note on Figure 3.1 for details on the interpretation of boxplots.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

3.2.3 Emissions by category

To explore the alignment of category emissions across datasets, the composition of emission categories in unofficial data is compared against the composition in UNFCCC data. The contributions of the first level IPCC categories to total emissions per country and year are calculated as a share; these shares are then compared to UNFCCC shares[29].

Figure shows the comparison across datasets of the shares of each first level IPCC category in total emissions for the sum of the three main GHGs. The category shares are more aligned for Annex I than

[28] The difference in growth rates is calculated as $G\_diffsource_{ijt} = \ln\left(\frac{Source_{ijt}}{Source_{ijt-1}}\right) - \ln\left(\frac{UNFCCC_{ijt}}{UNFCCC_{ijt-1}}\right)$

[29] The shares of subsectors are calculated as $sharesource_{ikt} = \frac{Source_{ikt}}{\sum_{k=1}^{K} Source_{ikt}}$ where $k = \{1, \dots, K\}$ are all subsectors of a category. The difference in compositions are calculated as simple differences: $Diffsharesource_{ikt} = sharesource_{ikt} - shareunfccc_{ikt}$.

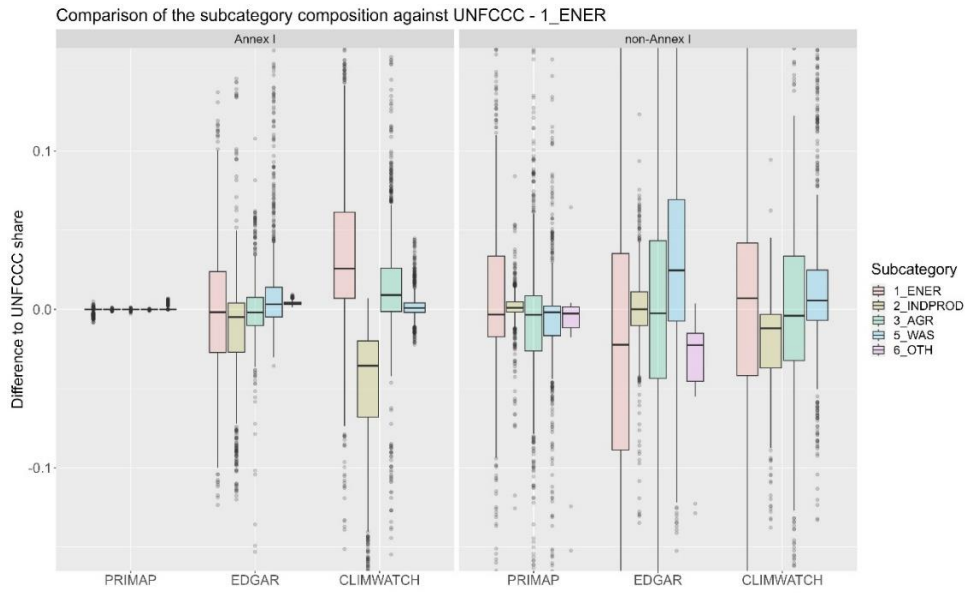
non-Annex I across almost all categories and datasets. PRIMAP is almost entirely aligned for Annex I countries and better aligned for non-Annex I countries than EDGAR and CLIMWATCH. Interestingly, EDGAR composition is significantly more aligned to UNFCCC composition than CLIMWATCH for Annex-I countries, while the opposite can be observed for non-Annex I countries.

Regarding specific emission categories, the share of energy from total emissions shows the largest discrepancies relative to official data across categories and datasets. This can be observed by the wider spread in its distribution and a higher interquartile range. The energy category accounts for the bulk of emissions and small relative differences could have a high impact on the comparisons. Industrial processes and product-use emissions seem relatively aligned for PRIMAP and EDGAR, while CLIMWATCH shows the largest discrepancies for this category.

Unofficial data on emissions from the agriculture and waste categories show a higher discrepancy for non-Annex I countries than for Annex-I countries. Such difference could respond to the inherent complexity of accounting for emissions from these categories in a harmonised manner, which may be aggravated for countries with limited institutional capacity on environmental information systems[30]. Finally, the category *other* generally makes up only a small share of total emissions. The observed differences are big relative to its size, nevertheless, comparing the *other* category is complicated, as it may have different scopes depending on the alignment of the dataset to official data and possible adjustments done by countries to match totals.

Figure 5. PRIMAP shows the best alignment to official data at the first IPCC category level.

Distribution differences in category shares from total emissions, 1990-2021, total CO_2, CH_4, and N_2O, percentage points.



Note: See note on Figure 3.1 for details on the interpretation of boxplots. See Annex A for additional comparisons by categories incl. LULUCF.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Annex A presents comparisons by category and gas (Figure A.6 for CO_2, Figure A.7 for CH_4, and Figure A.8 for N_2O). PRIMAP and EDGAR have smaller discrepancies in the energy category and industrial

[30] This could also explain the large discrepancies observed when comparing LULUCF data, as shown in Figure A.5.

processes category for CO_2 emissions compared to CH_4 and N_2O but higher discrepancies in the agriculture and waste category. Disparities are generally higher for category-gas combinations which have smaller relative contribution to total emissions, and category-gas emission processes which are generally subject to higher measurement uncertainty.

Additional comparisons are carried out in Annex A using the shares categories in total emissions to group countries. Such comparisons shed light on the alignment of datasets for countries according to the emission intensity of the category. Results show that PRIMAP is better aligned across the energy (Figure A.9), industry (Figure A.10) and agriculture (Figure A.11) category for countries where the relative size of the category is the highest. Conversely, differences in levels of PRIMAP with respect to UNFCCC data are most pronounced in countries where the share of the category is the smallest. These results suggest that PRIMAP is also a good proxy dataset of official data for the combination of country-sector with higher relative importance.

3.3 Discussion

The analysis above reveals that the PRIMAP series consistently exhibit a stronger alignment to official data than EDGAR or CLIMWATCH for total emission levels for both Annex-I and non-Annex-I countries. The analysis also shows that the data trends of PRIMAP also align more closely with the observed trends of UNFCCC data. In addition, PRIMAP is also aligned with official data, when comparisons are done across similar types of emission profiles of countries. This alignment indicates that PRIMAP may also serve as the basis for estimates of national emissions for interpolating and extrapolating official values on total emissions.

Regarding data by more granular IPCC categories, a preliminary analysis shows that data gaps at the first level of the IPCC classification could also be estimated using PRIMAP. Unfortunately, PRIMAP's granularity does not allow to assess its suitability to fill gaps beyond the first level category. EDGAR data, on the other hand, are available with higher disaggregation (at second, third or even fourth level for some IPCC categories), but initial comparisons show substantial differences with respect to official values. A deeper assessment of the suitability of EDGAR to fill data gaps would require accessing and analysing the underlying input data and assumptions behind EDGAR's model for each emission category.

Data gaps could be also filled for detailed emission categories and subcategories using sector-specific datasets from specialised data-producing agencies. Fortunately, in the case of energy emissions, the upcoming edition of the IEA GHG Emissions from Energy database will include a dedicated dataset with a complete mapping to IPCC categories down to the lowest level of disaggregation possible.[31] As such, the IEA estimates of emissions from energy by country could feed directly into the compiled global GHG emission inventories dataset. Beyond energy, other sector-specific datasets (e.g., from FAO) could be explored to assess their alignment to IPCC methodologies and their suitability to fill data gaps in official inventories.

[31] The dataset on GHG Emissions from Energy by IPCC categories is expected to be published by Q2-2024 (Roberta Quadrelli, personal communication, June 13, 2024)

4 Compiling a global dataset of GHG inventories

A global dataset of GHG inventories can be constructed by implementing a data gap-filling strategy based on PRIMAP data for total emissions for the first level of the IPCC categories. The strategy includes the following steps:

1. Use data reported to OECD by non-Annex-I countries.
2. Use UNFCCC data for Annex-I countries and remaining non-Annex-I.[32]
3. Missing values in OECD and UNFCCC data would be interpolated and extrapolated using PRIMAP trends.
4. PRIMAP estimates would be used to fill the missing values in cases where OECD or UNFCCC series are completely missing for the time series, i.e., when no official value serves as an anchor for the interpolation or extrapolation.
5. Disaggregation of the compiled totals, by category and gases, would be done:
    a) Using the shares in OECD and UNFCCC data when these are complete for all subcategories.
    b) Estimated using PRIMAP shares otherwise.

The data gap-filling strategy implies that missing data points are interpolated using PRIMAP data trends scaled to match available official values. If applicable, PRIMAP's values are used for those countries without official UNFCCC historical data. The compilation strategy also guarantees that data points subsequently reported to UNFCCC or OECD would fully replace any estimated value in the compiled dataset. Following this strategy and using currently available data, Annex B presents a proof of concept of the compiled estimates for country group aggregates.

To ensure transparency, the dataset clearly distinguishes official data points from the imputed (estimated) values, and indicates the source and estimation performed. This is done by labelling estimated data points with flags and including additional dataset attributes showing – for each data point – the primary source of data (OECD, UNFCCC, or PRIMAP) and, if applicable, the type of estimation performed (e.g., disaggregation using OECD shares of categories, backward extrapolation using PRIMAP trend).

The data gap-filling strategy will need to be refined over time. First, as official data becomes more available, the suitability of unofficial datasets to fill data gaps will have to be assessed periodically. Second, the dataset requires higher granularity on emission categories. The first level category in the IPCC classification helps to understand broad emission profiles of countries; however, analytical products require much more disaggregated information that is consistent with official totals. The use of sector-specific datasets, such as IEA's GHG Emissions from Energy dataset or FATOSTAT's Emissions Totals dataset

[32] UNFCCC data would include GHG emission inventories from the Biennial Transparency Reports (BTR) under the Paris agreement when they become available. The first submission of BTRs will take place by the end of 2024.

on GHG emissions from Agriculture and LULUCF, would help with such disaggregation. These options are proposed as next steps in the following section.

5 Next steps

A complete and timely dataset on GHG emissions based on official inventories is crucial for evidence-based analysis of climate change. GHG emission inventories are used extensively in country reviews, in the estimation of Air Emission Accounts and demand-based (or footprint) emission indicators, for conducting empirical analysis of policy effectiveness and environment-economic modelling. Improved availability and consistency of emissions data anchored to official data will facilitate international comparability, calculation of aggregated data, and enable tracking progress towards the reduction of greenhouse gas emissions.

This paper identifies three unofficial datasets of GHG emissions to expand the coverage and timeliness of official inventories. Comparing the datasets of PRIMAP, EDGAR and CLIMWATCH, against official data reported to OECD and UNFCCC, PRIMAP is identified as a proxy dataset that can be used for filling data gaps in emission levels and in trends. Filling data gaps allows the compilation of a global dataset with comprehensive coverage of categories, gases, and years.

Going forward, four objectives could be pursued:

- First, pilot testing the adjustment of the OECD SOE questionnaire on GHG inventories prefilled with the compiled dataset outlined here, would inform on the suitability of unofficial data to complement national inventories. This exercise would primarily concern countries with important data gaps on inventories and limited availability of national information to fill such gaps. Pending approval of countries piloting the prefilled SOE questionnaire, the resulting compiled dataset could be made publicly available.
- A second short-term objective would be to explore the options for filling data gaps on more disaggregated emission categories for the compiled dataset. The unofficial datasets examined in this paper show important discrepancies with respect to official data on key subcategories such as fuel manufacturing, electricity, and heat production. In this respect, collaboration among OECD and IEA committees involved in climate policy analysis would allow to 1) identify the level of granularity needed from inventories to carry out analytical work such as the estimation of Air Emission Accounts, 2) prioritise the categories and subcategories of interest for disaggregation, 3) identify potential sources of discrepancies across available unofficial estimates, and 4) develop a data gap-filling strategy to improve disaggregation and alignment to official data. With respect to the latter, options for further disaggregating subcategories of interest would include the use of sector-specific datasets produced by specialised agencies such as the GHG Emissions from Energy dataset by the IEA[33] and FAOSTAT's Emissions Totals dataset on GHG emissions from Agriculture and LULUCF.
- In the medium-term, a collaboration with the institutions producing the datasets studied in this paper would be sought. For example, OECD data collection and validation could be incorporated as underlying official data in unofficial datasets. Such alignment will improve coherence and transparency across publicly available datasets on inventories. The collaboration could also explore the feasibility of improving the granularity of the emission categories of interest to OECD and IEA products.

