Global Trade Analysis Project

GTAP 12 Data Base

Virtual Seminar Series, Vol 7, No 1 (2026)

Center for Global Trade Analysis
Department of Agricultural Economics
Purdue University

Global Trade Analysis Project

Welcome and Introduction

Channing Arndt
carndt@purdue.edu

Outline

- **Introduction**
  - Input-Output Tables
  - Macroeconomic conditions, including extra balance of payments, and
  - Services Trade Data
- **New build and Macro projections (SSP)**
- **Mainstreaming Land Use and Cover**
- **Energy data within GTAP, energy volumes and emissions**
- **Labor splits, Income/factor taxes, GTAPAgg3**

Global Trade Analysis Project

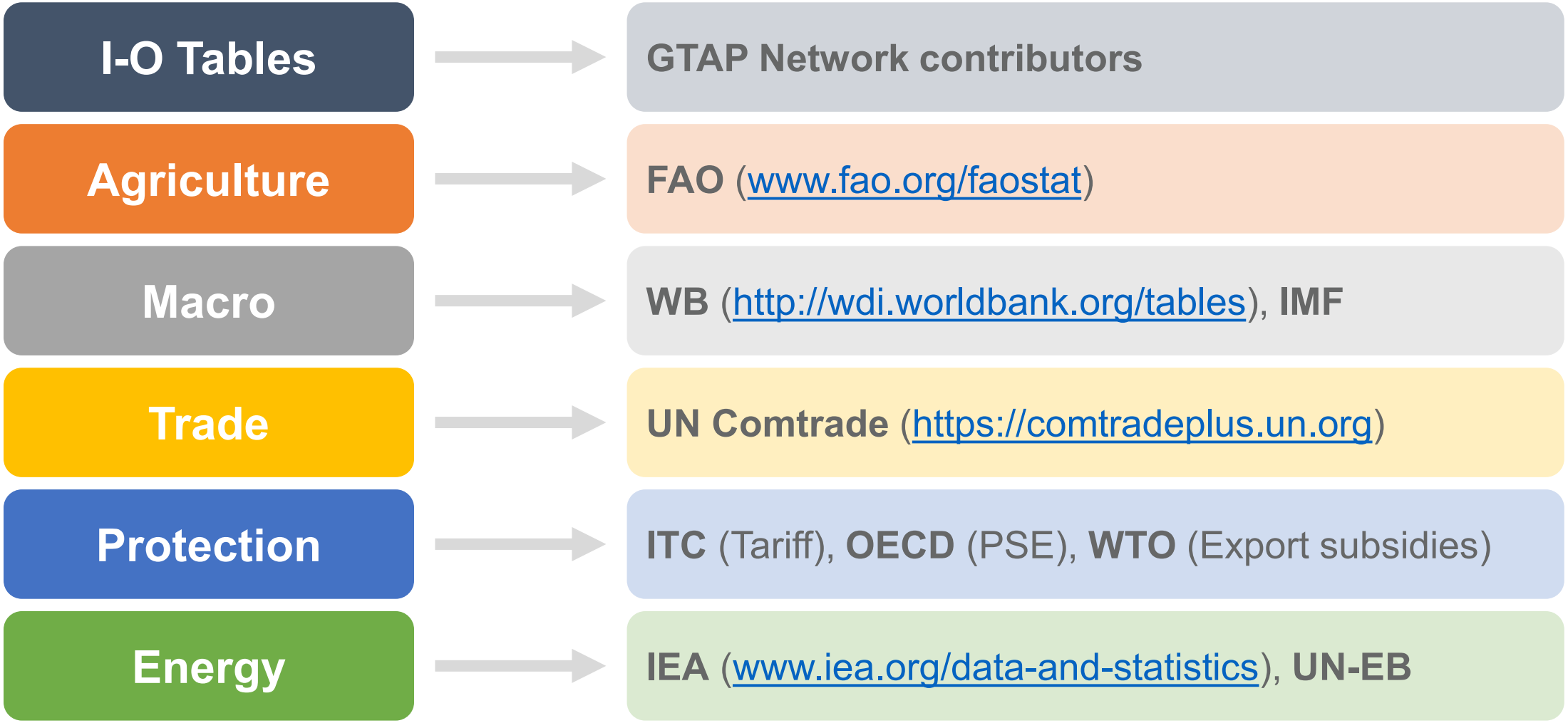
Overview of the GTAP 12 Data Base

Angel Aguiar
aaguiar@purdue.edu

Introduction

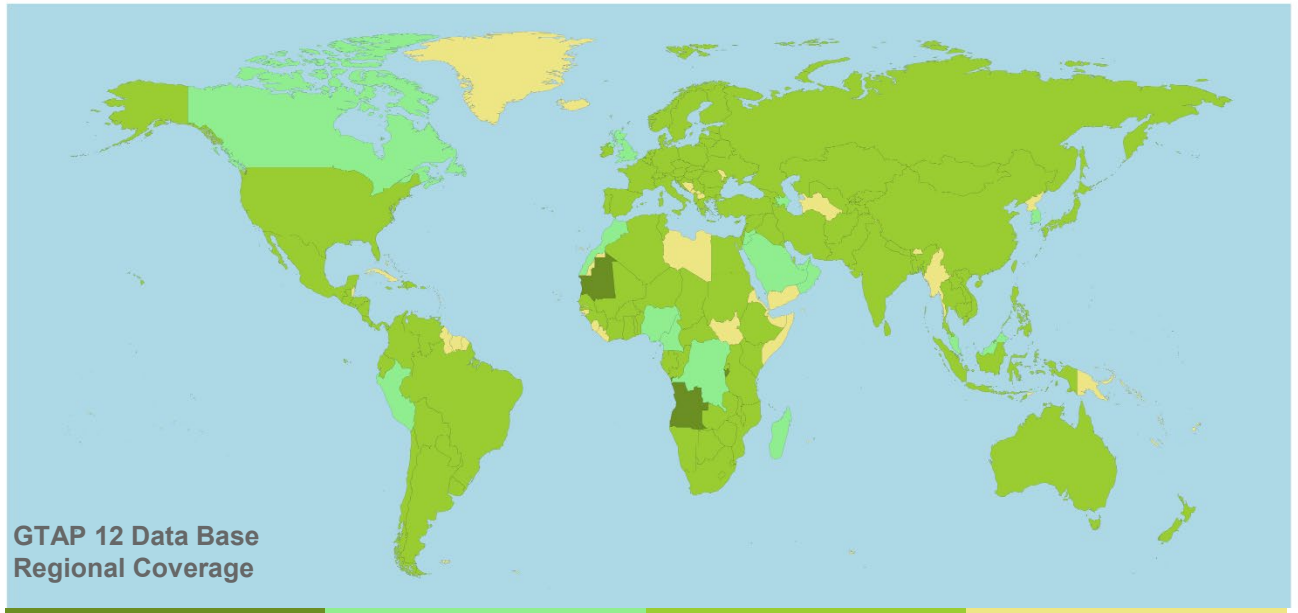
- **GTAP provides a representation of the world economy at a point in time**
  - Trade, transport, and protection data characterizing economic linkages among countries; the (I-O) data account for inter-sectoral linkages within countries
  - 7 reference years, 163 regions, 65 sectors
- **Documentation**
  - Aguiar, A., Baldos, U., Chepeliev, M., Corong, E., & Simonato, T. (2025). The Global Trade Analysis Project (GTAP) Data Base: Version 12. Journal of Global Economic Analysis, 10(2), 1-45. Retrieved from https://doi.org/10.21642/JGEA.100201AF
  - GTAP website: www.gtap.org/databases/v12/v12_doco.aspx

Data Sources



Input-Output Data

- **I-O tables contributed by users** → www.gtap.org/databases/contribute/
- **New reference years**
  - 2019 and 2023
- **163 regions**
  - **145 individual country tables:**
  - **18 composite regions**



Macroeconomic condition

- **GDP = C + G + I + X - M**
  - GDP and its expenditures are retrieved from WB, IMF and UN
- **S – I = X – M**

- **In GTAP 12 we report additional elements of the balance of payments**
  - Foreign investment income from IMF
  - Official development aid from OECD
  - Remittances from WB
- **S – I = X – M + Net remittances + Net Aid + Net investment income**

Services Trade Data

- **Balanced Trade in Services (BaTiS)**
  - Maintenance and repair services (trd),
  - Construction (cns),
  - Insurance and pension services (ins),
  - Financial services (ofi),
  - Telecommunications, computer, and information services (cmn),
  - Government goods and services (osg)
- **Trade in Services by mode of supply (TiSMoS)**
  - Transport (atp, wtp, otp, whs, cmn),
  - Travel (hht, edu, afs),
  - Other business services (obs, trd, rsa),
  - Personal, cultural, and recreational services (cmn, hht, edu, ros)

