



GLOBAL CHANGE INTERSECTORAL MODELING SYSTEM

# Bridging the Gap between Land and Food: Leveraging FAO Supply Utilization Accounts to Enhance Food System Modeling

June 5, 2024

Xin Zhao

Maksym Chepeliev, Pralit Patel

Page Kyle, Marshall Wise

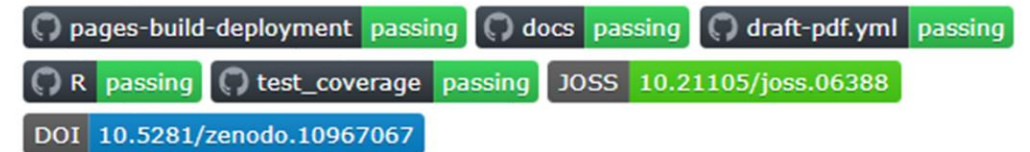
Joint Global Change Research Institute (JGCRI)



PNNL is operated by Battelle for the U.S. Department of Energy



- R package: *gcamfaostat* (v1.0.0)
  - Facilitate the use of FAOSTAT data



pages-build-deployment passing docs passing draft-pdf.yml passing  
R passing test\_coverage passing JOSS 10.21105/joss.06388  
DOI 10.5281/zenodo.10967067

**gcamfaostat** : An R package to prepare, process, and synthesize FAOSTAT data for global agroeconomic and multisector dynamic modeling

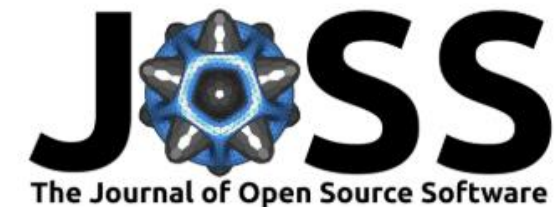
- As a function in *gcamfaostat*, we compile Food Balance Sheets using detailed FAO Supply Utilization Accounts
  - Connect land and food: a new approach for primary equivalent

- FAOSTAT is the major source of data for global agroeconomic modeling.
  - Agriculture development, trade, food security, nutrient, dietary health, environment, etc.
- Data need to be downloaded, cleaned, processed, and aggregated for modeling.
- Every modeling team is doing this with different assumptions.
  - Check out section #41 for AgMIP
- All the modeling groups can benefit from the concerted efforts of the open-source community.
- Need for traceable, transparent, consistent, and easy-to-update tool.

The screenshot displays the FAOSTAT website, which is part of the Food and Agriculture Organization of the United Nations. The header includes the FAO logo and the text 'Food and Agriculture Organization of the United Nations'. Below the header, there are navigation links in multiple languages (Arabic, Chinese, English, French, Russian, Spanish) and a search bar. The main content area is titled 'FAOSTAT' and features a 'Data' section. Under 'Data', there are two tabs: 'DOMAINS' and 'DOMAINS TABLE'. The 'DOMAINS' tab is active, showing a list of various data domains. These domains are organized into two columns. The left column includes: Production, Food Security and Nutrition (with an SDG indicator icon), Food Balances (with a dropdown arrow), Trade, Prices, Cost and Affordability of a Healthy Diet, Food and Diet, and Land, Inputs and Sustainability. The right column includes: Population and Employment, Investment (with an SDG indicator icon), Macro-Economic Indicators, Food Value Chain, Climate Change: Agrifood systems emissions, Forestry, SDG Indicators (with an SDG indicator icon), World Census of Agriculture, and Discontinued archives and data series. The 'Food Balances' domain is expanded, showing a list of specific data series: Food Balances (2010-), Supply Utilization Accounts (2010-), Commodity Balances (non-food) (2010-), Food Balances (-2013, old methodology and population), and Commodity Balances (non-food) (-2013, old methodology).