[33] The IEA plans to release a dedicated dataset on GHG emissions from energy with a complete mapping to IPCC categories down to the lowest level of disaggregation possible. Such a dataset would be essential for the disaggregation of the global compiled dataset.

- As a longer-term objective, the OECD data collection on national GHG emission inventories could be optimised, modernised, and refined. For example, the reporting templates could focus on filling missing values and missing disaggregation of priority gases and priority categories for OECD and IEA analysis, standardise the global warming potential applied and include details on the tier methodology used for data compilation. Improving the data collection and validation tools would also decrease the reporting burden and facilitate the processing of data.

References

Andrew, R. (2023), “Global CO2 emissions from cement production”, https://doi.org/10.5281/zenodo.8339353. [16]

Chepeliev, M. (2024), *Chapter 13B: Complementary Greenhouse Gas Emissions Database*, Center for Global Trade Analysis. [20]

Climate Watch (2022), *Climate Watch*, https://www.climatewatchdata.org/ (accessed on December 2023). [22]

D’Arcangelo, F., T. Kruse and M. Pisu (2023), “Identifying and tracking climate change mitigation strategies: A cluster-based assessment”, *OECD Economics Department Working Papers*, No. 1786, OECD Publishing, Paris, https://doi.org/10.1787/a23b43c5-en. [5]

European Commission, E. (2024), “Methodological note on FIGARO - CO2 estimates”. [21]

Flachenecker, F., E. Guidetti and P. Pionnier (2018), “Towards global SEEA Air Emission Accounts: Description and evaluation of the OECD methodology to estimate SEEA Air Emission Accounts for CO2, CH4 and N2O in Annex-I countries to the UNFCCC”, *OECD Statistics Working Papers*, No. 2018/11, OECD Publishing, Paris, https://doi.org/10.1787/7d88dfdd-en. [7]

Gütschow, J. et al. (2016), “The PRIMAP-hist national historical emissions time series”, *Earth Syst. Sci. Data*, Vol. 8, pp. 571-603, https://doi.org/10.5194/essd-8-571-2016. [15]

Gütschow, J. and M. Pflüger (2023), *The PRIMAP-hist national historical emissions time series (1750-2022) v2.5. Zenodo. https://doi.org/10.5281/zenodo.10006301*. [14]

Hefner, M. and G. Marland (2023), “Global, Regional, and National Fossil-Fuel CO2 Emissions: 1751-2020 CDIAC-FF”, *Research Institute for Environment, Energy, and Economics, Appalachian State University*, https://energy.appstate.edu/research/work-areas/cdiac-appstate. [24]

IEA (2023), *Greenhouse Gas Emissions from Energy*, https://www.iea.org/data-and-statistics/data-product/greenhouse-gas-emissions-from-energy. [19]

IPCC (2022), *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [P.R. Shukla, J. Skea, A. Reisinger, R. Slade, R. Fradera, M. Pathak, A. Al Khourdajie,*, https://doi.org/10.1017/9781009157926.001. [1]

IPCC (2019), *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Calvo Buendia, E., Tanabe, K., Kranjc, A., Baasansuren, J., Fukuda, M., Ngarize S., Osako, A., Pyrozhenko, Y., Shermanau, P. and Federici, S. (eds)*, IPCC, Switzerland. [23]

IPCC (2006), *2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds).*, IGES, Japan. [12]

IPCC ((2019), "2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Calvo Buendia, E., Tanabe, K., Kranjc, A., Baasansuren, J., Fukuda, M., Ngarize S., Osako, A., Pyrozhenko, Y., Shermanau, P. and Federici, S. (eds)", *IPCC*, Vol. 4. [25]

Janssens-Maenhout, G. (2019), *EDGAR v4.3.2 Global Atlas of the three major Greenhouse Gas Emissions for the period 1970-2012*, https://doi.org/doi:10.5194/essd-11-959-2019. [18]

Joint Research Centre (JRC), E. (2023), *Emissions Database for Global Atmospheric Research) Community GHG Database (EDGAR) version 8.0*. [17]

OECD (2023), "Air Emission Accounts - OECD Estimates", *OECD.Stat*, https://stats.oecd.org/Index.aspx?DataSetCode=OECD-AEA (accessed on 12 October 2023). [8]

OECD (2023), *Environment at a Glance Indicators*, OECD Publishing, Paris, https://doi.org/10.1787/ac4b8b89-en. [3]

OECD (2023), "Greenhouse gas emissions by source", *OECD.Stat*, https://stats.oecd.org/Index.aspx?DataSetCode=AIR_GHG (accessed on 12 October 2023). [2]

OECD (2021), *Carbon Dioxide Emissions embodied in International Tade*, http://oe.cd/io-co2 (accessed on 29 September 2023). [10]

OECD (2012), *OECD Environmental Outlook to 2050: The Consequences of Inaction*, OECD Publishing, Paris, https://doi.org/10.1787/9789264122246-en. [6]

Pizarro, R. et al. (2024), *GHG Emission Trends and Targets (GETT) – Harmonised quantification methodology and indicators*, https://doi.org/10.1787/decef216-en. [4]

Sakata, S., A. Aklilu and R. Pizarro (2024), *Greenhouse Gas Emission Data: Concepts and data availability*, https://doi.org/10.1787/b3e6c074-en. [11]

UNFCCC (2023), *GHG data from UNFCCC*, https://unfccc.int/topics/mitigation/resources/registry-and-data/ghg-data-from-unfccc (accessed on January 2023). [13]

Yamano, N. and J. Guilhoto (2020), "CO2 emissions embodied in international trade and domestic final demand: Methodology and results using the OECD Inter-Country Input-Output Database"*, OECD Science, Technology and Industry Working Papers*, No. 2020/11, OECD Publishing, Paris, https://doi.org/10.1787/8f2963b8-en. [9]

Annex A. Additional comparisons

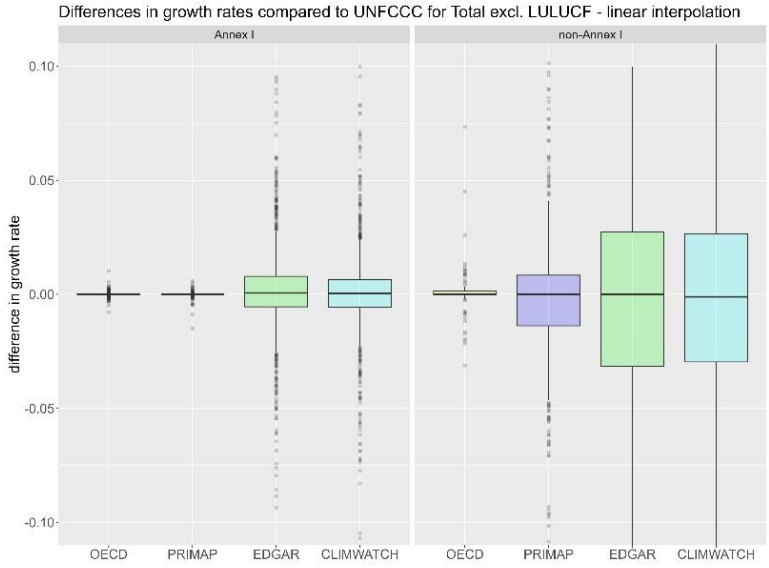
This Annex presents additional comparisons of the three unofficial datasets on GHG inventories. Availability of granular data on emission categories is shown by groups of countries, and the graphical analysis conducted for emission levels in 3.2 Alignment of emissions data is also done by IPCC categories and gases. Finally, data for selected countries are presented for illustration purposes.

A.1 Interpolated official data

UNFCC data for several non-Annex I countries are only available for a handful of years. For example, Figure C.1 in (Sakata, Aklilu and Pizarro, 2024[11]) shows that countries such as China and India have about four to six years of values, and they are available at irregular intervals. Irregular data gaps inevitably impact the comparability of trends across datasets done in the paper. Figure A.1 explores how unofficial datasets would compare against UNFCCC data where data gaps have been linearly interpolated.

Figure A.1. PRIMAP also aligns with linearly interpolated official data.

Emissions from total CO_2, CH_4, and N_2O, for all IPCC categories excluding LULUCF.

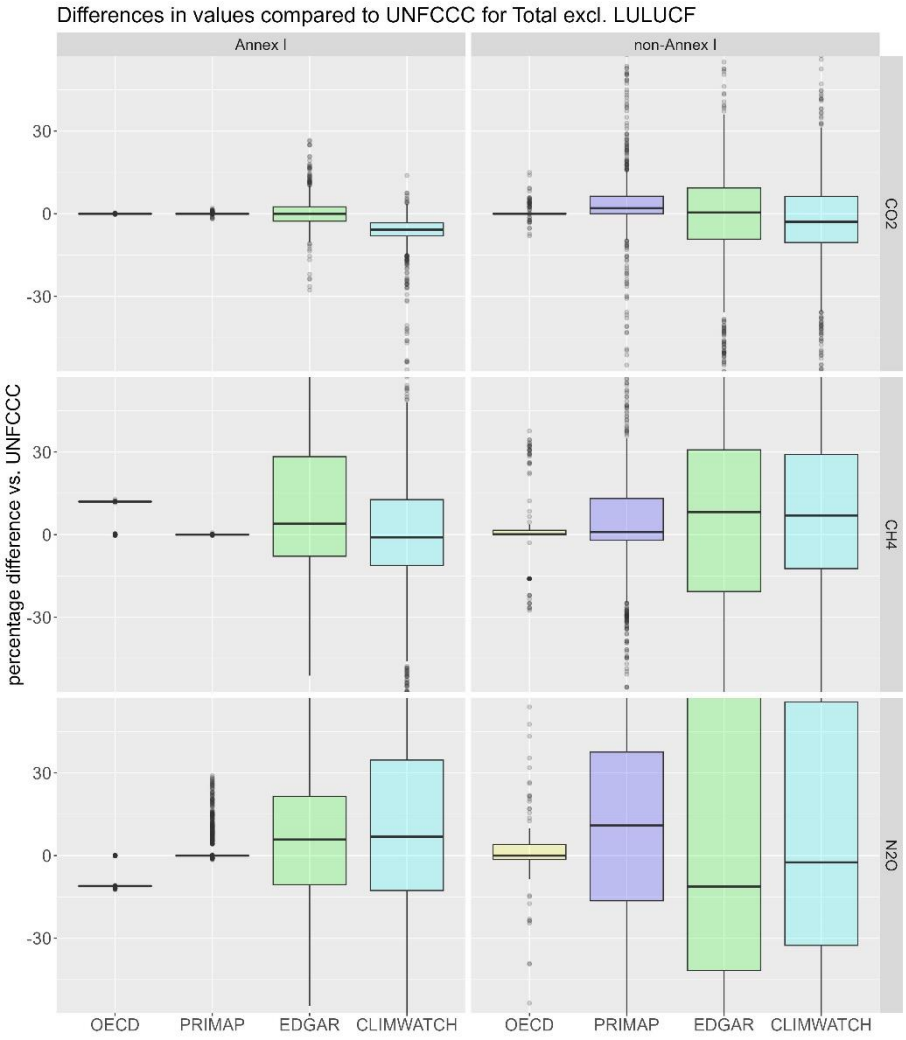


Note: See note on Figure 3.1 for details on the interpretation of boxplots. Extrapolations (i.e., estimation of missing values before (after) the first (last) available observation) are not included in the comparison.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

A.2 Emissions by gas

Figure A.2. PRIMAP emission levels align best with official data for each CO_2, CH_4 and N_2O gas.