Global Trade Analysis Project

Overview of the GTAP 12 Data Base

Angel Aguiar
aaguiar@purdue.edu

Global Trade Analysis Project

New build and Macro projections (SSP)

Thiago Simonato
tsimonat@purdue.edu

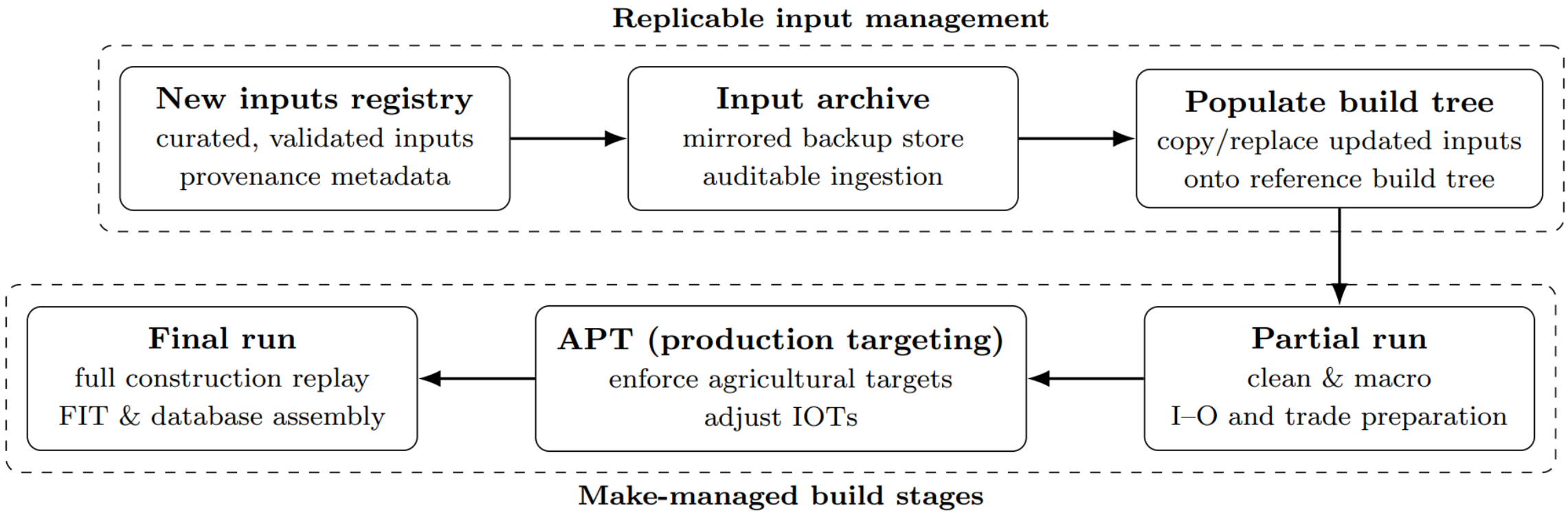
GTAP Data Build Overview

- **Goal: produce one consistent global GTAP Data Base by combining**
  - domestic input–output tables (regional databases)
  - international datasets (macro, trade, protection, energy)
- **High-level flow (simple story)**
  - **Prepare/clean** regional tables and international inputs
  - **Update** everything to a common reference year (APT & FIT-style reconciliation)
  - **Assemble** into the final global database + summaries (sets, parameters, tax rates, BaseView)
- **What's new in the GTAP 12 Build System?**
  - Better input management, more automation, faster runs

Input provenance + safer speed-ups (automation & parallelism)

- **Problem before:** inputs copied manually, no logs, making it hard to track source and data characteristics
  - **Solution**: Adopt an online sharable input registry records
    - where each new input comes from
    - where it goes in the build
    - key metadata (year, update date, validation status)
  - **Benefits**: Adopt an online sharable input registry records
    - repeatable builds, fewer “it worked on my machine”
    - easier auditing (“what changed since last version?”)
    - faster onboarding for new team members
- **Parallel execution:** faster, but controlled
  - **Parallel region runs with safeguards:** bounded concurrency keeps multi-region execution stable
  - **More reliable completion**: timeouts and retry logic reduce “stuck” runs and handle transient failures automatically

Reproducible Pipeline with Parallel Execution



Open-source SSP pipeline

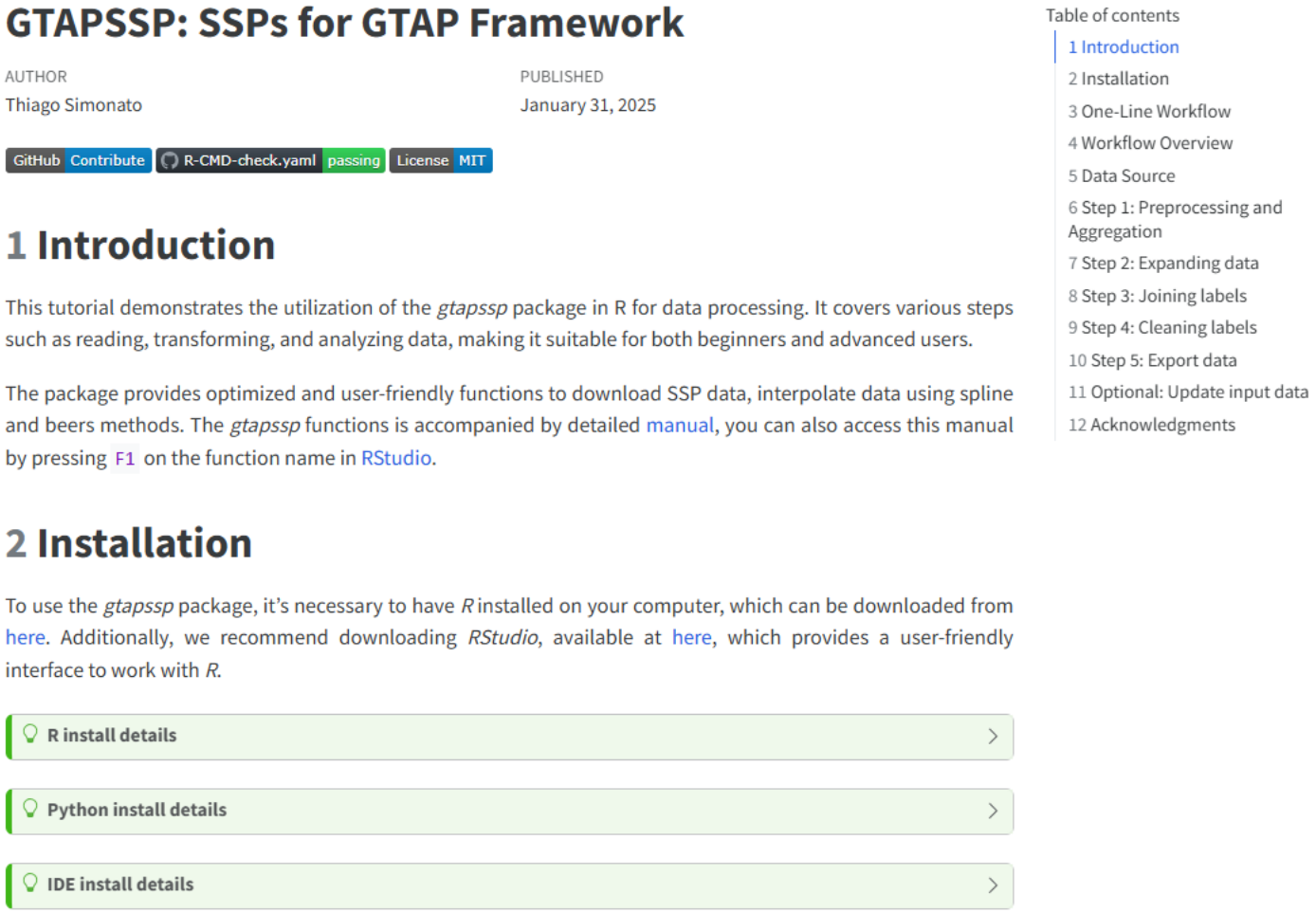
- **What SSPs provide:** long-run scenario paths for key drivers such as GDP and population.
- **Why GTAP uses SSPs:** to build consistent baselines and compare alternative futures using a common set of macro-demographic assumptions (narratives).
- **Typical data challenge:** SSP datasets often come in 5-year steps (and may have gaps), while GTAP workflows usually need annual, gap-free series aligned to GTAP regions and variables.
- **What's new in the GTAP 12 SSP integration?**
  - **Updated projections**: inputs updated to SSP Release 3.2.
  - **Open + reproducible**:
    - Full preprocessing pipeline is open-source on GitHub https://github.com/tsimonato/gtapssp
    - Compiled SSP datasets are built into GtapAgg3.