# And here is one: **gcamfaostat**

- We developed an R package
- **GCAM**
  - Widely used PE/MSD/IAM model maintained by PNNL-JGCRI
  - R package: **gcamdata**
    - ✓ Translate data into formats (XMLs) GCAM needs
    - ✓ Cool function to trace data *drake*
  - Will explain more later
- But the name doesn't matter
  - New features do



gcamfaostat: An R package to prepare, process, and synthesize FAOSTAT data for global agro-economic and multisector dynamic modeling

Xin Zhao <sup>1</sup>, Maksym Chepeliev <sup>2</sup>, Pralit Patel <sup>1</sup>, Marshall Wise <sup>1</sup>, Katherine Calvin <sup>1</sup>, Kanishka Narayan <sup>1</sup>, and Chris Vernon <sup>1</sup>

<sup>1</sup> Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA <sup>2</sup> Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University, West Lafayette, IN, USA

JOSS: <https://joss.theoj.org/papers/10.21105/joss.06388>

GitHub: <https://github.com/JGCRI/gcamfaostat>

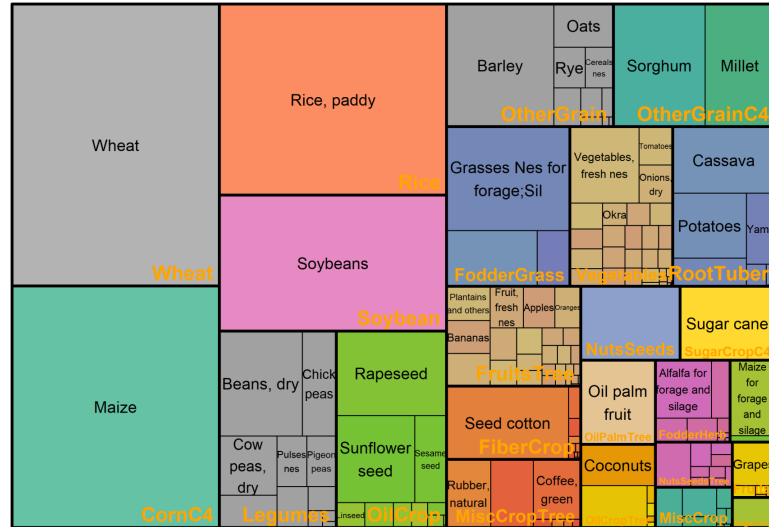
Website: <https://jgcri.github.io/gcamfaostat/>

# IMS

GLOBAL CHANGE INTERSECTORAL MODELING SYSTEM

## Some functions Visualization

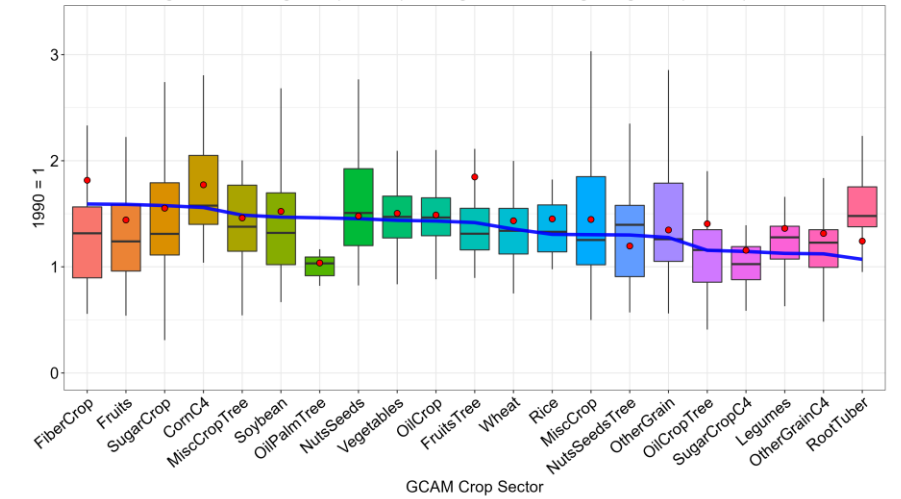
World area harvested share by GCAM-FAO crops in 2010-2019 (mean = 1570 Mha)