Differences in emission levels of unofficial data relative to UNFCCC, by gas, in percentage.



Note: See note on Figure 3.1 for details on the interpretation of boxplots.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

A.3 Emission levels for categories

Figure A.3. Availability of granular data for Annex-I countries

Percentage of available (non-missing) values, 1990-2022, 43 Annex-I countries for CO_2, CH_4, and N_2O

2006 IPCC source categories	UNFCCC	OECD ENV	PRIMAP	EDGAR	ClimWatch
Total excluding LULUCF	94%	97%	100%	95%	92%
1. Energy	94%	97%	100%	95%	92%
A. Fuel combustion (sectoral approach)	94%	97%	100%	95%	90%
3rd level	94%	92%	0%	89%	90%
4th level	94%	0%	0%	62%	0%
B. Fugitive emissions from fuels	94%	94%	100%	89%	92%
3rd level	94%	0%	100%	85%	0%
C. CO_2 transport and storage	94%	6%	100%	0%	0%
2. Industrial processes and product use	94%	97%	100%	95%	92%
A. Mineral industry	94%	100%	95%	0%	0%
3rd level	94%	0%	0%	73%	0%
B. Chemical industry	94%	0%	100%	90%	0%
C. Metal industry	94%	0%	100%	88%	0%
D. Non-energy products from fuels and solvent use	94%	0%	100%	95%	0%
E. Electronic Industry	n/a	n/a	n/a	n/a	n/a
F. Product uses as ODS substitutes	n/a	n/a	n/a	n/a	n/a
G. Other product manufacture and use	94%	0%	100%	95%	0%
H. Other	90%	0%	67%	0%	0%
3. Agriculture	94%	95%	100%	95%	92%
2nd level	93%	0%	20%	57%	0%
4. Land use, land-use change and forestry	94%	97%	100%	0%	92%
2nd level	94%	0%	0%	0%	0%
5. Waste	94%	97%	100%	95%	92%
2nd level	94%	0%	0%	75%	0%
6. Other	87%	9%	100%	95%	0%

Note: n/a indicates that the category is not applicable to any of the three GHGs. If data is available for CO_2, CH_4 or N_2O, the observation is considered available. Zero percent availability indicates that the source does not report the category. Incompleteness might stem from categories, years or countries being unavailable, for example, UNFCCC does not reach more than 94% as the 2022 is unavailable at the time of writing of this paper.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.4. Availability of granular data for non-Annex-I countries

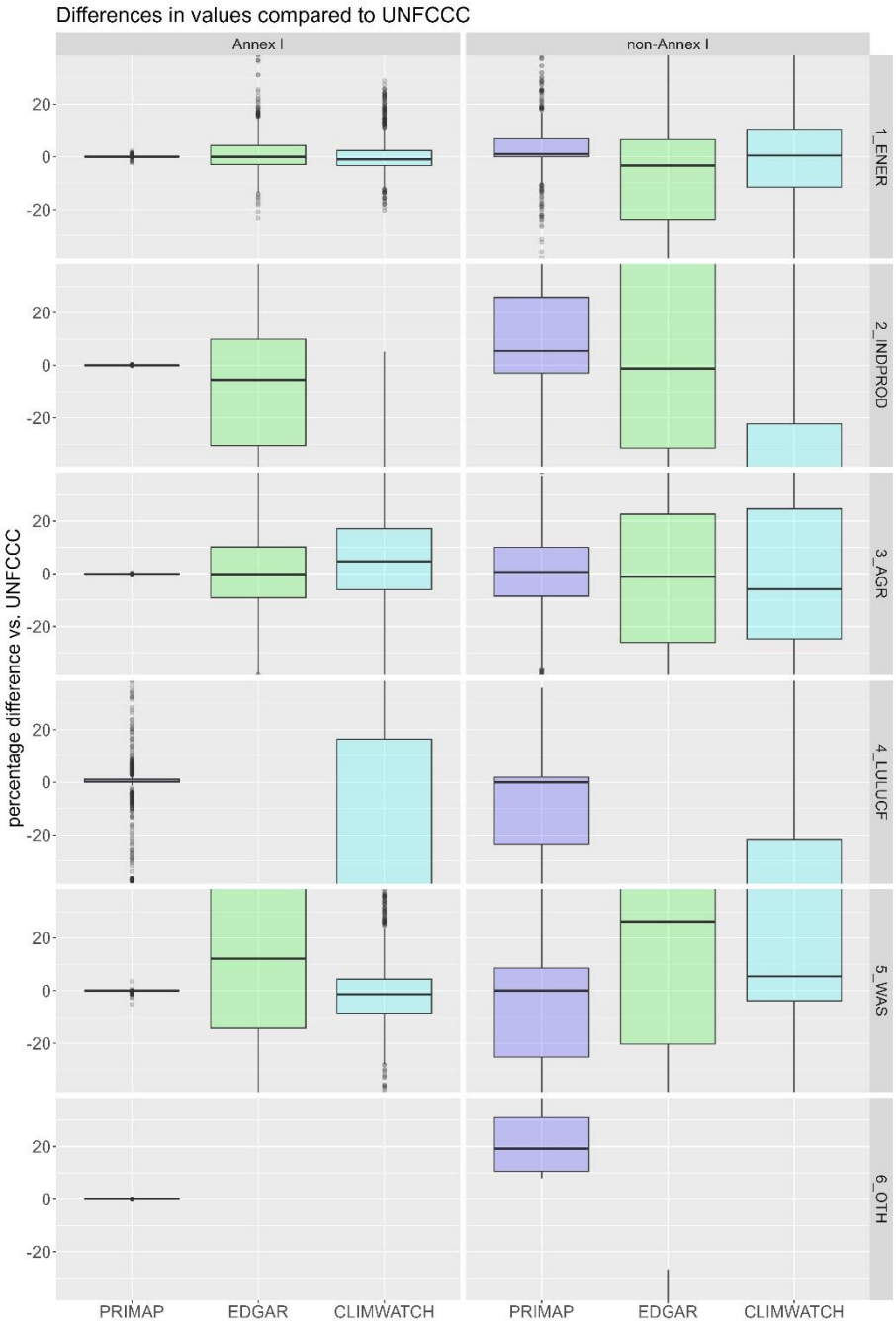
Percentage of available (non-missing) values, 1990-2022, 155 non-Annex I countries for CO_2, CH_4, and N_2O.

2006 IPCC source categories	UNFCCC	OECD ENV	PRIMAP	EDGAR	ClimWatch
Total excluding LULUCF	14%	5%	88%	96%	76%
1. Energy	14%	5%	87%	88%	76%
A. Fuel combustion (sectoral approach)	13%	4%	87%	88%	75%
3rd level	10%	4%	0%	81%	75%
4th level	0%	0%	0%	58%	0%
B. Fugitive emissions from fuels	8%	2%	85%	76%	76%
3rd level	6%	0%	80%	49%	0%
C. CO_2 transport and storage	0%	0%	12%	0%	0%
2. Industrial processes and product use	11%	5%	87%	96%	76%
A. Mineral industry	10%	0%	87%	89%	0%
3rd level	0%	0%	0%	48%	0%
B. Chemical industry	5%	0%	42%	29%	0%
C. Metal industry	7%	0%	49%	37%	0%
D. Non-energy products from fuels and solvent use	1%	0%	83%	96%	0%
E. Electronic Industry	n/a	n/a	n/a	n/a	n/a
F. Product uses as ODS substitutes	n/a	n/a	n/a	n/a	n/a
G. Other product manufacture and use	1%	0%	84%	92%	0%
H. Other	1%	0%	17%	0%	0%
3. Agriculture	13%	5%	84%	87%	76%
2nd level	6%	0%	19%	52%	0%
4. Land use, land-use change and forestry	13%	5%	87%	0%	76%
2nd level	5%	0%	0%	0%	0%
5. Waste	13%	5%	87%	96%	76%
2nd level	5%	0%	0%	61%	0%
6. Other	0%	0%	84%	91%	0%

Note: If data is available for CO_2, CH_4 or N_20, the observation is considered available. Incompleteness might stem from categories, years or countries being unavailable; for example: OECD data is not collected at the 4th level category disaggregation. N/A indicates that the category is not applicable to any of the three GHGs.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.5. Differences in emission levels are higher for the first level of IPCC categories.

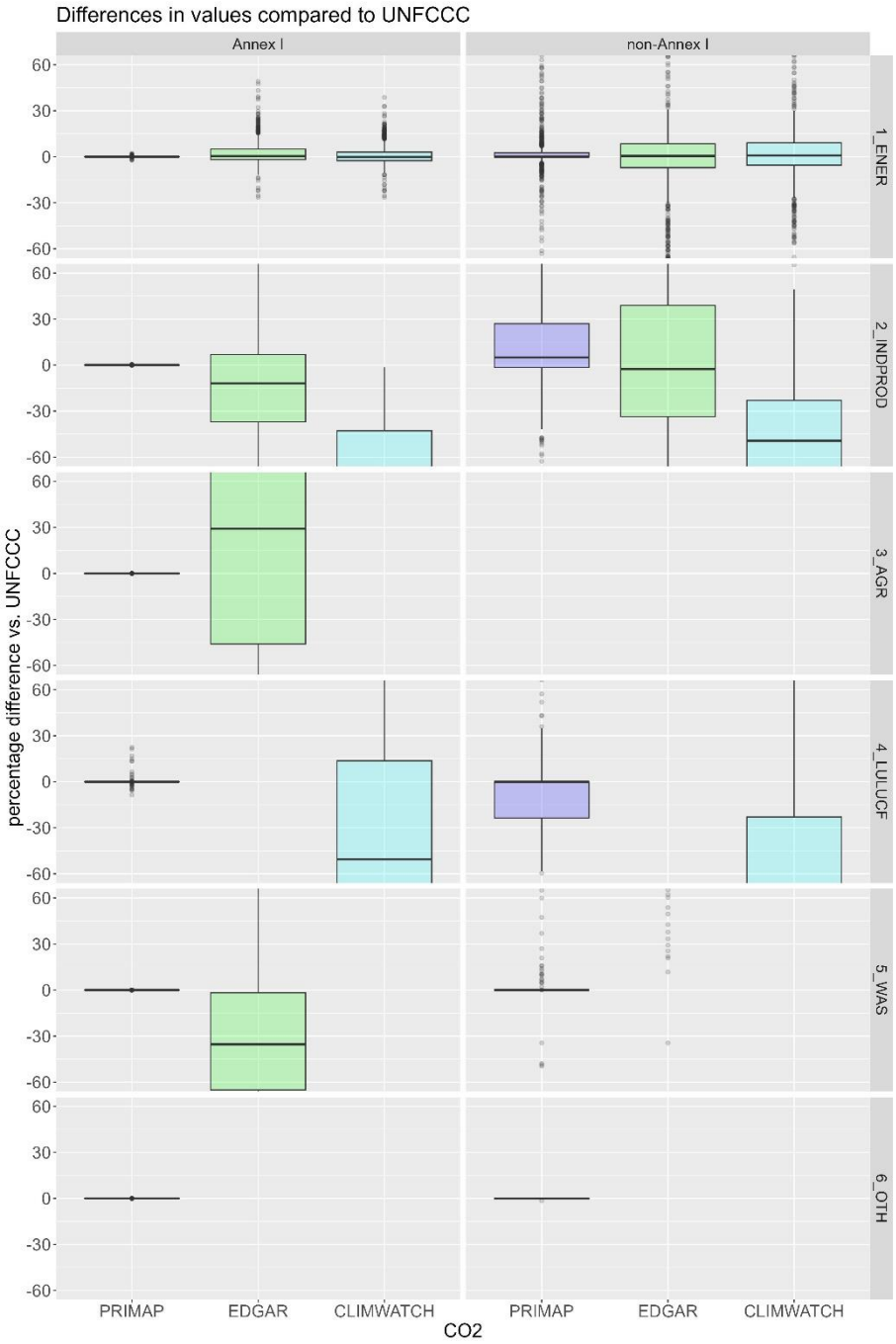
Differences in emission levels (sum of CO_2, CH_4, and N_2O) of unofficial data relative to UNFCCC, by IPCC category, in percentage.