End-to-End Pipeline on GitHub

- **Transparency:**
  end-to-end pipeline is public. See how SSP data are produced in https://tsimonato.github.io/gtapssp/

- **Reproducibility:**
  rerun the same steps to recreate the exact outputs.

- **Collaboration & customization:**
  fork/contribute and adapt settings to your needs.



GTAPSSP: SSPs for GTAP Framework

AUTHOR
Thiago Simonato

PUBLISHED
January 31, 2025

GitHub Contribute | R-CMD-check.yaml passing | License MIT

Table of contents
1 Introduction
2 Installation
3 One-Line Workflow
4 Workflow Overview
5 Data Source
6 Step 1: Preprocessing and Aggregation
7 Step 2: Expanding data
8 Step 3: Joining labels
9 Step 4: Cleaning labels
10 Step 5: Export data
11 Optional: Update input data
12 Acknowledgments

1 Introduction

This tutorial demonstrates the utilization of the *gtapssp* package in R for data processing. It covers various steps such as reading, transforming, and analyzing data, making it suitable for both beginners and advanced users.

The package provides optimized and user-friendly functions to download SSP data, interpolate data using spline and beers methods. The *gtapssp* functions is accompanied by detailed manual, you can also access this manual by pressing `F1` on the function name in RStudio.

2 Installation

To use the *gtapssp* package, it's necessary to have *R* installed on your computer, which can be downloaded from here. Additionally, we recommend downloading *RStudio*, available at here, which provides a user-friendly interface to work with *R*.

R install details

Python install details

IDE install details

Global Trade Analysis Project

New build and Macro projections (SSP)

Thiago Simonato
tsimonat@purdue.edu

Global Trade Analysis Project

Mainstreaming Land Use and Land Cover

Uris Baldos
ubaldos@purdue.edu

About the GTAP LULC Database

- GTAP LULC is used in the GTAP Agro-Ecological Zone (GTAP-AEZ) model which extends the standard GTAP model by including competition in land endowments across different sectors (crops, pasture, forestry sectors) and within the crop sector

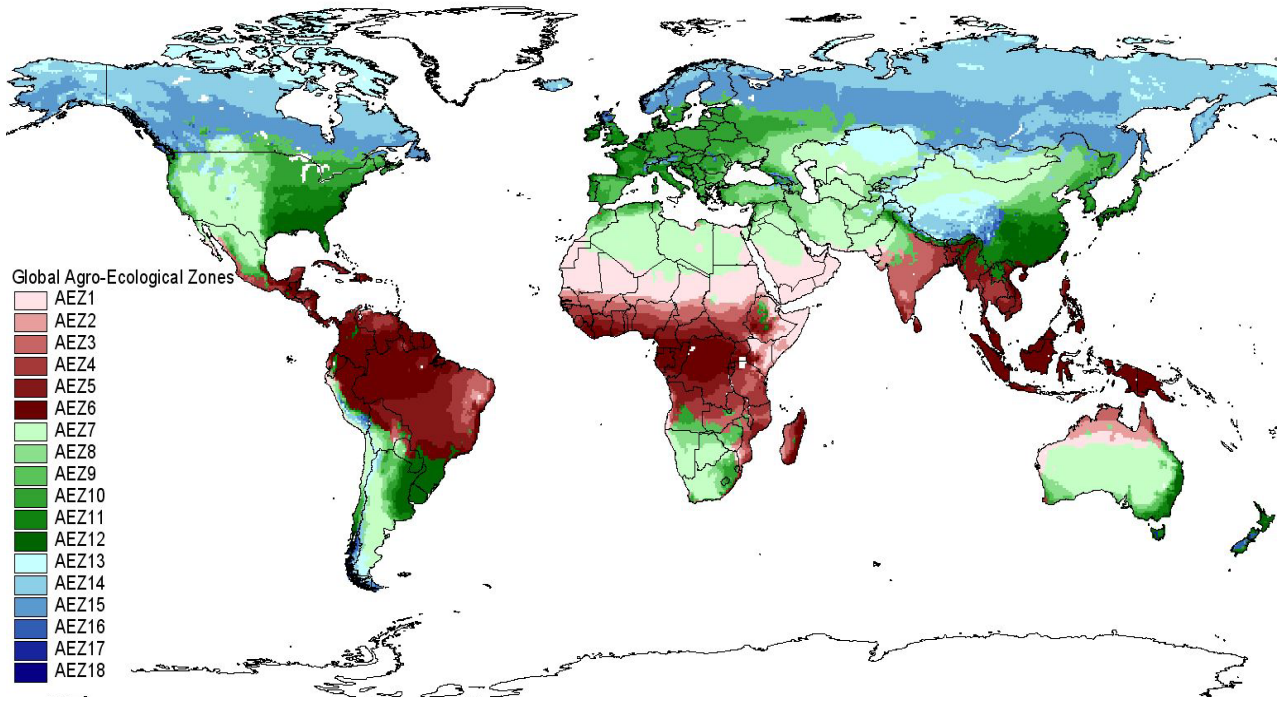
- GTAP LULC expands the standard database by:
  - Incorporating spatially explicit information on land use and land cover
  - Reallocating payments to land endowments based on land use information

- It is now part of the main build to ensure consistency and rapid construction

GTAP LULC: Agro-Ecological Zones

- **18 Agro-Ecological Zones**
  - 6 growing periods (6 categories x 60 day intervals)
  - 3 climatic zones (tropical, temperate and boreal)

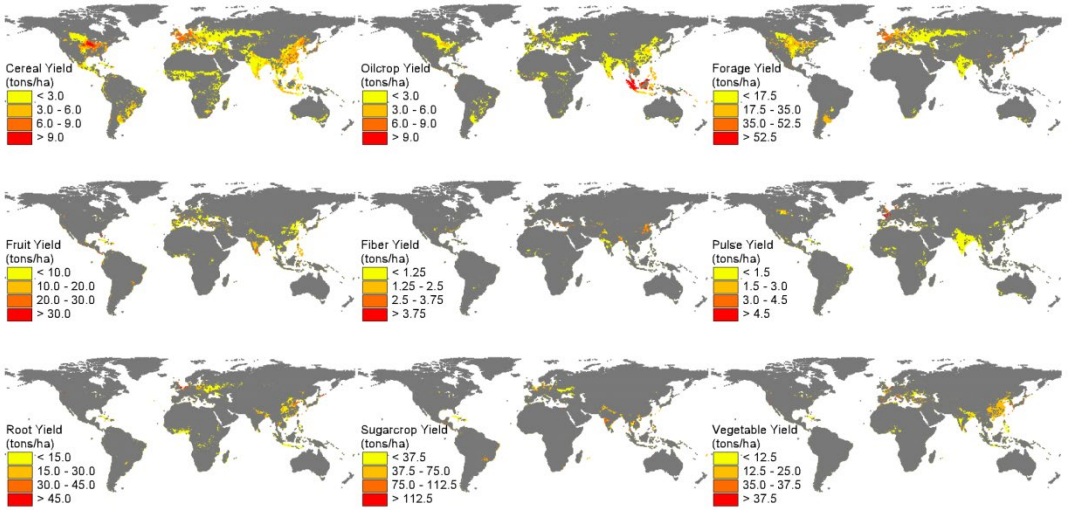
- **The competition for land within a given AEZ across uses is constrained to include activities that have been observed to take place in that AEZ**



FAO / IIASA (2025). Global Agro-Ecological Zoning version 5

GTAP LULC: Gridded Datasets

Land Use Data



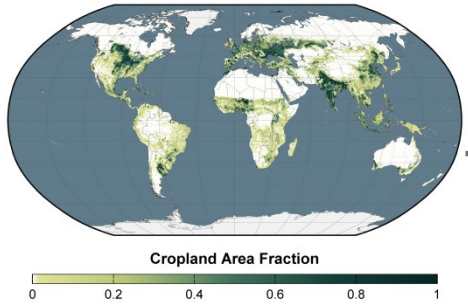
Monfreda et al (2008). Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000.