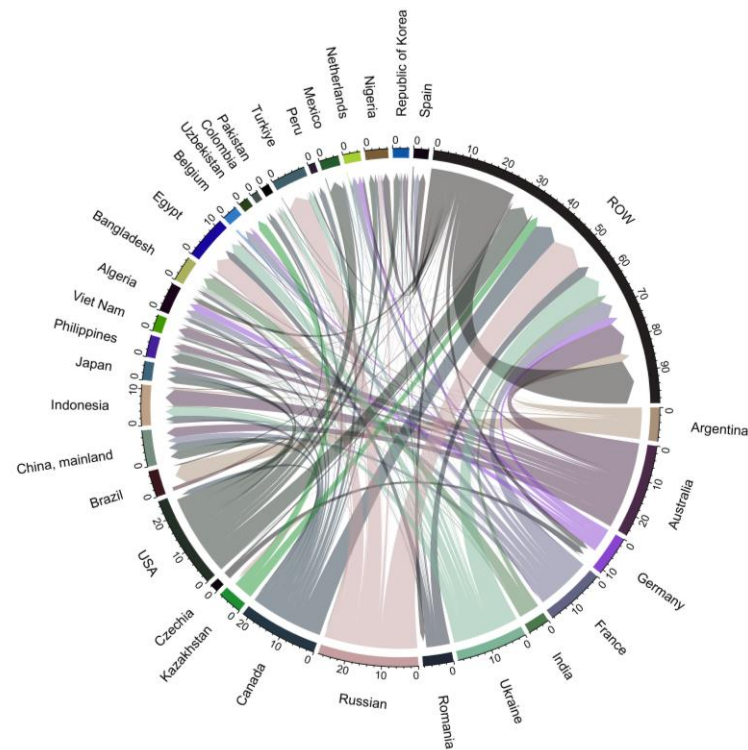
Per ca. per day food calories availability by region and commodity, 2013-2017 mean (world: 2902 Kcal/ca/d)



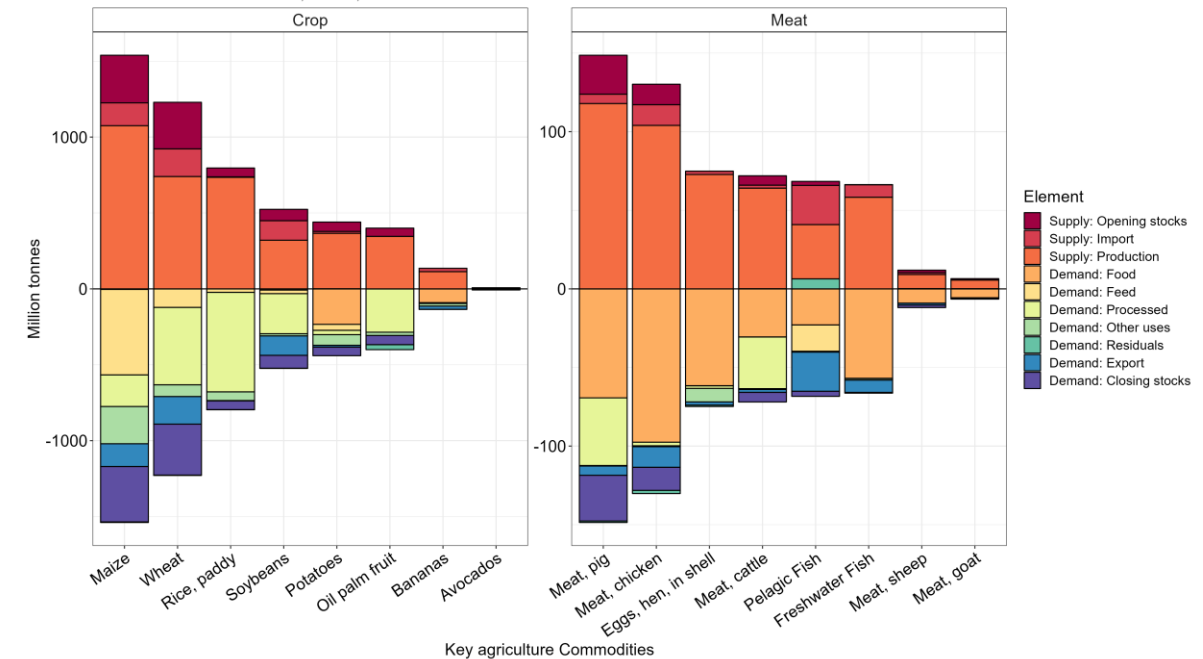
Yield growth in 1990 - 2020, distribution (Tukey style) across GCAM regions.  
Extensive margin: world mean growth (blue line) vs. weighted mean of regional growth (red dots)