Note: See note on Figure 3.1 for details on the interpretation of boxplots.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.6. Discrepancies in CO_2 emission levels are lowest for energy.

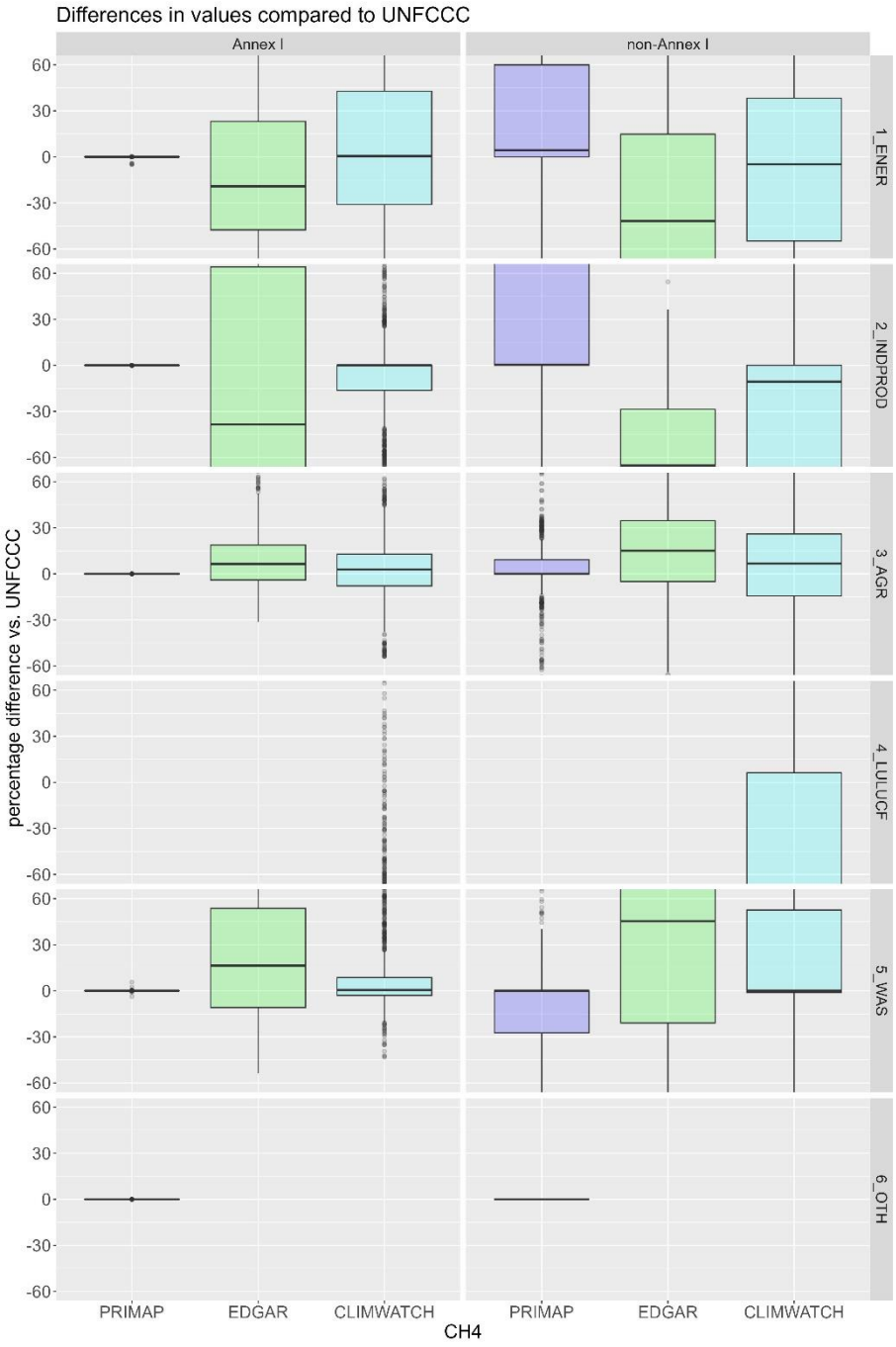
Differences in CO_2 levels of unofficial data relative to UNFCCC, by IPCC category, in percentage.



Note: See note on Figure 3.1 for details on the interpretation of boxplots.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.7. Discrepancies in CH_4 emission levels are lowest for agriculture.

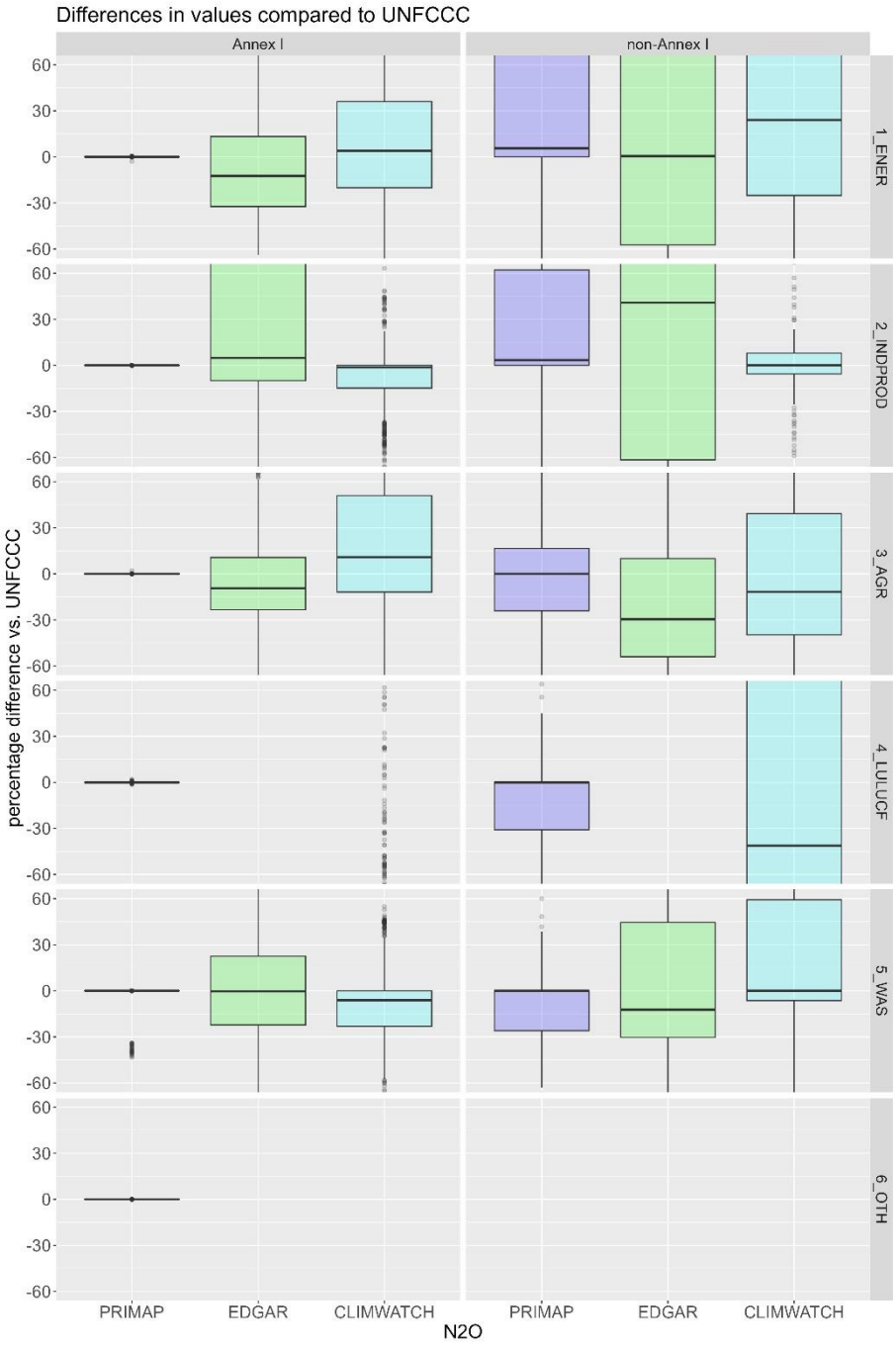
Differences in CH_4 levels of unofficial data relative to UNFCCC, by IPCC category, in percentage.



Note: See note on Figure 3.1 for details on the interpretation of boxplots.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.8. Discrepancies for N_2O are lowest for agriculture and waste.

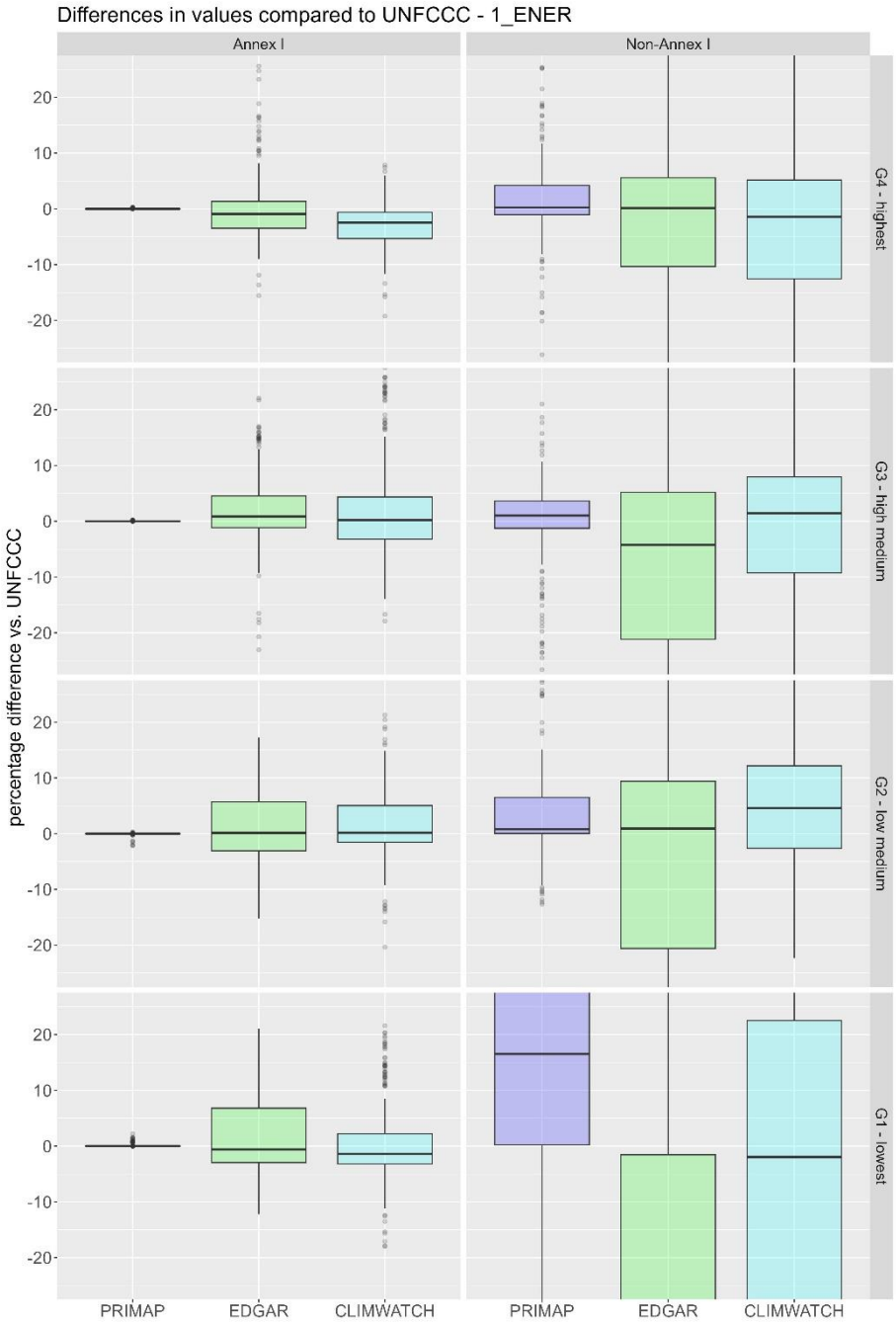
Differences in N_2O levels of unofficial data relative to UNFCCC, by IPCC category, in percentage.