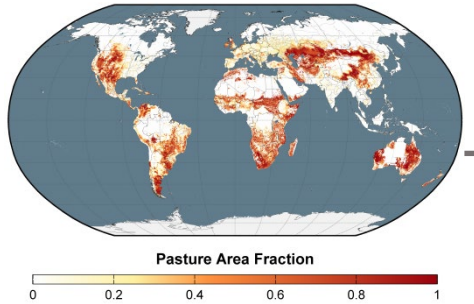
- **Crop output and harvested area (175 crops)**

Land Cover Data

Agricultural Lands

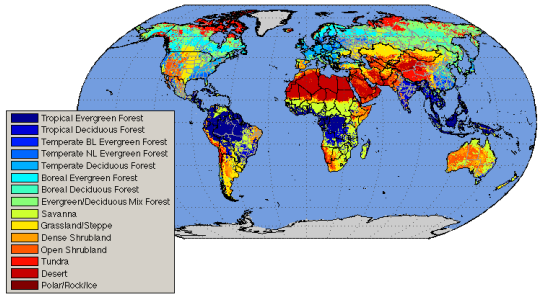
Ramankutty et al. (2008). Geographic distribution of global agricultural lands in the year 2000.

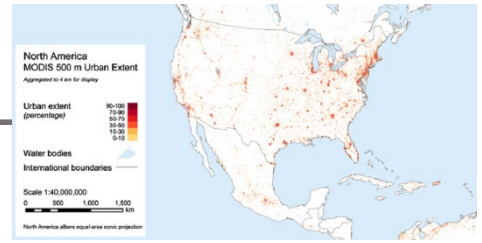




Potential Vegetation

Ramankutty, N., and J.A. Foley (1999). Estimating historical changes in global land cover: croplands from 1700 to 1992





Urban Extent

Schneider et al (2010). Mapping global urban areas using MODIS 500-m data: New methods and datasets based on urban ecoregions.

FAOSTAT (GTAP Reference Years)

GTAP LULC Database

GTAP LULC Header Arrays

Updated Headers in Base Data file

Header	Coefficient Name	Description
VFM	VFM	Endowments - Firms Purchases at Market Prices
EVOA	EVOA	Endowments - Output at Agents Prices
EVFA	EVFA	Endowments - Firms Purchases at Agents Prices

New Headers in Base Data file

Header	Coefficient Name	Description
TONS	TONS	Production (1000 metric tonnes): 8 crop sectors
AREA	AREA	Harvested Area (1000 hectares): 8 crop sectors
LCOV	LCOV	Land Cover (1000 hectares): 4 land cover types