Data: FAOSTAT  
Package: JGCR/gcamfaostat

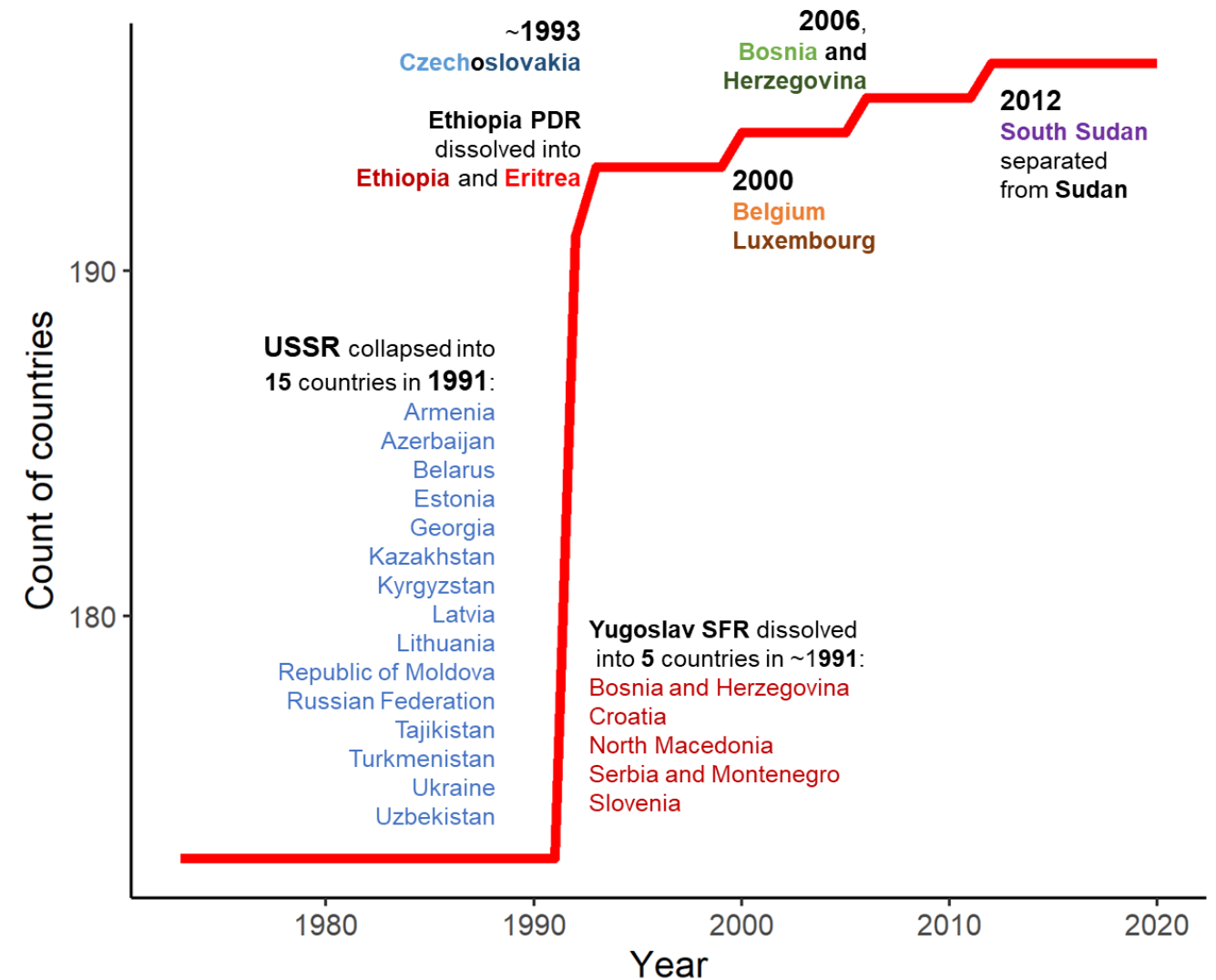


World supply utilization for key commodities, 2013 - 2017 mean  
Other uses include seed use, losses, and industrial use



Data: FAOSTAT  
Package: JGCR/gcamfaostat

- Still remember the pain of dealing with region/country mappings
- Not even mentioning the name changes of countries/sector
  - area\_code & item\_code (or CPC) are prioritized



# FAO data currently included in gcamfaostat

- *gcamfaostat\_metadata()*

**Table 1. FAOSTAT dataset processed in gcamfaostat**

| Dataset Code | Dataset Name                                                         | Exist_Local | Exist_Prebuilt | FAO update date | FAO size |
|--------------|----------------------------------------------------------------------|-------------|----------------|-----------------|----------|
| CB           | Food Balances: Commodity Balances (non-food) (2010-)                 | TRUE        | TRUE           | 2023-08-08      | 1MB      |
| FBSH         | Food Balances: Food Balances (-2013, old methodology and population) | TRUE        | TRUE           | 2023-03-10      | 69MB     |
| OA           | Population and Employment: Annual population                         | TRUE        | TRUE           | 2022-10-24      | 2MB      |
| FO           | Forestry: Forestry Production and Trade                              | TRUE        | TRUE           | 2023-12-21      | 15MB     |
| QCL          | Production: Crops and livestock products                             | TRUE        | TRUE           | 2023-12-18      | 32MB     |
