MAFAP Emissions Database Construction

Datasets and approaches

Renato Vargas Martin Cicowiez

Table of contents

Introduction

The Monitoring and Analyzing Food and Agricultural Policies (MAFAP) program seeks to establish country owned and sustainable systems to monitor, analyze, and reform food and agricultural policies to enable more effective, efficient and inclusive policy frameworks. Since policies are only enacted when they are properly funded, MAFAP tools are particularly geared towards helping countries—the agricultural public sector, in particular—optimize public budgets in the face of competing objectives, like economic growth, poverty reduction, better nutrition, or increasing yields, for example.

In the face of increasing vulnerability, climate risk mitigation and adaptation objectives have also become a part of the policy dialogue, as different international commitments (e.g. the Paris Agreement) present new demands and challenges to governments. To be able to provide objective functions within MAFAP tools, it has become necessary to extend its socioeconomic reach to include environmental objectives. As an extension of the MAFAP CGE modelling tool, the Global Emissions Database aims to provide a seamless integration of data on anthropogenic emissions of greenhouse gases and air pollution with structural analysis frameworks like Computable General Equilibrium and Input-Output models.

Here we describe the approach used to construct an anthropogenic emissions accounting satellite framework by economic activity and driver, using globally available datasets. We aim to maintain compatibility with the System of National Accounts (SNA) with which economic performance is measured (European Commission et al. 2009). Where possible, we adopt approaches suggested by previous efforts at harmonization between Greenhouse Gas Inventories and SEEA pilot implementations (Eurostat 2015; Flachenecker, Guidetti, and Pionnier 2018; Gutman et al. 2024).

Emissions from combustion are estimated at the economic activity level using the International Industrial Standard Classification—ISIC—(United Nations 2008) and driver, using the Tier 1 method of the Intergovernmental Panel on Climate Change—IPCC—(IPCC 1996, 2006) and energy use information from the Energy Statistics Database (UNSD 2023). Emissions from Industrial Processes and Waste come from EDGAR (Crippa et al. 2023) and are mapped to economic activities following a broadly tested global approach that facilitates use with CGE frameworks; i.e. that of the Global Trade Analysis Project—GTAP—(Aguiar et al. 2019; Chepeliev 2024). Emissions from Agriculture are sourced from FAOSTAT (FAO 2024) and mapped to ISIC economic activities. Individual drivers are left for the practitioner to map to their own structure (i.e. output by industries, endowment by industries, input use by industries, and consumption by households).

Methodology

Overview

As our goal is to leverage valuable emissions data from different sources, we have scripted data processing routines to ingest original datasets in tabular format (mainly Excel Files and Comma Separated Values) and manipulate them to conform to a single flat file structure, using the R programming language (R Core Team 2024)[1]. The original datasets, translated into the much smaller size RDS binary format, as well as the processing scripts, are included with the resulting emissions database repository[2]. The flow of data follows the diagram in Figure 1.

MAFAP Complementary Emissions Database Construction Process

Figure 1

[1]R was chosen for its information management and data compression facilities. However, the principles used can be easily implemented in Python, Stata, or any other language with similar affordances.

[2]The data is available at: https://github.com/renatovargas/gears/tree/master/emdb/data

Total Emissions from EDGAR

We start our methodological discussion with the sourcing of The Emissions Database for Global Atmospheric Research—EDGAR—(Crippa et al., 2023) datasets published by the European Commission because they are comprehensive, covering most of the greenhouse gas emissions inventories domains as shown in Table 1, belonging to 1) energy; 2) Industrial processes and product use; 3) Agriculture, forestry, and other land use; 4) Waste; and 5) Other (mainly indirect emissions from deposition of Nitrogen). This allows us to provide an overview of the coverage needed by our dataset. However, due to the form in which the data is presented in the EDGAR data and our database needs, we only take advantage of emissions data on industrial processes, waste, and "other" (which refers mainly to emissions from the atmospheric deposition of nitrogen). To avoid double counting, we replace emissions from the use of energy (Chapter 1A) with data from the UN Energy Database (UNSD 2023) and emissions from agriculture, forestry, and other land use with data from FAOSTAT (FAO 2024). For comparison purposes, it is important to note that large scale biomass burning with Savannah burning, forest fires, and sources and sinks from land-use, land-use change and forestry (LULUCF) are not covered in EDGAR, but are in FAOSTAT.