Global Trade Analysis Project

Mainstreaming Land Use and Land Cover

Uris Baldos
ubaldos@purdue.edu

Global Trade Analysis Project

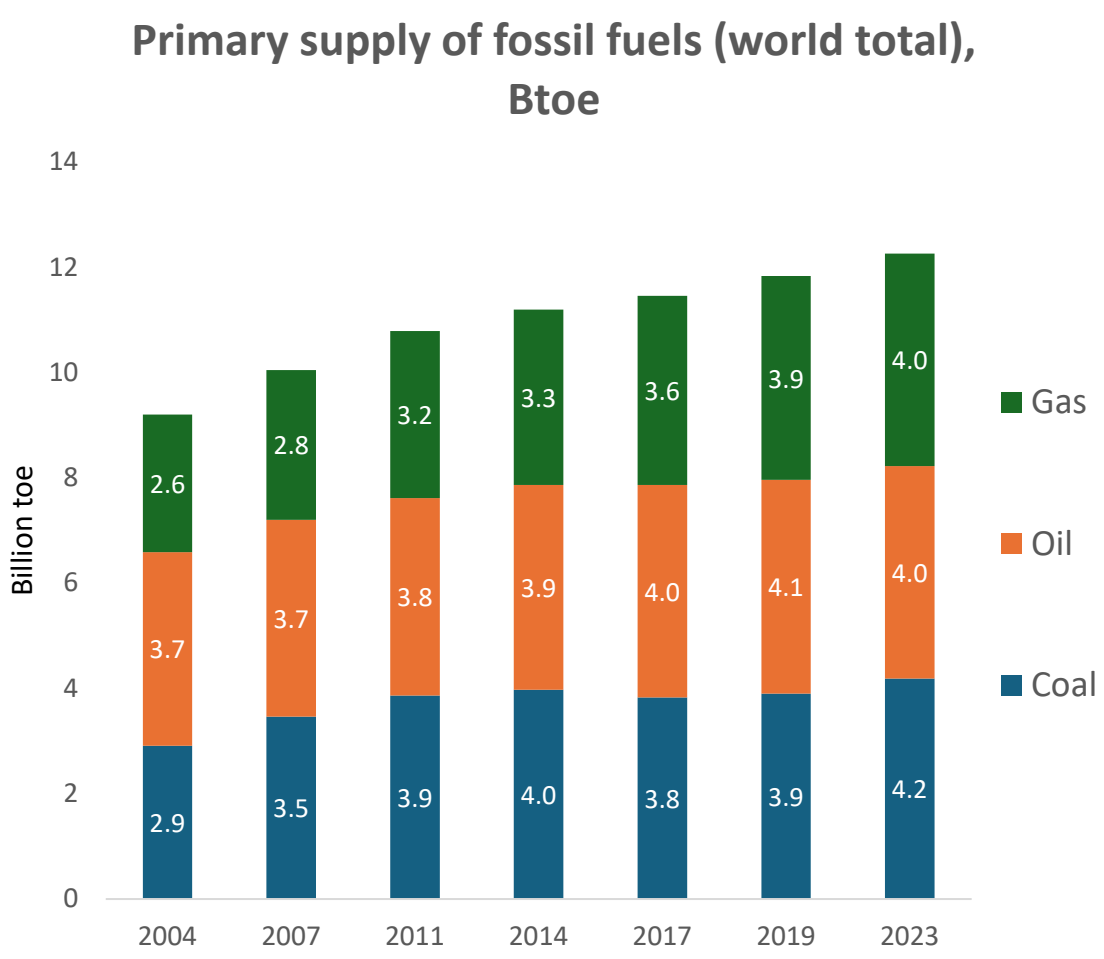
Energy and Emissions

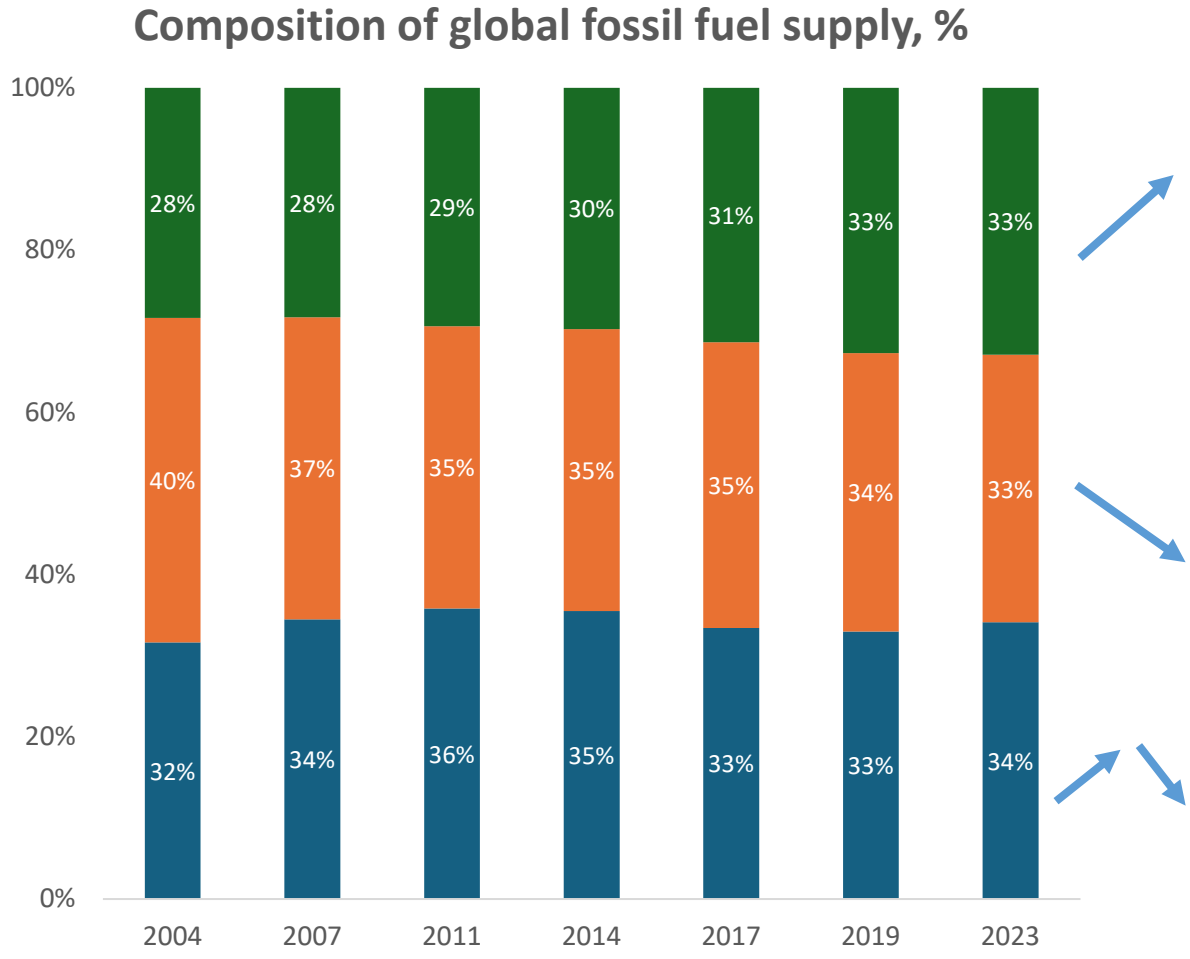
Maksym Chepeliev
mchepeli@purdue.edu

Energy database for GTAP: key features

- **Key features of the GTAP energy database**
  - A special data set that overrides energy flows, prices, and taxes in GTAP.
  - Primarily relies on the extended energy balances from IEA as well as several other data sources (e.g. UN, BP, Eurostat, OLADE, COMTRADE, US DOE, China Energy Databook, etc.).
- **GTAP 12 energy data**
  - Updated bilateral energy flows
    - Rely on British Petroleum and Eurostat to complement the COMTRADE bilateral data.
    - Primarily addresses the issue of bilateral trade in natural gas. Unilateral totals are taken from IEA.
  - UN energy balances are used to complement IEA for selected countries.
    - IEA explicitly reports data for 152 individual countries.
    - UN energy balances help to expand the coverage to 214 countries.
  - Fossil fuel consumption subsidies (based on IEA and IMF data) remain part of the standard GTAP Data Base.

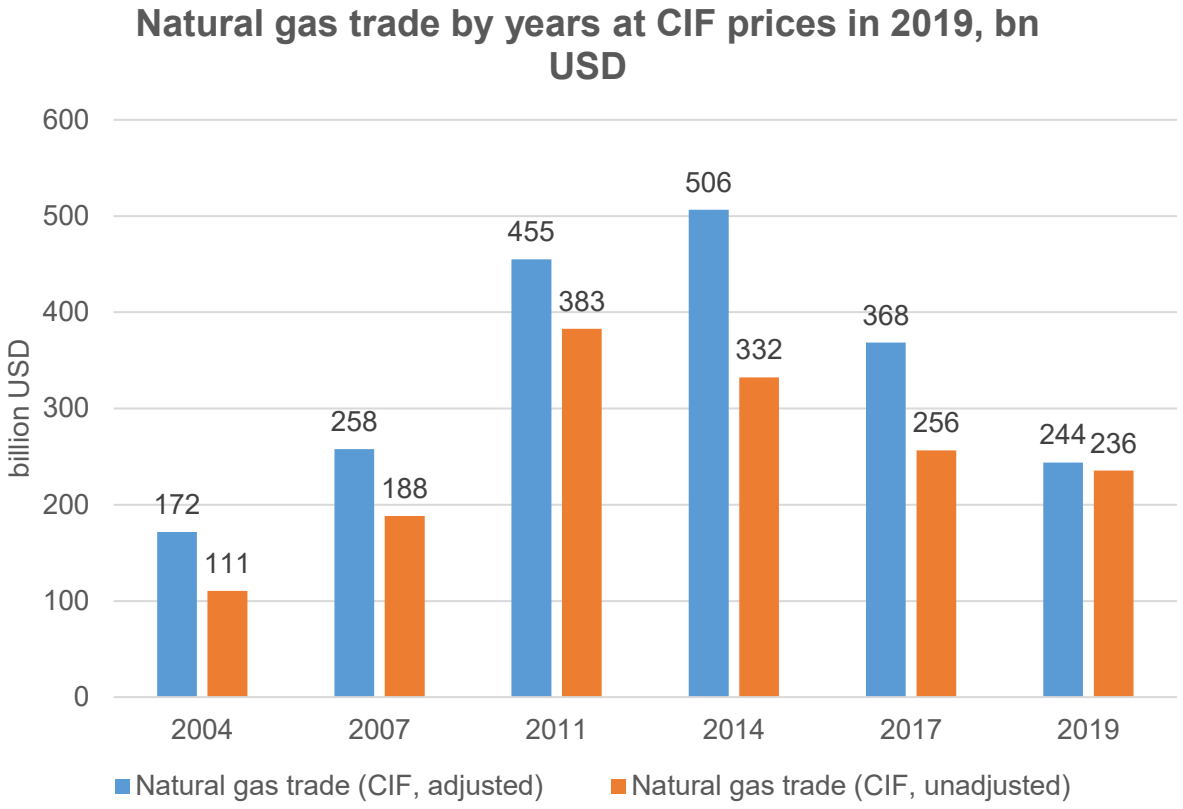
Global fossil fuel energy supply gradually increased over time with major structural shifts

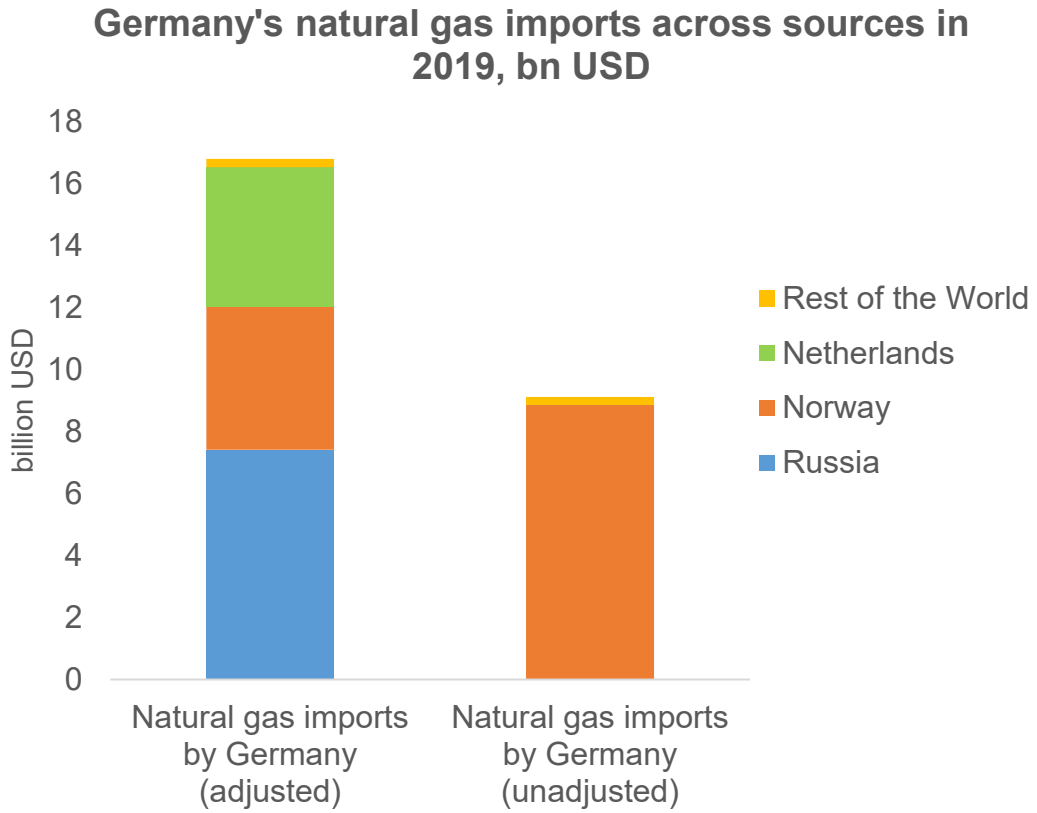




Updated bilateral energy trade data: a case of natural gas

- Complementary bilateral energy flows data are used to split unilateral totals from IEA.
- COMTRADE reports significant values/shares of unclassified sources and destinations.
- This is particularly an issue in the case of natural gas.
- Incorporating selected bilateral flows using additional data sources (BP, Eurostat) improves the representation of corresponding flows.