| PD           | Prices: Deflators                                                    | TRUE        | TRUE           | 2023-11-07      | 1MB      |
| TCL          | Trade: Crops and livestock products                                  | TRUE        | TRUE           | 2023-12-13      | 244MB    |
| FBS          | Food Balances: Food Balances (2010-)                                 | TRUE        | TRUE           | 2023-09-04      | 55MB     |
| RFN          | Land, Inputs and Sustainability: Fertilizers by Nutrient             | TRUE        | TRUE           | 2023-07-05      | 2MB      |
| RL           | Land, Inputs and Sustainability: Land Use                            | TRUE        | TRUE           | 2024-02-15      | 2MB      |
| PP           | Prices: Producer Prices                                              | TRUE        | TRUE           | 2023-12-19      | 10MB     |
| SCL          | Food Balances: Supply Utilization Accounts (2010-)                   | TRUE        | TRUE           | 2023-09-04      | 72MB     |
| TM           | Trade: Detailed trade matrix                                         | TRUE        | TRUE           | 2023-11-23      | 446MB    |



Food and Agriculture Organization  
of the United Nations

Home

Data

Selected Indicators

Compare Data

Rankings

Definitions and Standards

FAQ

Search an Indicator or Commodity

# FAOSTAT

## Data

DOMAINS
DOMAINS TABLE

▶

Production

▶

Food Security and Nutrition

SDG indicators

▼

Food Balances

Food Balances (2010-)
Supply Utilization Accounts (2010-)
Commodity Balances (non-food) (2010-)
Food Balances (-2013, old methodology and population)
Commodity Balances (non-food) (-2013, old methodology)