Table 1: EDGAR greenhouse gas inventory domain coverage

Code	IPCC Greenhouse Gas Inventory domain
1.A.1.a	Main Activity Electricity and Heat Production
1.A.1.bc	Petroleum Refining - Manufacture of Solid Fuels and Other Energy Industries
1.A.2	Manufacturing Industries and Construction
1.A.3.a	Civil Aviation
1.A.3.b_noRES	Road Transportation no resuspension
1.A.3.c	Railways
1.A.3.d	Water-borne Navigation
1.A.3.e	Other Transportation
1.A.4	Residential and other sectors
1.A.5	Non-Specified
1.B.1	Solid Fuels
1.B.2	Oil and Natural Gas
2.A.1	Cement production
2.A.2	Lime production
2.A.3	Glass Production
2.A.4	Other Process Uses of Carbonates
2.B	Chemical Industry
2.C	Metal Industry

Code	IPCC Greenhouse Gas Inventory domain
2.D	Non-Energy Products from Fuels and Solvent Use
2.E	Electronics Industry
2.F	Product Uses as Substitutes for Ozone Depleting Substances
2.G	Other Product Manufacture and Use
3.A.1	Enteric Fermentation
3.A.2	Manure Management
3.C.1	Emissions from biomass burning
3.C.2	Liming
3.C.3	Urea application
3.C.4	Direct N2O Emissions from managed soils
3.C.5	Indirect N2O Emissions from managed soils
3.C.6	Indirect N2O Emissions from manure management
3.C.7	Rice cultivations
4.A	Solid Waste Disposal
4.B	Biological Treatment of Solid Waste
4.C	Incineration and Open Burning of Waste
4.D	Wastewater Treatment and Discharge
5.A	Indirect N2O emissions from the atmospheric deposition of nitrogen in NOx and NH_3
5.B	Fossil fuel fires

Source: authors based on Crippa, et al. (2023).

EDGAR datasets contain disaggregated information for the elements in the table above in Gigagrams for CO_2, CO_2 from biofuels, CH_4, N_2O, and Fluorinated greenhouse gases (F-Gases) PFC, HFC, HCFC, and NF_3. However, while they provide a single dataset with aggregate GHG emissions expressed in CO_2eq using the Global Warming Potential values of the Fifth Assessment Report of the IPCC, all gases are summed into a single figure for each element in Table 1, not distinguished from the individual shares of CO_2 tons equivalent contributed by each gas. For that reason, we used the individualized datasets per gas in Gigagrams and performed the Global Warming Potential AR5 calculation ourselves using coefficients from Myhre et al. Myhre et al. (2013), except for F-Gases, for which the conversion was made by EDGAR and kept separate as needed.

Table 2: Conversion coefficients for methane and nitrous oxide

Gas	Global Warming Potential Values AR5 100-year horizon
Methane (CH4)	28 CO_2 tons equivalent / Gigagram
Nitrous Oxide (N2O)	265 CO_2 tons equivalent / Gigagram

Source: Myhre et al. (2013, 731).

As explained above, while the Edgar dataset is very comprehensive, it lacks the disaggregation by driver (intermediate input, output, endowment, or consumption) to satisfy the needs of CGE studies, when it comes to energy and certain agricultural aspects. That means that it is possible to identify to which volume of the IPCC a certain emission value belongs, but not which fuel product is responsible for it. For example, we can identify that CO_2 emissions are emitted by the Road Transportation sector, but not whether the emissions come from the burning of Diesel or from Bioethanol or any other fuel type. For that reason, we replace the values related to energy with those from our own estimates using the United Nations Energy Database and default emission factors, as well as those elements related to agriculture and LULUCF with emissions data from FAOSTAT at the driver level (explained in the FAOSTAT section below). For completeness, the remaining domains; i.e. industrial processes, waste, and others are linked to total output of specific economic activities, using the International Standard Industrial Classification (United Nations, 2008) where possible (see Table 3).

Table 3: Mapping from EDGAR to individual ISIC economic activities

IPCC Code	IPCC	ISIC Code	ISIC
2.A.1	Cement production	2394	Manufacture of cement, lime and plaster
2.A.2	Lime production	2394	Manufacture of cement, lime and plaster
2.A.3	Glass Production	2310	Manufacture of glass and glass products
2.A.4	Other Process Uses of Carbonates	9999	Undetermined
2.B	Chemical Industry	2011	Manufacture of basic chemicals

IPCC Code	IPCC	ISIC Code	ISIC
2.C	Metal Industry	2410	Manufacture of basic iron and steel
2.D	Non-Energy Products from Fuels and Solvent Use	9999	Undetermined
2.E	Electronics Industry	2640	Manufacture of consumer electronics
2.F	Product Uses as Substitutes for Ozone Depleting Substances	9999	Undetermined
2.G	Other Product Manufacture and Use	9999	Undetermined
4.A	Solid Waste Disposal	3821	Treatment and disposal of non-hazardous waste
4.B	Biological Treatment of Solid Waste	3822	Treatment and disposal of hazardous waste
4.C	Incineration and Open Burning of Waste	3822	Treatment and disposal of hazardous waste