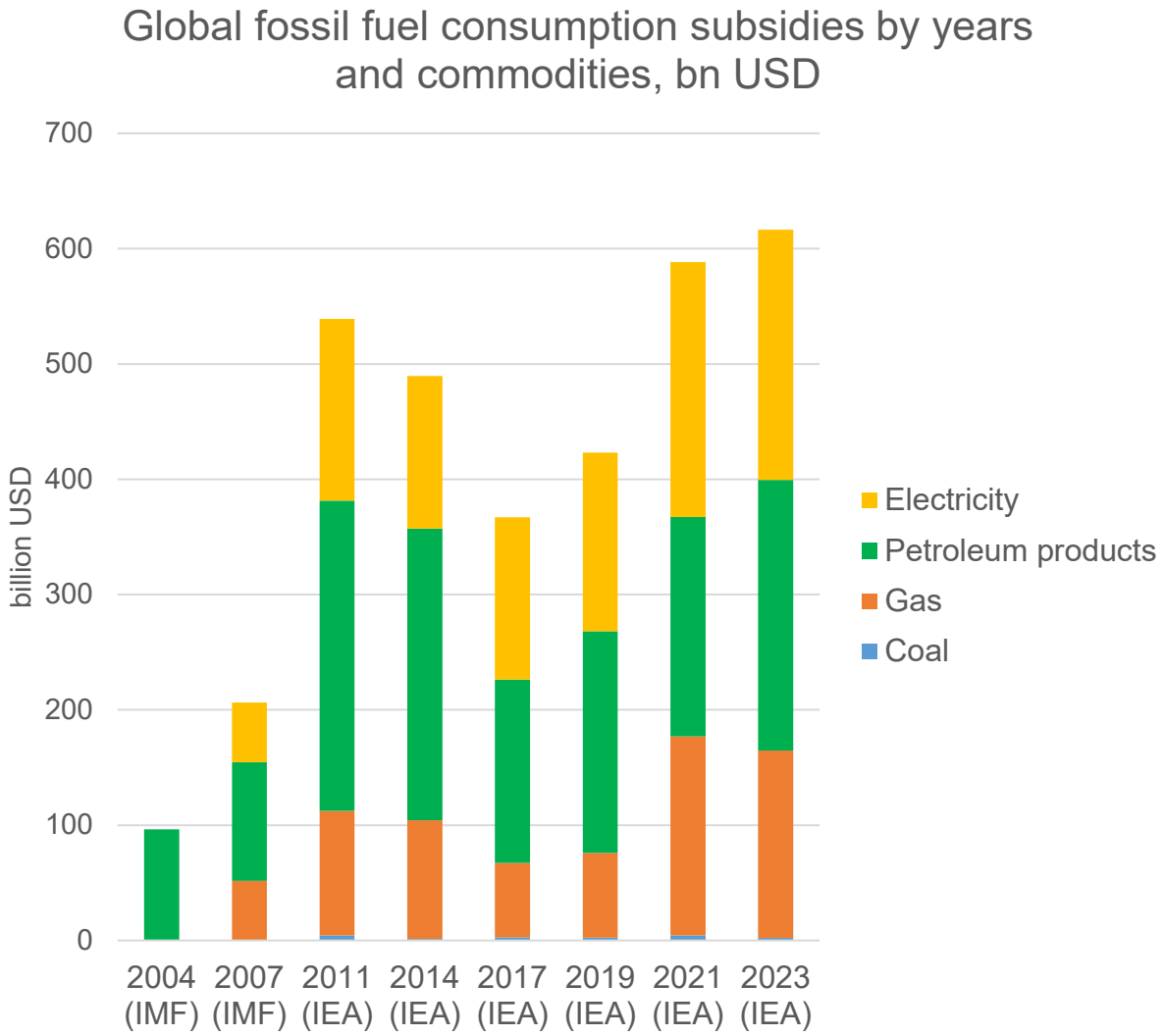




Source: GTAP v12 energy data.

Fossil fuel consumption subsidies are sourced from two core data sources

- For the pre-2010 period (2004 and 2007 reference years), energy subsidies are sourced from the **IMF data:**
  - IMF data has higher country coverage and is accompanied by both energy volumes and prices, however, the data is not representative of some of the recent subsidy reforms.
- The post-2010 period (2011, 2014, 2017, 2019 and 2023) is covered using **IEA energy subsidies data:**
  - IEA data reports subsidies starting from 2010 on the annual basis in constant prices.
  - US CPI is used to convert subsidy values to current prices.