▶

Trade

▶

Prices

▶

Cost and Affordability of a Healthy Diet

▶

Food and Diet

▶

Land, Inputs and Sustainability

▶

Population and Employment

▶

Investment

SDG indicator

▶

Macro-Economic Indicators

▶

Food Value Chain

▶

Climate Change: Agrifood systems emissions

▶

Forestry

▶

SDG Indicators


▶

World Census of Agriculture

▶

Discontinued archives and data series

- *gcamfaostat\_metadata()*

**Table 1. FAOSTAT dataset processed in gcamfaostat**

| Dataset Code | Dataset Name                                                         | Exist_Local | Exist_Prebuilt | FAO update date | FAO size |
|--------------|----------------------------------------------------------------------|-------------|----------------|-----------------|----------|
| CB           | Food Balances: Commodity Balances (non-food) (2010-)                 | TRUE        | TRUE           | 2023-08-08      | 1MB      |
| FBSH         | Food Balances: Food Balances (-2013, old methodology and population) | TRUE        | TRUE           | 2023-03-10      | 69MB     |
| OA           | Population and Employment: Annual population                         | TRUE        | TRUE           | 2022-10-24      | 2MB      |
| FO           | Forestry: Forestry Production and Trade                              | TRUE        | TRUE           | 2023-12-21      | 15MB     |
| QCL          | Production: Crops and livestock products                             | TRUE        | TRUE           | 2023-12-18      | 32MB     |
| PD           | Prices: Deflators                                                    | TRUE        | TRUE           | 2023-11-07      | 1MB      |
| TCL          | Trade: Crops and livestock products                                  | TRUE        | TRUE           | 2023-12-13      | 244MB    |
| FBS          | Food Balances: Food Balances (2010-)                                 | TRUE        | TRUE           | 2023-09-04      | 55MB     |
| RFN          | Land, Inputs and Sustainability: Fertilizers by Nutrient             | TRUE        | TRUE           | 2023-07-05      | 2MB      |
| RL           | Land, Inputs and Sustainability: Land Use                            | TRUE        | TRUE           | 2024-02-15      | 2MB      |
| PP           | Prices: Producer Prices                                              | TRUE        | TRUE           | 2023-12-19      | 10MB     |
| SCL          | Food Balances: Supply Utilization Accounts (2010-)                   | TRUE        | TRUE           | 2023-09-04      | 72MB     |
| TM           | Trade: Detailed trade matrix                                         | TRUE        | TRUE           | 2023-11-23      | 446MB    |


**Food and Agriculture Organization of the United Nations**

العربية
中文
English
Français
Русский
Español

FAOSTAT

[Home](#)
[Data](#)
[Selected Indicators](#)
[Compare Data](#)
[Rankings](#)
[Definitions and Standards](#)
[FAQ](#)

Data

[DOMAINS](#)
[DOMAINS TABLE](#)