Note: See note on Figure 3.1 for details on the interpretation of boxplots.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.9. Unofficial data on emissions from energy align best with official data for countries with a higher relative size of the category.

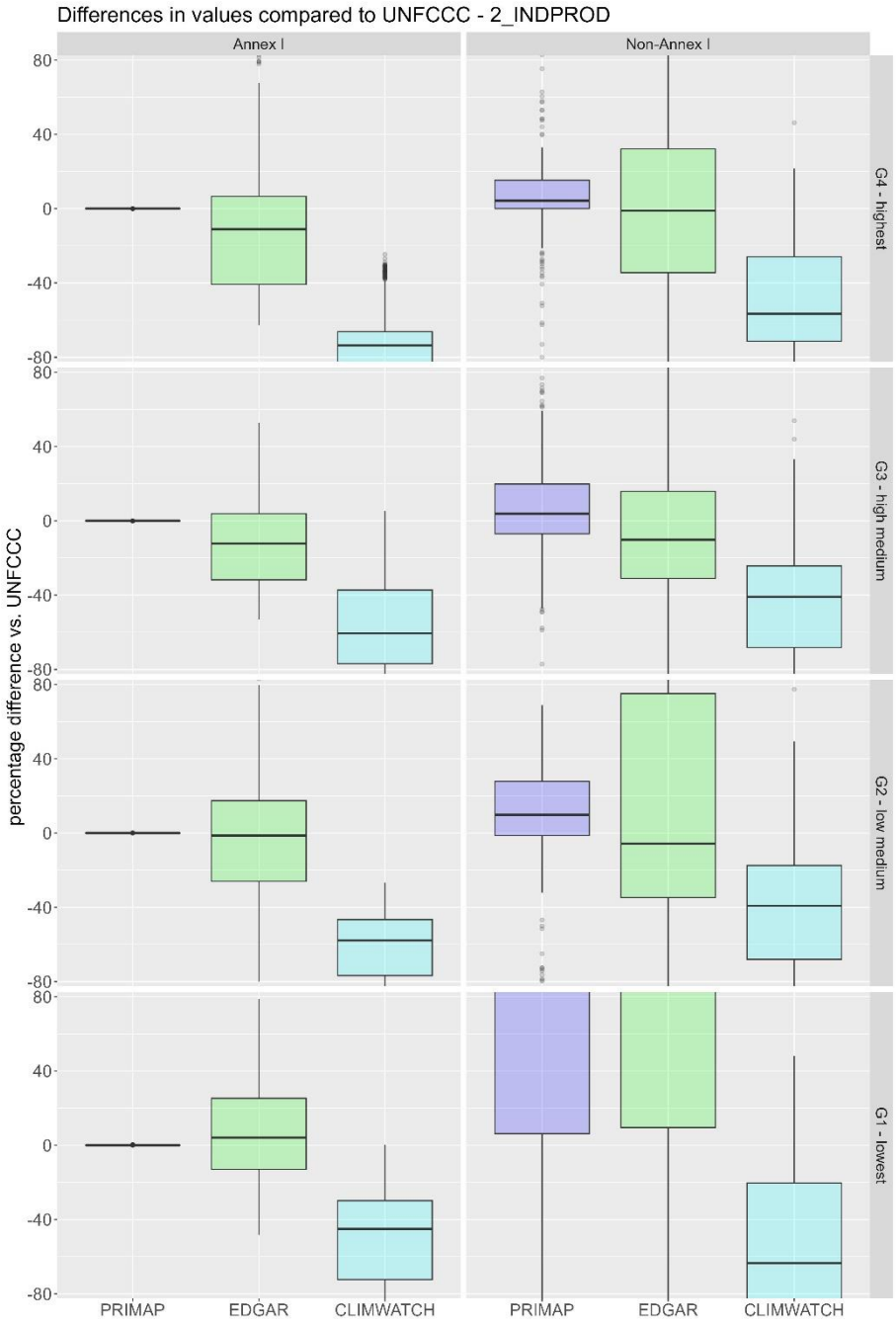
Differences in emission levels of the energy category in percentage, unofficial data relative to UNFCCC. Countries are grouped based on the relative size of energy in total emissions.



Note: See note on Figure 3.1 for details on the interpretation of boxplots. Countries are grouped into four baskets according to the average share of the category in total emissions, based on UNFCCC data. The 4th group (G4) includes the countries where energy represents the highest share in total emissions relative to the sample (i.e., countries above the third quartile of the distribution), while the 1st group (G1) includes the countries with the lowest share of energy emissions (i.e., countries below the first quartile of the distribution).
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.10. Unofficial data on emissions from industrial processes and product use align best with official data for countries with a higher relative size of the category.

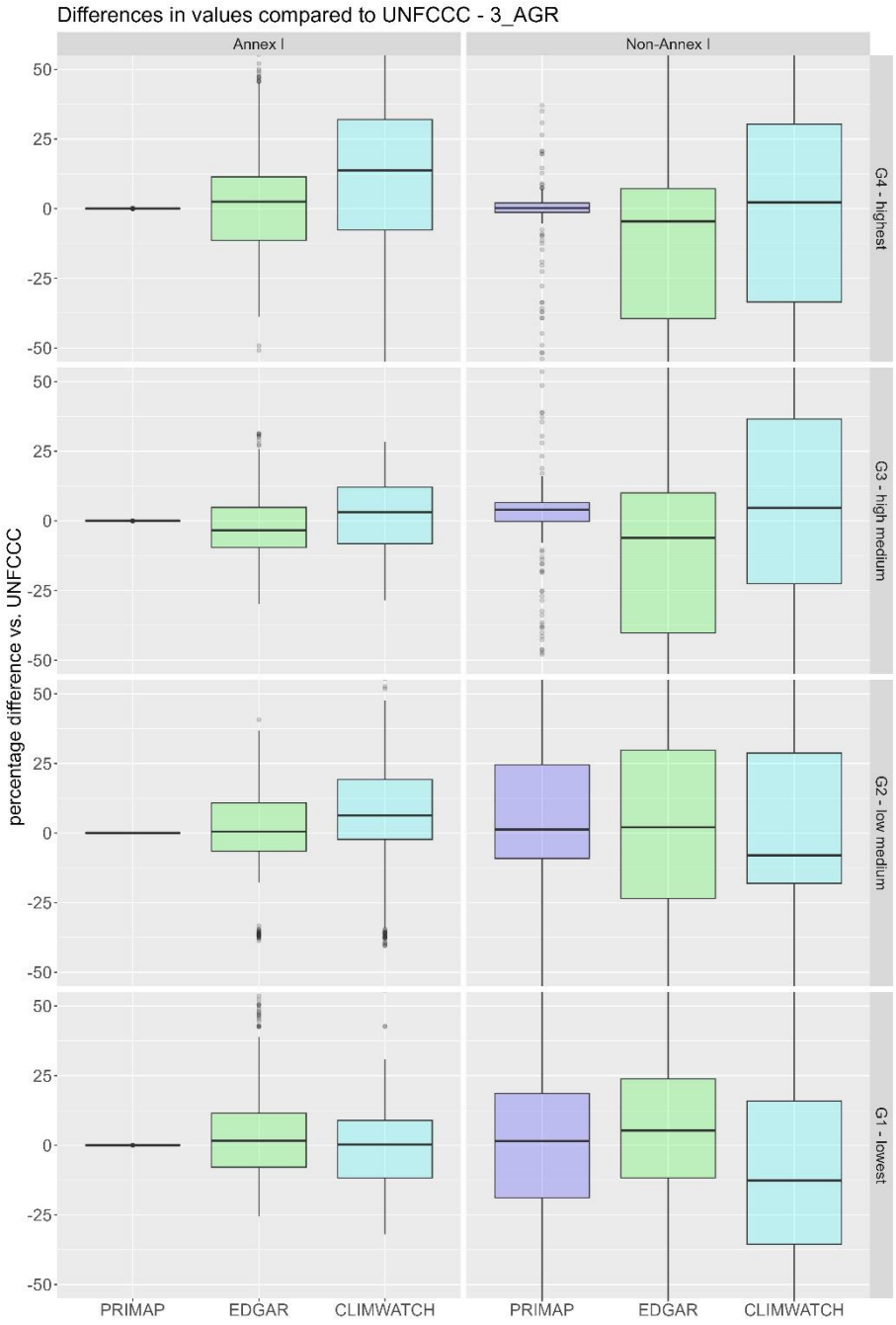
Differences in emission levels of the industrial processes and product use category in percentage, unofficial data relative to UNFCCC. Countries are grouped based on the relative size of the category in total emissions.



Note: See note on Figure 3.1 for details on the interpretation of boxplots. Countries are grouped into four baskets according to the average share of the category in total emissions, based on UNFCCC data. The 4th group (G4) includes the countries where industrial emissions represent the highest share in total emissions relative to the sample (i.e., countries above the third quartile of the distribution), while the 1st group (G1) includes the countries with the lowest share of industrial emissions (i.e., countries below the first quartile of the distribution).
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.11. PRIMAP data on emissions from agriculture align best with official data for countries with a higher relative size of the category.

Differences in emission levels of the agriculture category in percentage, unofficial data relative to UNFCCC. Countries are grouped based on the relative size of the category in total emissions.