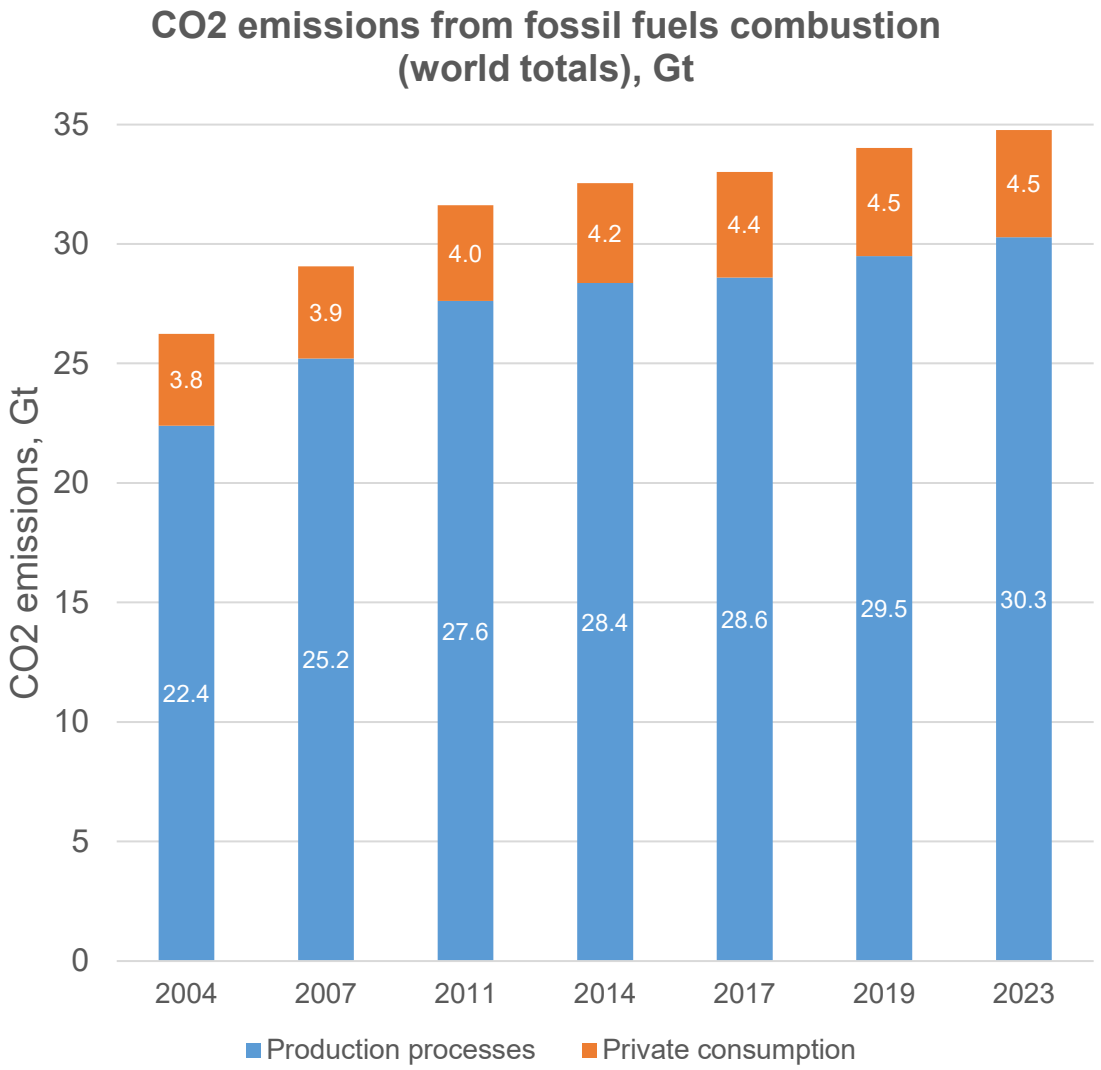


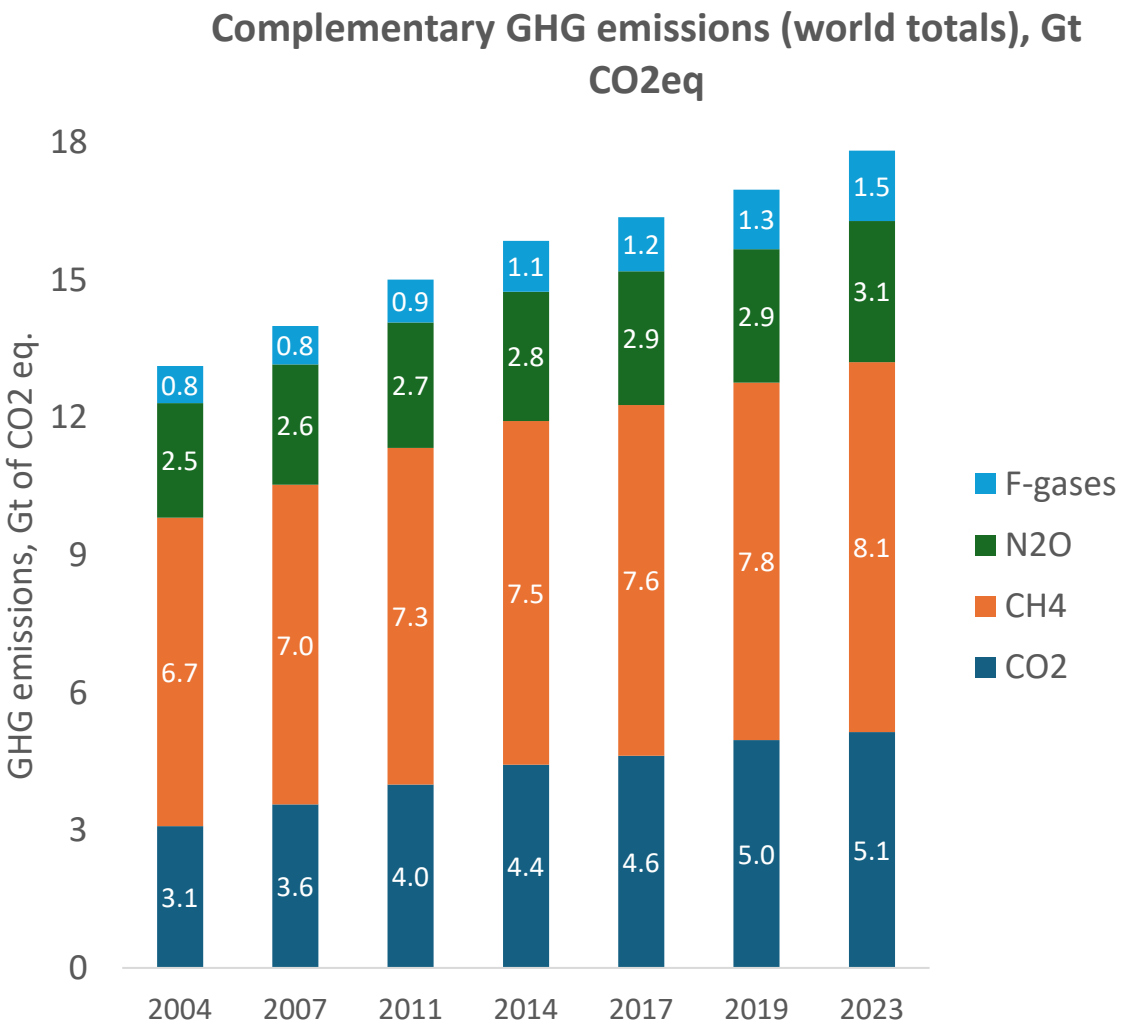
Source: author's estimates based on the IEA and IMF datasets.

GTAP 12 emissions methodology builds on the updated GTAP 11 framework

- **Fossil fuel combustion CO_2 emissions utilize 2006 IPCC guidelines (introduced in GTAP 11) and combine IPCC-derived emission factors with GTAP fossil fuel energy data.**
- **Complementary GHG emissions and air pollutants data rely on EDGAR and FAO databases:**
  - In the case of air pollutant emissions GTAP 12 utilizes the EDGAR v8.1 database (currently the latest available year for air pollution data is 2022).
  - The latter preserves a split between bio- and fossil-based emissions allowing for a more refined mapping of emission flows by drivers.
- **In the final database complementary GHG emission and air pollutants are combined in a single set.**
  - Emission flows are reported in million metric tons.
  - Global warming potentials (GWPs) for GHGs are reported in a separate header providing GWPs across AR2, AR4 and AR5.

Global GHG emissions show consistent growth over the years





Global Trade Analysis Project

Energy and Emissions

Maksym Chepeliev
mchepeli@purude.edu

Global Trade Analysis Project

Labor splits, Income/factor taxes, and GTAPAgg3

Erwin Corong

ecorong@purdue.edu

Income and Factor Taxes

- **GTAP Data Base includes income and factor tax payments, which can be calculated from three headers: EVOS, EVFB and EVFP**
  - Income Taxes: INCTAX(e,a,r) = EVFB(e,a,r) – EVOS(e,a,r)
  - Factor Taxes: ETAX(e,a,r) = EVFP(e,a,r) – EVFB(e,a,r)

- **Since GTAP DB version 6, income and factor tax data is based on International Monetary Fund's Government Finance Statistics**
  - Tax coverage:
    - Corporate taxes on income and profits (G111: G1111)
    - Individual taxes on income and profits (G111: G1112)
    - Social Security contributions (G12)
    - Payroll and manpower taxes (G112)
    - Property taxes (G113)

- **GTAP DB v12: 187 countries**
  - IMF GFS data (1997-2023) and OECD Global Revenue Statistic Database (1990-2023)

Income and Factor Tax processing

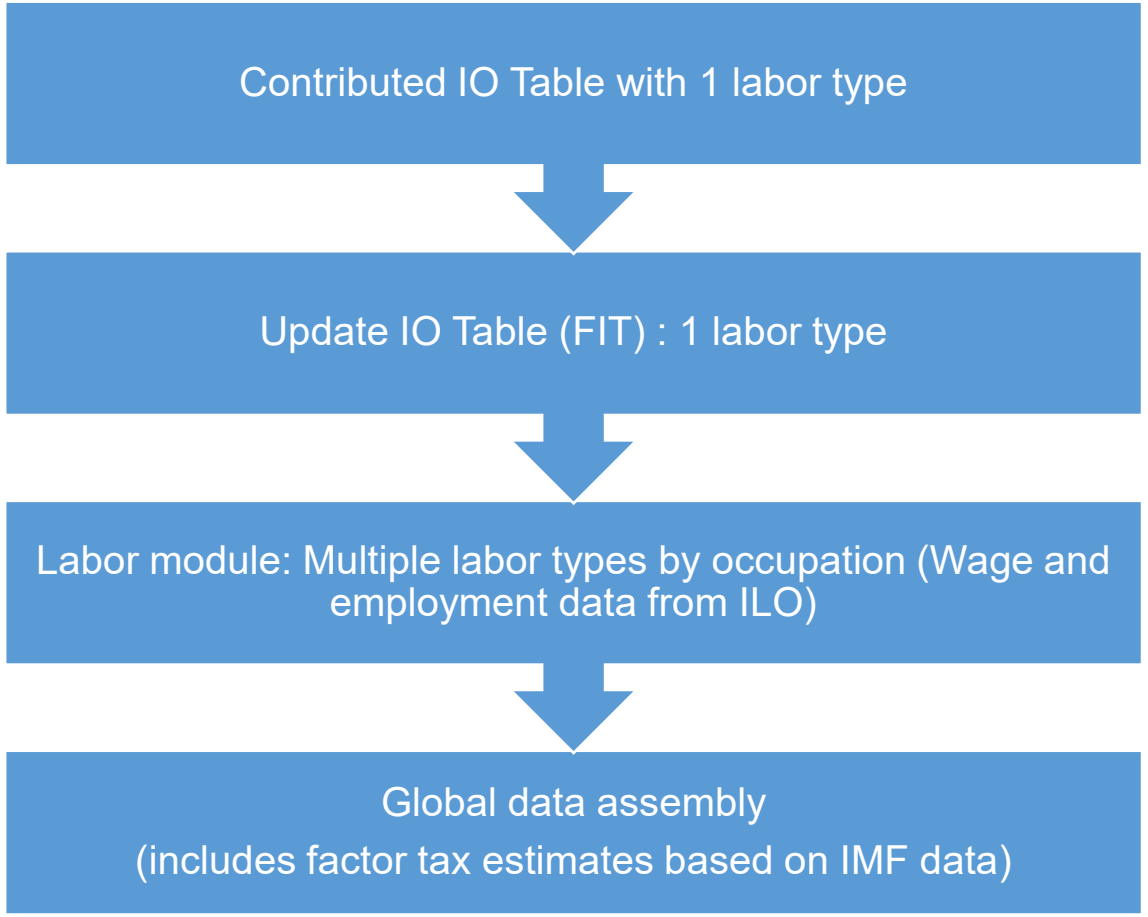
- **Update data extraction and processing (Years 1997-2023, by Yutaro Inoue)**
  - Extract all available data in GFS and OECD Global Revenue Statistics Database
  - Convert from money values to ratios to GDP, with GDP from the IMF's International Financial Statistics(187 countries)
  - Fill in data for the missing countries by calculating average tax-to-GDP ratios across 187 countries, and applying them to 251 standard GTAP countries in the data build.