Emissions from Agriculture and LULUCF

Data on CO_2, CH_4, and N_2O emissions from Agriculture, Forestry, and Other Land Use, including Land Use Change were sourced from FAOSTAT[3]. The level of disaggregation of its individual datasets allows us to identify the driving agricultural product (crop) or endowment (animal assets or cropland area) behind the emission of a specific gas. Since values are expressed in kilotonnes (kt) of each GHG, we converted them to Global Warming Potential AR5 for a 100-year horizon using the values in Table 2. To facilitate the mapping to CGE and Input-Output frameworks, we added a correspondence to individual ISIC economic activities responsible for the emission, where possible.

[3] Datasets are available at https://www.fao.org/faostat/en/#data under the heading "Climate Change: Agrifood systems emissions".

Table 4: FAOSTAT domains sourced

Code	Sector
3.B.1.a	LULUCF - Forestland
3.B.1.b	LULUCF - Net Forest conversion
3.B.2	LULUCF - Drained organic soils (CO2)
3.B.6.a	LULUCF - Fires in organic soils
3.B.6.b	LULUCF - Forest fires
3.C.1.c	Agriculture - Savanna fires
3.C.5	Agriculture - Drained organic soils (N2O)
3.A.1	Agriculture - Enteric fermentation (Emissions CH4)
3.A.2.a	Agriculture - Emissions (N2O) (Manure applied)
3.A.2.b	Agriculture - Manure management (Emissions N2O)
3.A.2.c	Agriculture - Manure management (Emissions CH4)
3.C.6	Agriculture - Manure left on pasture (Emissions N2O)
3.C.4	Agriculture - Synthetic Fertilizers
3.C.1.a	Agriculture - Burning crop residues (Emissions CH4)
3.C.1.b	Agriculture - Burning crop residues (Emissions N2O)
3.C.7	Agriculture - Rice cultivation (Emissions CH4)
3.C.8	Agriculture - Crop residues (Emissions N2O)

FAOSTAT datasets go beyond the elements stipulated in the third category of an IPCC inventory, and provide information on energy use by the agricultural sector, as well as emissions from industrial processes and waste. Since we source those from EDGAR for all economic sectors, we only use data from the domains shown in Table 4 from FAOSTAT.

More importantly, the reason why we replaced the data from EDGAR on agriculture emissions with FAOSTAT is because of the ability to differentiate the different drivers of those emissions, as well as the separate totals by gas. Table 5 shows the different elements from FAOSTAT available for practitioners to map to inputs, outputs, endowments, or consumption.

Greenhouse Gas Emissions from Stationary and Mobile Combustion

A choice was made to estimate emissions from combustion, although the global datasets that we use already have this information in the aggregate, because it was important to be able to distinguish between the different fuels driving the combustion and the sectors in which this happened. Greenhouse gas emissions from fuel combustion and fugitive emissions were estimated applying Tier 1 coefficients from the 2006 IPCC Guidelines for National Greenhouse Gas Inventories to national information on fuel uses from the United Nations Energy Database (UNSD 2023) using Equation 1. We first applied conversion factors from units of volume

Table 5: FAOSTAT domains sourced

Code	Driver
0015	Wheat
0027	Rice
0044	Barley
0056	Maize (corn)
0071	Rye
0075	Oats
0079	Millet
0083	Sorghum
0116	Potatoes
0156	Sugar cane
0176	Beans, dry
0236	Soya beans
0946	Buffalo
0960	Cattle, dairy
0961	Cattle, non-dairy
0976	Sheep
1016	Goats
1049	Swine, market
1051	Swine, breeding
1052	Chickens, layers
1053	Chickens, broilers
1068	Ducks
1079	Turkeys
1096	Horses
1107	Asses
1110	Mules and hinnies
1126	Camels
1177	Llamas
2012	Manufacture of fertilizers and nitrogen compounds
9997	Undetermined from LULUCF
9998	Undetermined from Agriculture

(metric tons) to Terajoules from *Definitions, Units of Measure and Conversion Factors* (United Nations 1987).

$$E_{GHG,fuel} = C_{fuel} * F_{GHG,fuel} \tag{1}$$

Where:

- $E_{GHG,fuel}$ = Emissions of a given GHG by type of fuel (kg GHG).
- C_{fuel} = Amount of fuel combusted (TJ).
- $F_{GHG,fuel}$ = default emission factor of a given GHG by type of fuel (kg gas/TJ).

The UN dataset is organized according to the standard construction of Energy Balances and it features transactions related to supply, imports, stock variation, exports, transformation, industrial, and final use of all energy resources. In order to estimate emissions we focused on 39 energy resources subject to combustion or fugitive emissions (shown in Table 6) and 88 transactions.