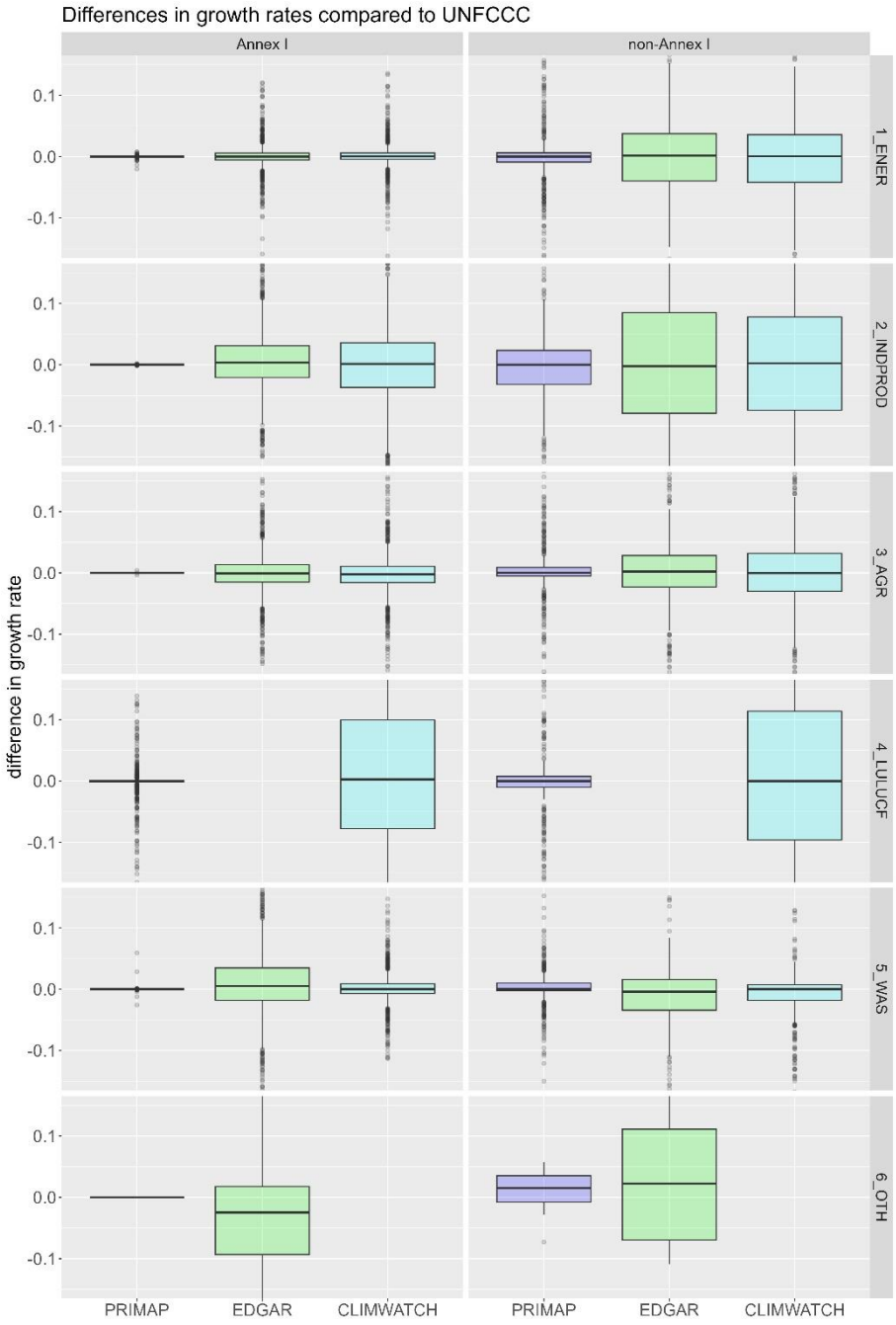


Note: EDGAR and CLIMWATCH show similar discrepancies with respect to official data across the four country groups. See note on Figure 3.1 for details on the interpretation of boxplots. Countries are grouped into four baskets according to the average share of the category in total emissions, based on UNFCCC data. The 4th group (G4) includes the countries where agriculture represents the highest share in total emissions relative to the sample (i.e., countries above the third quartile of the distribution), while the 1st group (G1) includes the countries with the lowest share of agriculture emissions (i.e., countries below the first quartile of the distribution).
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

A.4 Trends for categories

Figure A.12. PRIMAP emission trends align best with UNFCCC across all first level IPCC categories.

Differences in the growth rates of unofficial data relative to UNFCCC, by IPCC category, in percentage points.

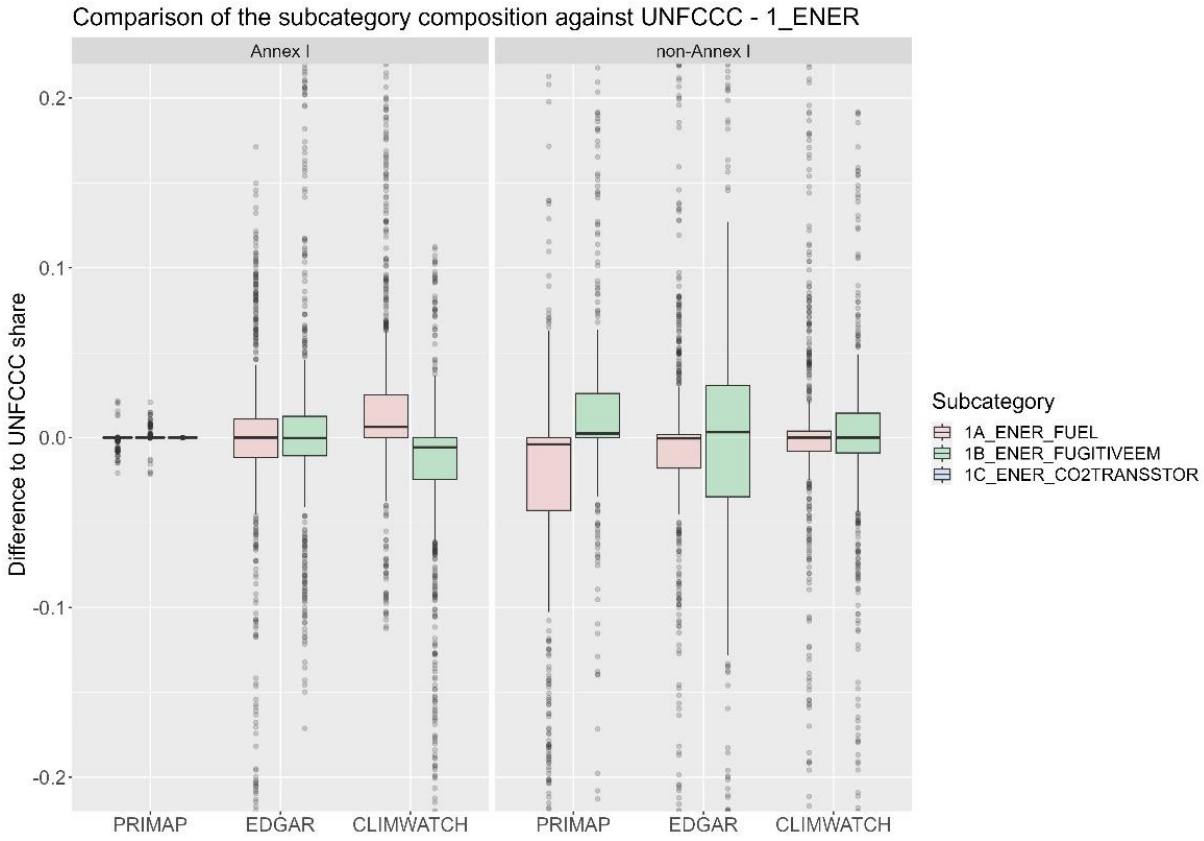


Note: See note on Figure 3.1 for details on the interpretation of boxplots.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

A.5 Composition of categories

Figure A.13. Energy subcategories show discrepancies across all datasets for Non-Annex I countries.

Differences in the composition of the energy category unofficial data relative to UNFCCC in percentage points.

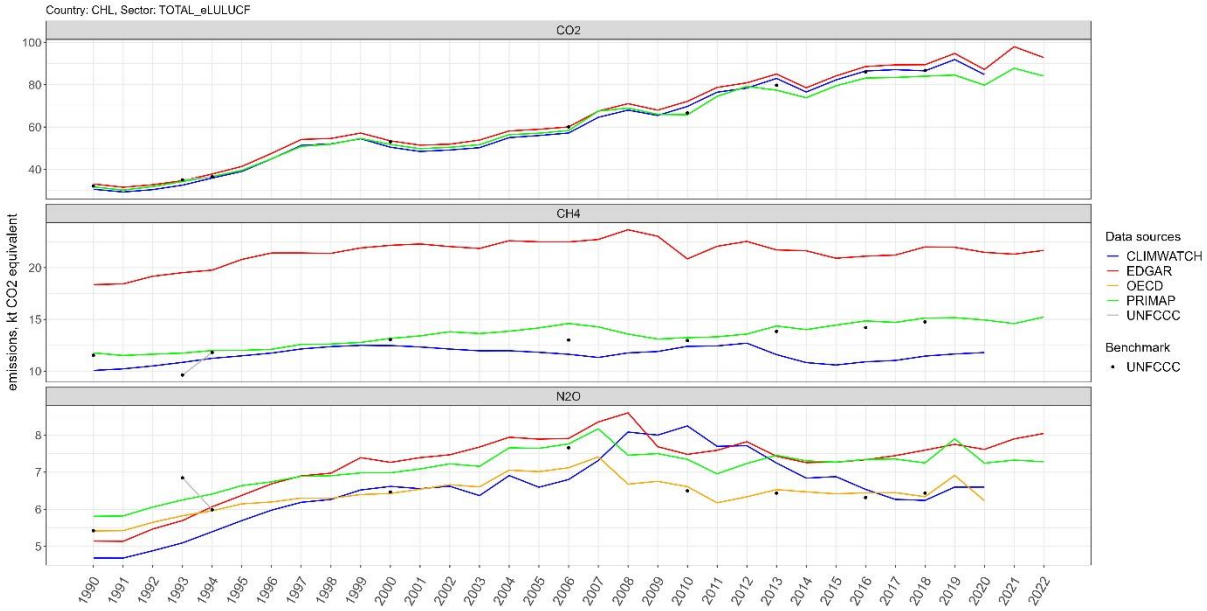


Note: See note on Figure 3.1 for details on the interpretation of boxplots.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

A.6 Selected country examples

Figure A.14. Chile's GHG emission inventories

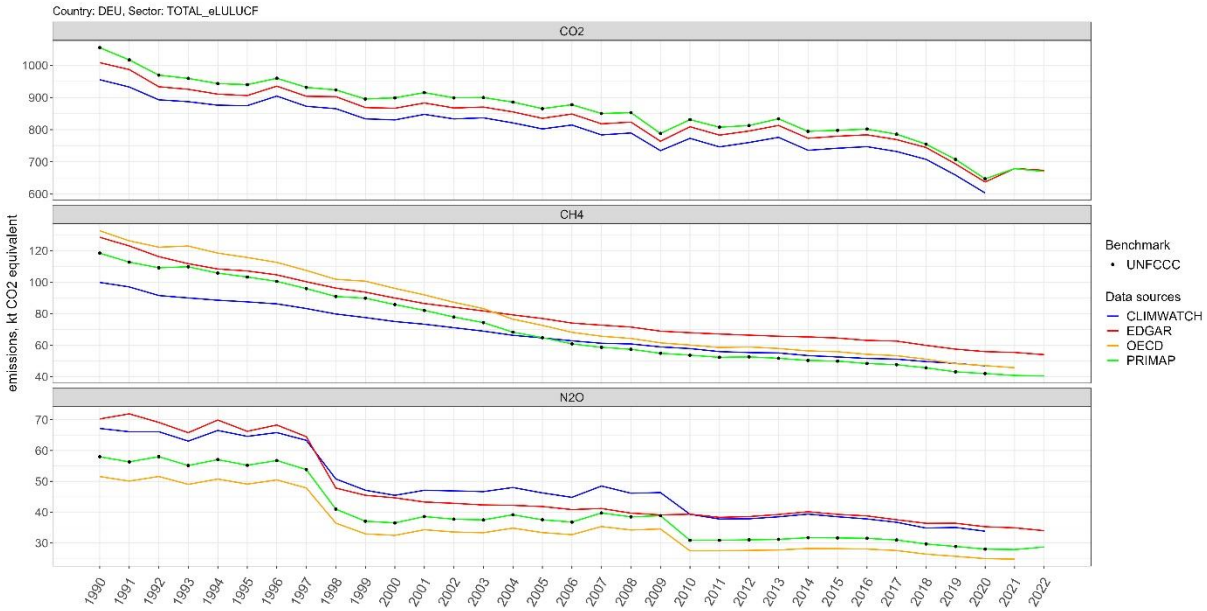
Emissions by gas for all categories excluding LULUCF.