- **Implementation**
  - Map GFS tax classifications to GTAP endowments
  - Factor Taxes
    - Social security contributions: Labor taxes
    - Other factor employment taxes: All GTAP endowments (i.e., labor, capital, land and natural resources)
    - Uniform across industries
  - Income Taxes
    - Personal income: Labor income taxes
    - Corporate income tax: Non-labor income taxes
  - Finally, we assume GFS does not account for Domestic Support, so final income taxes are GFS income tax plus land and capital-based Domestic Support (Domestic Support part is recorded in header FBEP)

GTAP DB Labor classification and procedure

- **Version 4-8:**
  - 2 labor types
    - Skilled (professional)
    - Unskilled (production and farm labor)
  - Procedure
    - Labor Force Surveys and Censuses from 13 (originally 15) countries
    - Regression to predict skilled labor shares for countries/regions with no data

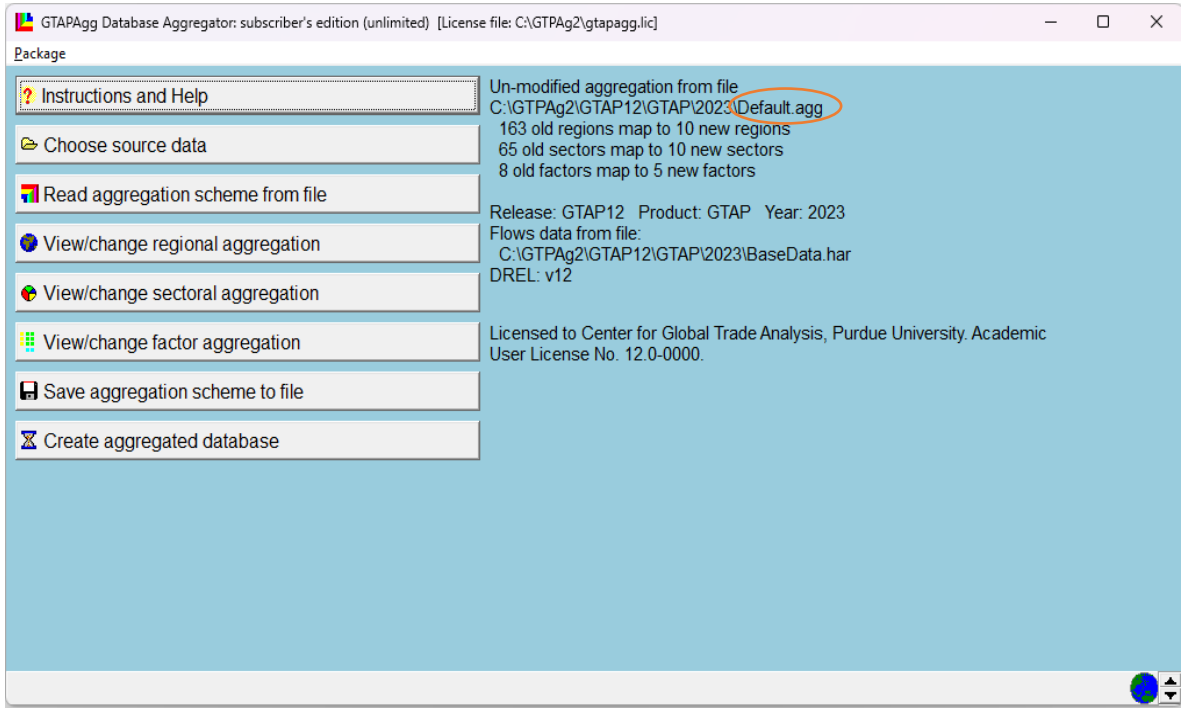
- **Version 9-present**
  - 5 Labor types by ILO occupation
    - 2 Skilled: Tech/Associate professionals and officials/managers/professionals
    - 3 Unskilled: Clerks, service/shop, agricultural and other workers
  - Details: Next slide

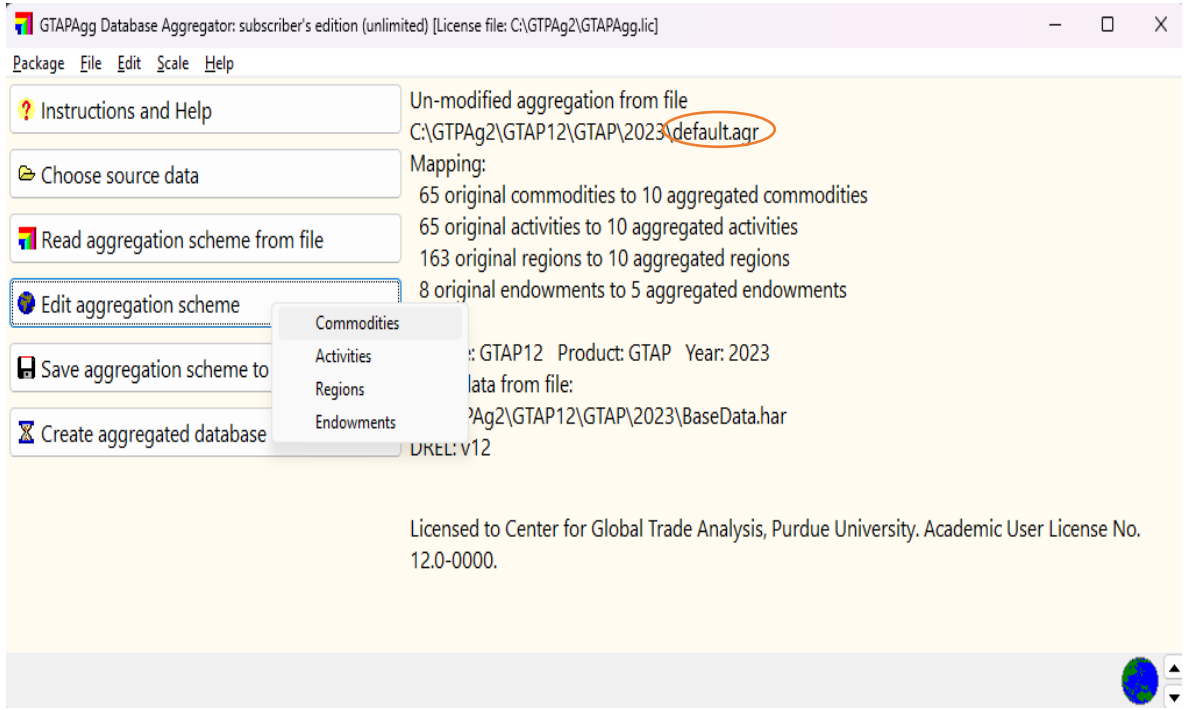


Labor data processing (since v11)

- **Updated ILO data extraction and processing (Years 2000-2023, by Puangchit Pattawee) using R software**
  - ilostat R package
  - Matched and merged wage and employment data: (a) wage by occupation (b) wage by industry; (c) employment by occupation and (d) employment by industry.
- **Updated wage and employment information for 70, mostly developing, countries.**
  - Wage optimization with degree of belief parameter (Tsigas and Weingarden, 2010) to impute wages for 5 occupations and 17 industries (ISIC Rev 4: A,B,C,D,E,F,G,H,I,J,K,L,MN,O,P,Q,RST)

GTAP Data Aggregation Utility (GTAPAgg2 and GTAPAgg3 by Mark Horridge and Dean Mustakinov, CoPS, Victoria University)





Global Trade Analysis Project

Labor splits, Income/factor taxes, protection, GTAPAgg3

Erwin Corong

ecorong@purdue.edu

Global Trade Analysis Project

Closing Remarks

Channing Arndt
carndt@purdue.edu

Global Trade Analysis Project

GTAP 12 Data Base

Virtual Seminar Series, Vol 7, No 1 (2026)

Center for Global Trade Analysis
Department of Agricultural Economics
Purdue University