Results

Database

Emissions

The resulting database reports GHG emissions in thousand metric tons CO_2eq using the 100-year Global Warming Potential (GWP) of the IPCC's Fifth Assessment Report (AR5). We provide three formats for practitioners' convenience. A long format (what other efforts term "normalized"), where ISIC sectors are laid out in a single column, with a single column of values, a pivot format, where ISIC sectors (including undetermined) are arranged in columns of data values, and a raw long format with all fields disaggregated.

Aguiar, Angel, Maksym Chepeliev, Erwin L. Corong, Robert McDougall, and Dominique van der Mensbrugghe. 2019. "The GTAP Data Base: Version 10." *Journal of Global Economic Analysis* 4 (1): 1–27. https://doi.org/10.21642/JGEA.040101AF.

Chepeliev, Maksym. 2024. "Chapter 13A: CO2 Emissions from Fossil Fuels Combustion." http://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=7130.

Crippa, Monica, Diego Guizzardi, Federico Pagani, Manjola Banja, Marilena Muntean, Edwin Schaaf, William Edward Becker, et al. 2023. "GHG Emissions of All World Countries." Luxembourg. https://publications.jrc.ec.europa.eu/repository/handle/JRC134504.

European Commission, International Monetary Fund, Organisation for Economic Co-operation, Development, United Nations, and World Bank. 2009. "System of National Accounts 2008." https://unstats.un.org/unsd/nationalaccount/sna2008.asp.

Table 6: Energy Commodities

Commodity
Anthracite
Aviation gasoline
Bagasse
Bio jet kerosene
Biodiesel
Biogases
Biogasoline
Black liquor
Blast furnace gas
Brown coal
Brown coal briquettes
Charcoal
Coal
Coke oven gas
Coke-oven coke
Coking coal
Ethane
Fuel oil
Fuelwood
Gas coke
Gas oil/diesel oil
Gasoline-type jet fuel
Gasworks gas
Hard coal
Kerosene-type jet fuel
Lignite
Lignite brown coal
Lignite brown coal- recoverable resources
Liquefied petroleum gas (lpg)
Motor gasoline
Naphtha
Natural gas
Natural gas (including lng)
Other bituminous coal
Other coal products
Other hydrocarbons
Other liquid biofuels
Peat (for fuel use)
Petroleum coke
Refinery gas

Eurostat. 2015. “Manual for Air Emissions Accounts.” Luxembourg. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-15-009.

FAO. 2024. “FAOSTAT Domain Emissions Totals.” Rome. https://www.fao.org/faostat/en/#data/GT.

Flachenecker, Florian, Emmanuelle Guidetti, and Pierre-Alain 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.” Paris. https://doi.org/10.1787/7d88dfdd-en.

Gutman, Verónica, María Priscila Ramos, Carlos Adrián Romero, Martina Chidiak, Gustavo Adolfo Ferro, María Laura Ojeda, Exequiel Romero Gómez, Juan Ignacio Mercatante, María Paula Covelli, and Verónica Fernández. 2024. “Integración de Información Climática y Económica En Los Sistemas Estadísticos Nacionales: Armonizando Los Inventarios Nacionales de Gases de Efecto Invernadero (INGEI) Con El Sistema de Contabilidad Ambiental y Económica (SCAE).” Buenos Aires. https://ojs.econ.uba.ar/index.php/DT-IIEP/issue/view/500.

IPCC. 1996. “Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories.” London. https://www.ipcc-nggip.iges.or.jp/public/gl/invs4.html.

———. 2006. “2006 IPCC Guidelines for National Greenhouse Gas Inventories.” Geneva. https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_02_Ch2_Generic.pdf.

Myhre, Gunnar, Drew Shindell, Francois-Marie Bréon, William Collins, Jan Fuglestvedt, Jianping Huang, Dorothy Koch, et al. 2013. “Anthropogenic and Natural Radiative Forcing.” Switzerland. https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter08_FINAL.pdf.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

United Nations. 1987. “Energy Statistics: Definitions, Units of Measure and Conversion Factors.” Studies in {Methods} 44. New York: United Nations Department of Economic; Social Affairs, Statistical Office. https://unstats.un.org/unsd/publication/SeriesF/SeriesF_44E.pdf.

———. 2008. “International Standard Industrial Classification of All Economic Activities: Revision 4.” New York: UN Department of Economic; Social Affairs. Statistics Division. https://unstats.un.org/unsd/classifications/Econ/isic.

UNSD. 2023. “Energy Statistics Database 2021: Notes on Sources.” https://unstats.un.org/unsd/energystats/pubs/yearbook/2021/metadata.pdf.