Note: Emission levels in $ktCO_2$-equivalent, expressed in Global Warming Potential of Assessment Report 4.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.15. Germany's GHG emission inventories

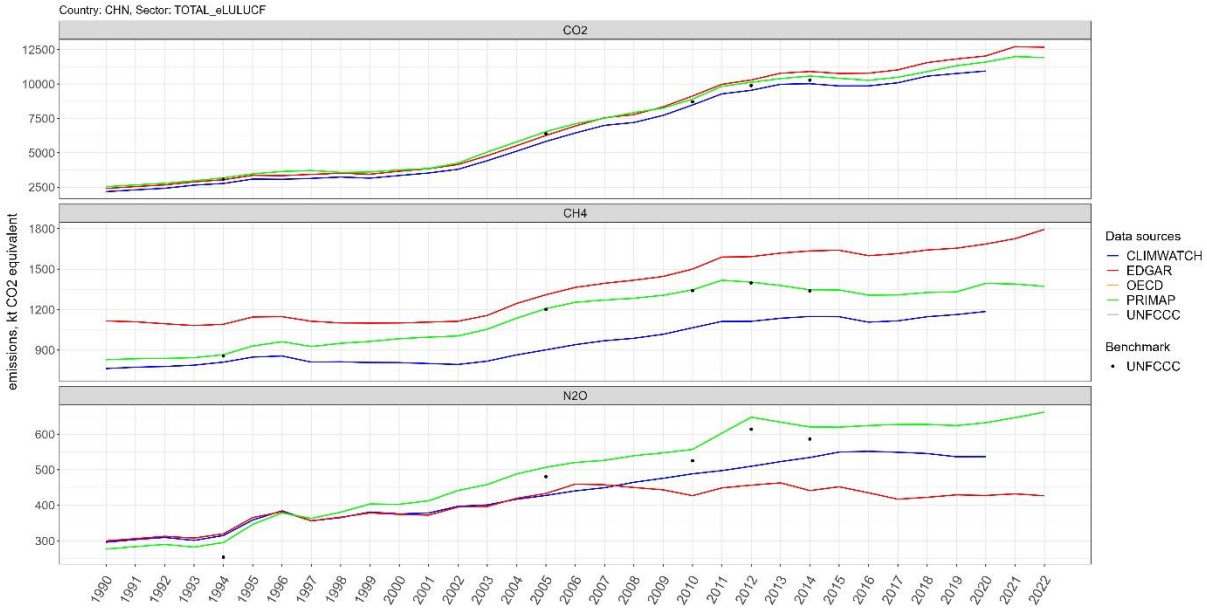
Emissions by gas for all categories excluding LULUCF.



Note: Emission levels in $ktCO_2$-equivalent, expressed in Global Warming Potential of Assessment Report 4.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.16. China's GHG emission inventories

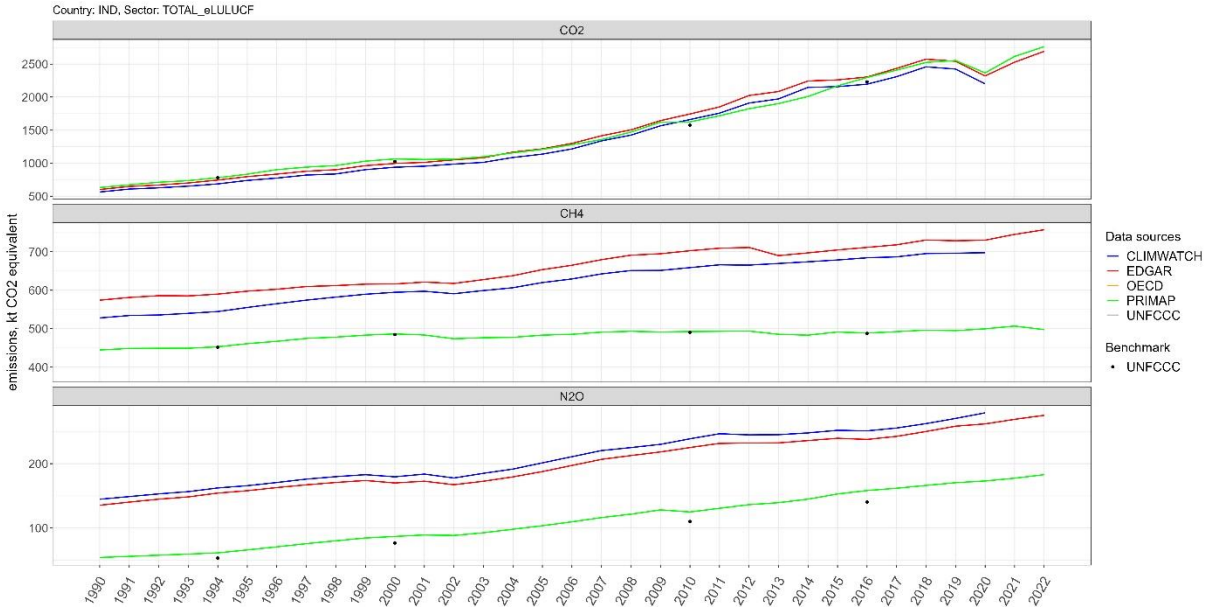
Total emissions by gas for all categories excluding LULUCF.



Note: Emission levels in $ktCO_2$-equivalent, expressed in Global Warming Potential of Assessment Report 4. Inventories are not reported to OECD.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Figure A.17. India's GHG emission inventories

Total emissions by gas for all categories excluding LULUCF.

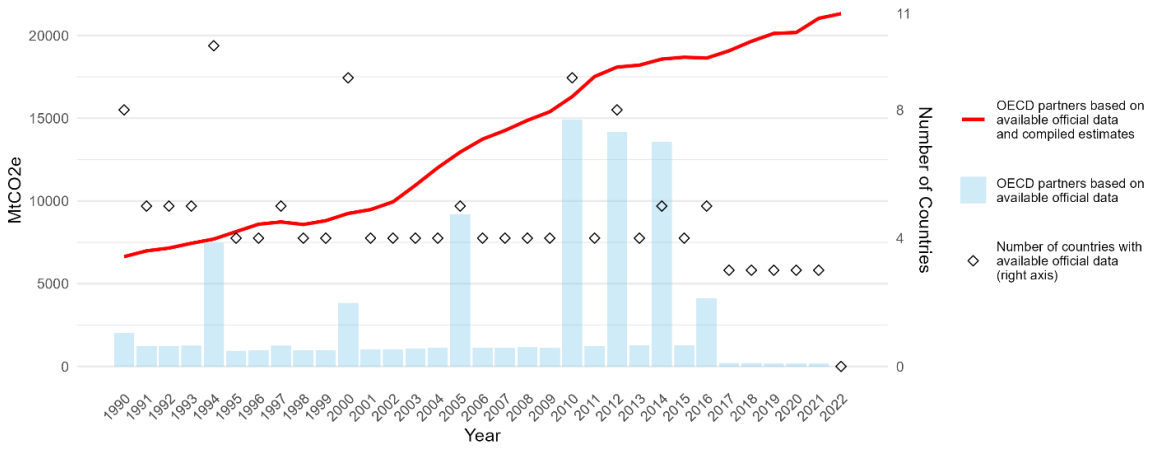


Note: Emission levels in $ktCO_2$-equivalent, expressed in Global Warming Potential of Assessment Report 4.
Source: EDGAR v8, PRIMAP v2.5, CLIMWATCH vDec2023, OECD data extracted December 2023, UNFCCC data extracted January 2024.

Annex B. Preliminary results

Figure B.1. The strategy allows to fill important gaps in official data for OECD partner countries.

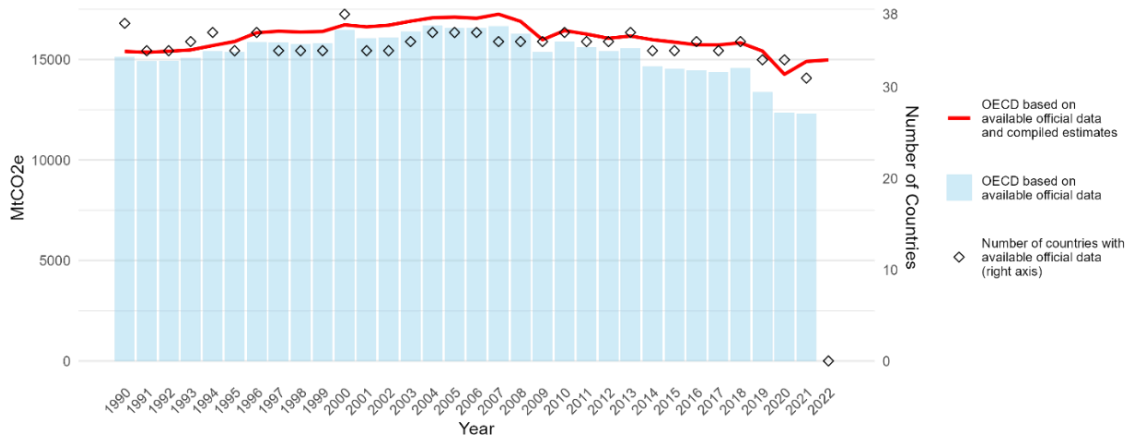
GHG emissions for all categories excluding LULUCF, Aggregate over available data for OECD partner countries.



Note: The partial sum over available official data (bars) includes only UNFCCC submissions. GHG emissions are expressed in million tonnes of CO_2-equivalent, in Global Warming Potential of Assessment Report 4. OECD partners include Argentina, Brazil, Bulgaria, China, Croatia, India, Indonesia, Peru, Romania, Saudi Arabia, South Africa.
Source: OECD estimations based on OECD data extracted December 2023, UNFCCC data extracted January 2024, and PRIMAP v2.5.

Figure B.2. The strategy allows to fill gaps for OECD countries.

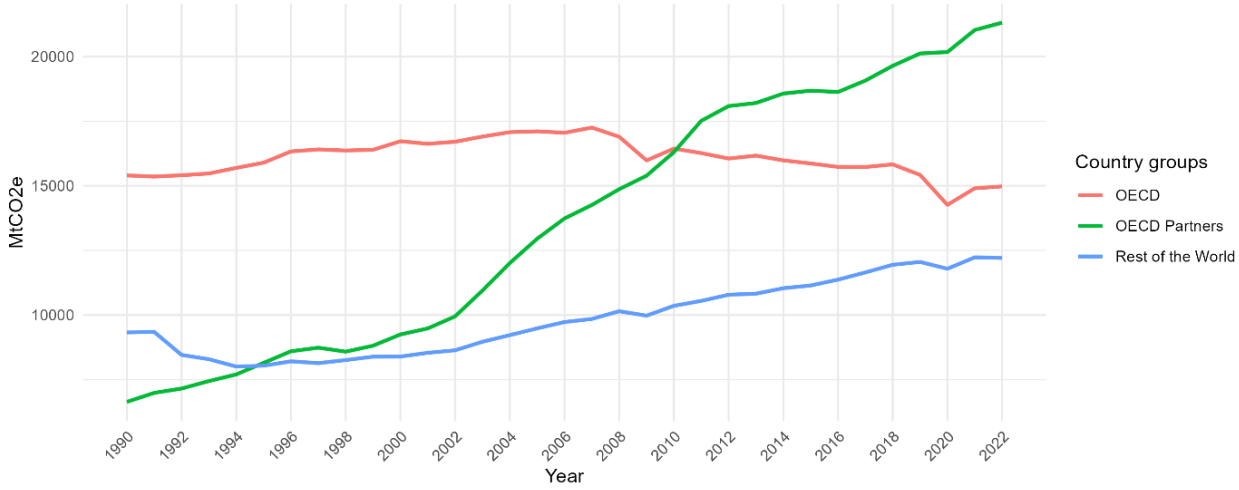
Compiled estimates of GHG emissions for all categories excluding LULUCF, Aggregate over available data.



Note: The partial sum over available official data (bars) includes only UNFCCC submissions. GHG emissions are expressed in million tonnes of CO_2-equivalent, in Global Warming Potential of Assessment Report 4.
Source: OECD estimations based on OECD data extracted December 2023, UNFCCC data extracted January 2024, and PRIMAP v2.5.

Figure B.3. Compiled GHG emission estimates by country groups.

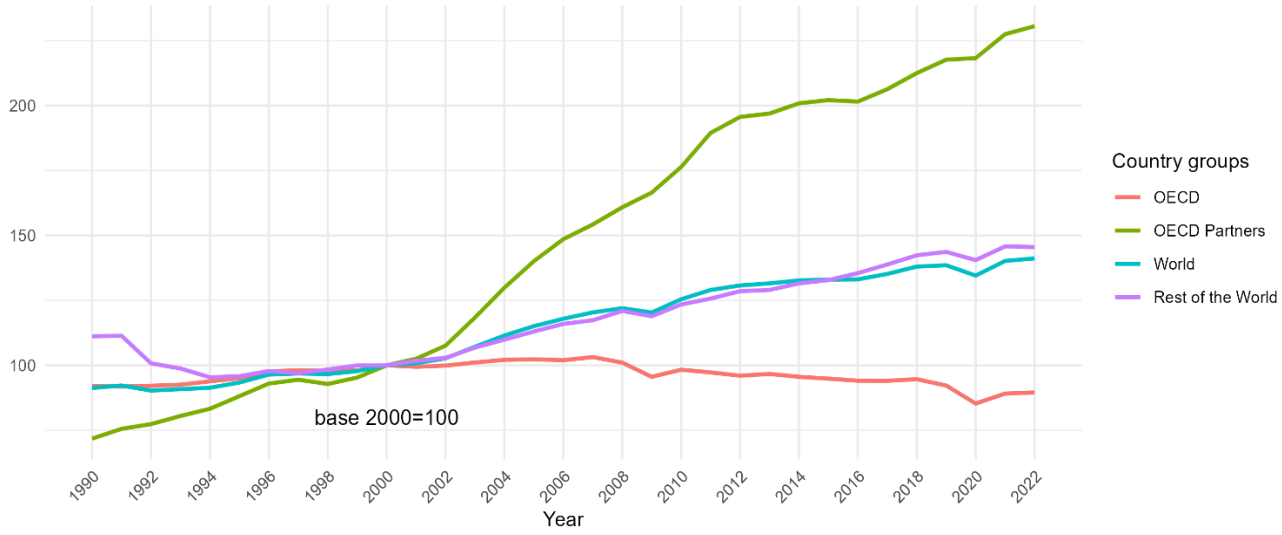
Compiled estimates of GHG emission levels for all categories excluding LULUCF, by country groups.



Note: GHG emissions are expressed in million tonnes of CO_2-equivalent, in Global Warming Potential of Assessment Report 4.
Source: OECD estimations based on OECD data extracted December 2023, UNFCCC data extracted January 2024, and PRIMAP v2.5.

Figure B.4. Trends in compiled GHG emission estimates by country groups.

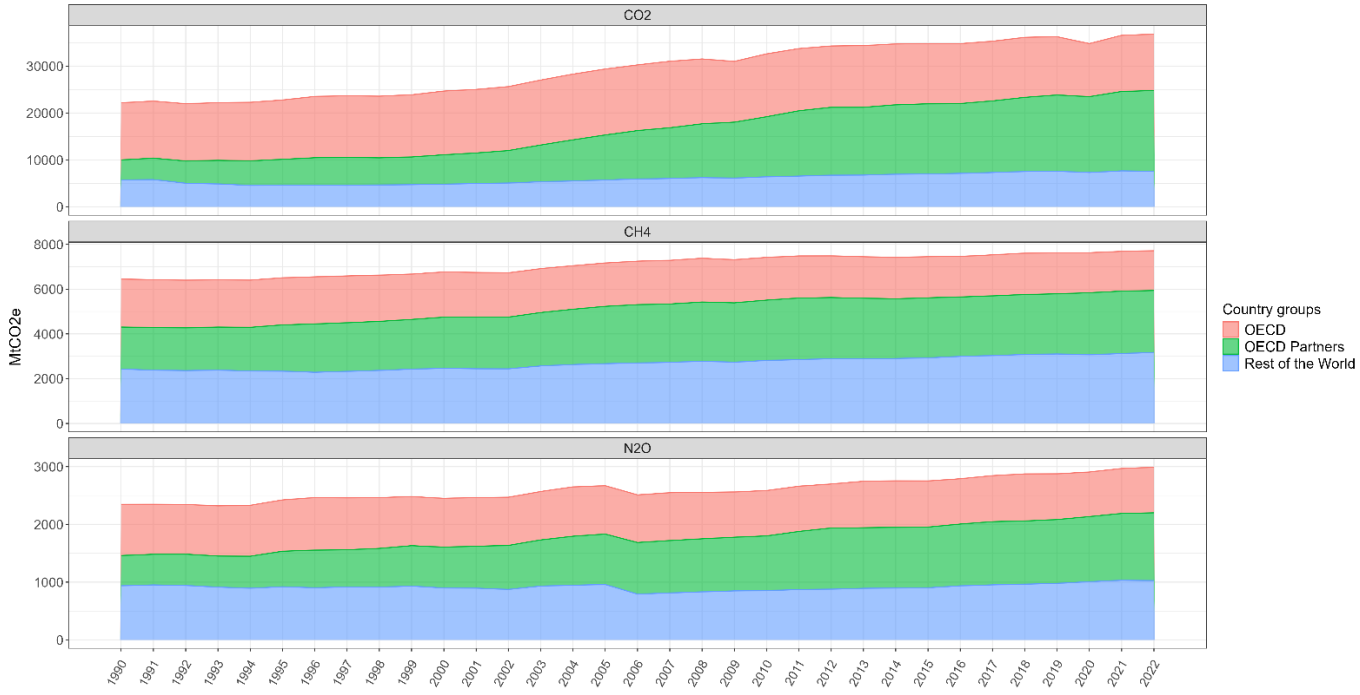
Compiled estimates of GHG emissions for all categories excluding LULUCF, index 2000=100.



Note: GHG emissions are expressed in million tonnes of CO_2-equivalent, in Global Warming Potential of Assessment Report 4.
Source: OECD estimations based on OECD data extracted December 2023, UNFCCC data extracted January 2024, and PRIMAP v2.5.

Figure B.5. CO_2, CH_4 and N_2O emissions by country groups

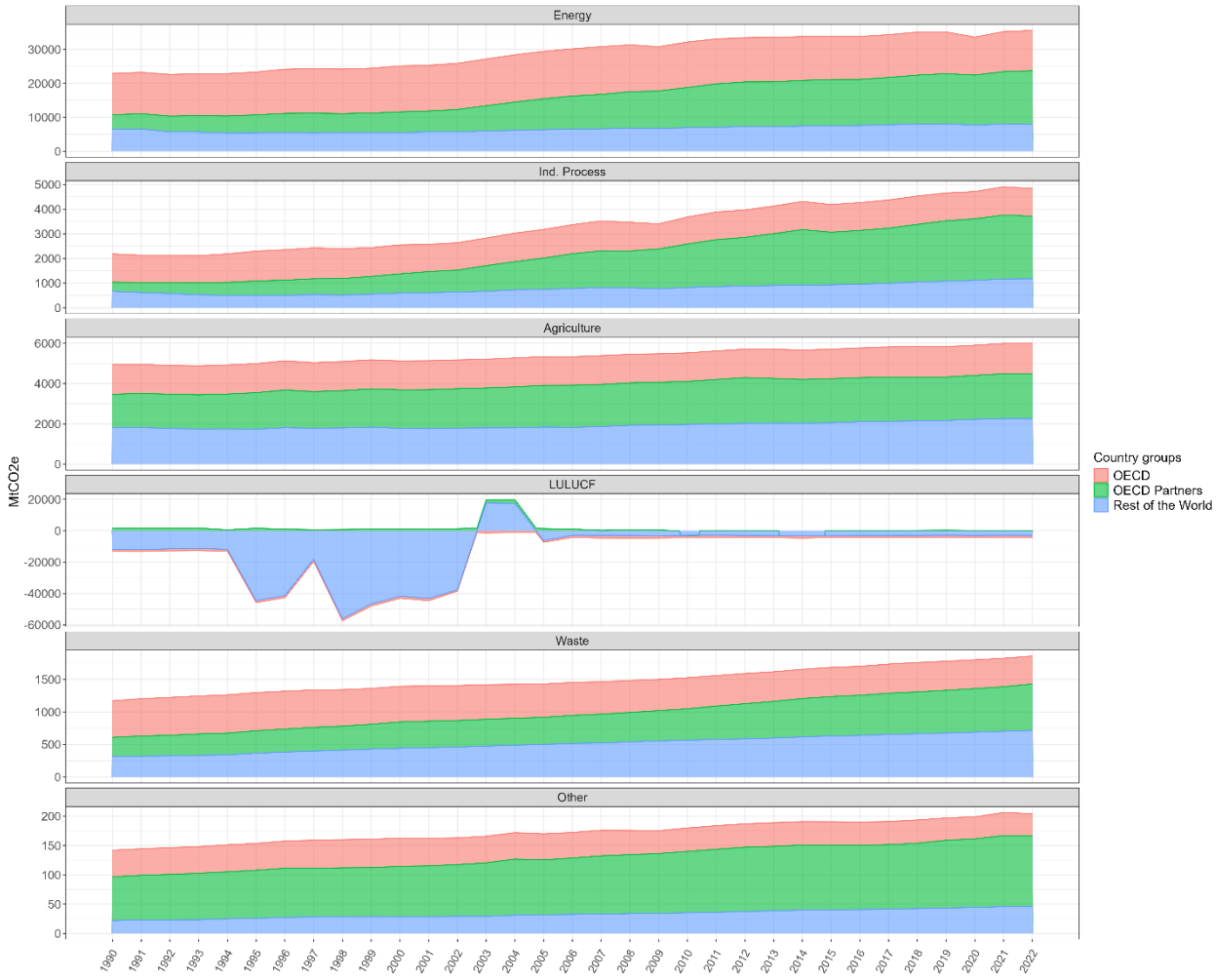
Compiled estimates of CO_2, CH_4 and N_2O emissions for all categories excluding LULUCF.



Note: GHG emissions are expressed in million tonnes of CO_2-equivalent, in Global Warming Potential of Assessment Report 4.
Source: OECD estimations based on OECD data extracted December 2023, UNFCCC data extracted January 2024, and PRIMAP v2.5.

Figure B.6. GHG emissions by IPCC categories and country groups

GHG emissions by categories and country groups



Note: GHG emissions are expressed in million tonnes of CO_2-equivalent, in Global Warming Potential of Assessment Report 4.
Source: OECD estimations based on OECD data extracted December 2023, UNFCCC data extracted January 2024, and PRIMAP v2.5.