▶ Production

▶ Food Security and Nutrition SDG indicators

▶ Trade

▶ Prices

▶ Cost and Affordability of a Healthy Diet

▶ Food and Diet

▶ Land, Inputs and Sustainability

▶ Population and Employment

▶ Investment SDG indicator

▶ Macro-Economic Indicators

▶ Food Value Chain

▶ Climate Change: Agrifood systems emissions

▶ Forestry

▶ SDG Indicators

▶ World Census of Agriculture

▶ Discontinued archives and data series

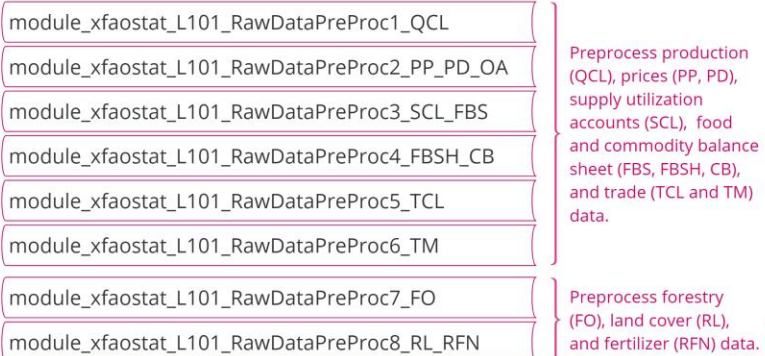
▶ Food Balances

Food Balances (2010-)  
Supply Utilization Accounts (2010-)  
Commodity Balances (non-food) (2010-)  
Food Balances (-2013, old methodology and population)  
Commodity Balances (non-food) (-2013, old methodology)

# Data processing modules (functions)

- *driver\_drake()*
  - Runs data processing modules sequentially to generate intermediate data outputs and final output
  - Trace data flow

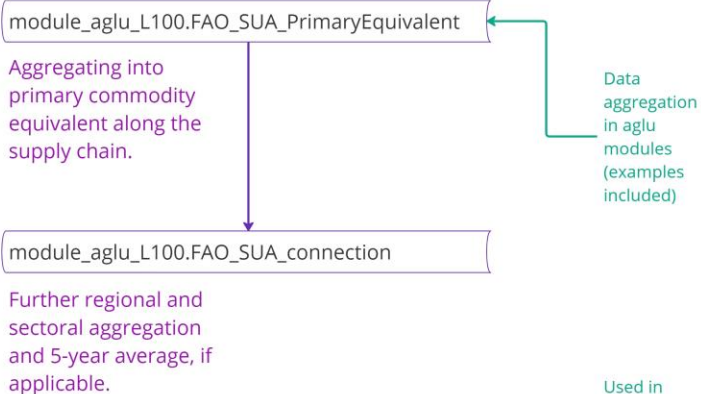
## 1. gcamfaostat preprocessing modules



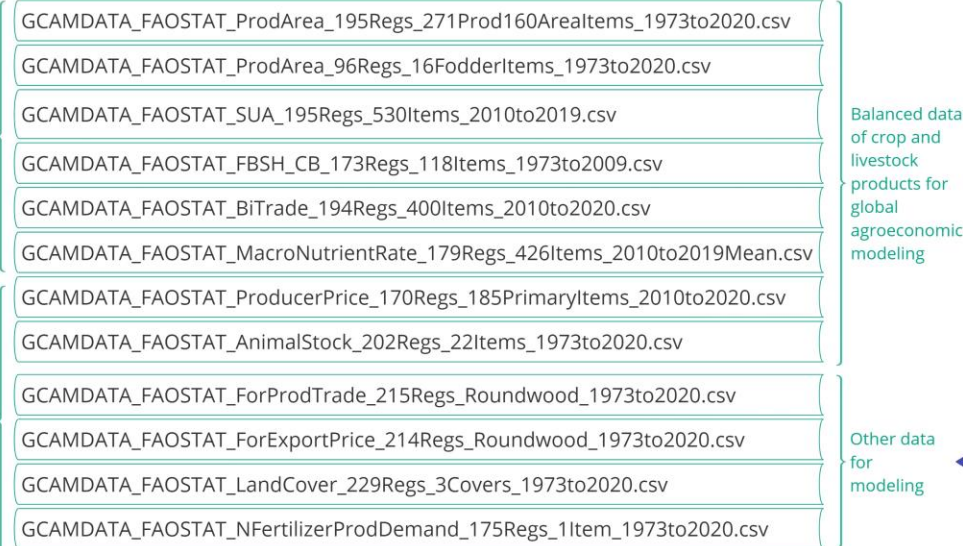
## 2. gcamfaostat data processing and balancing modules



## 4. gcamfaostat/gcamdata data aggregation for modeling



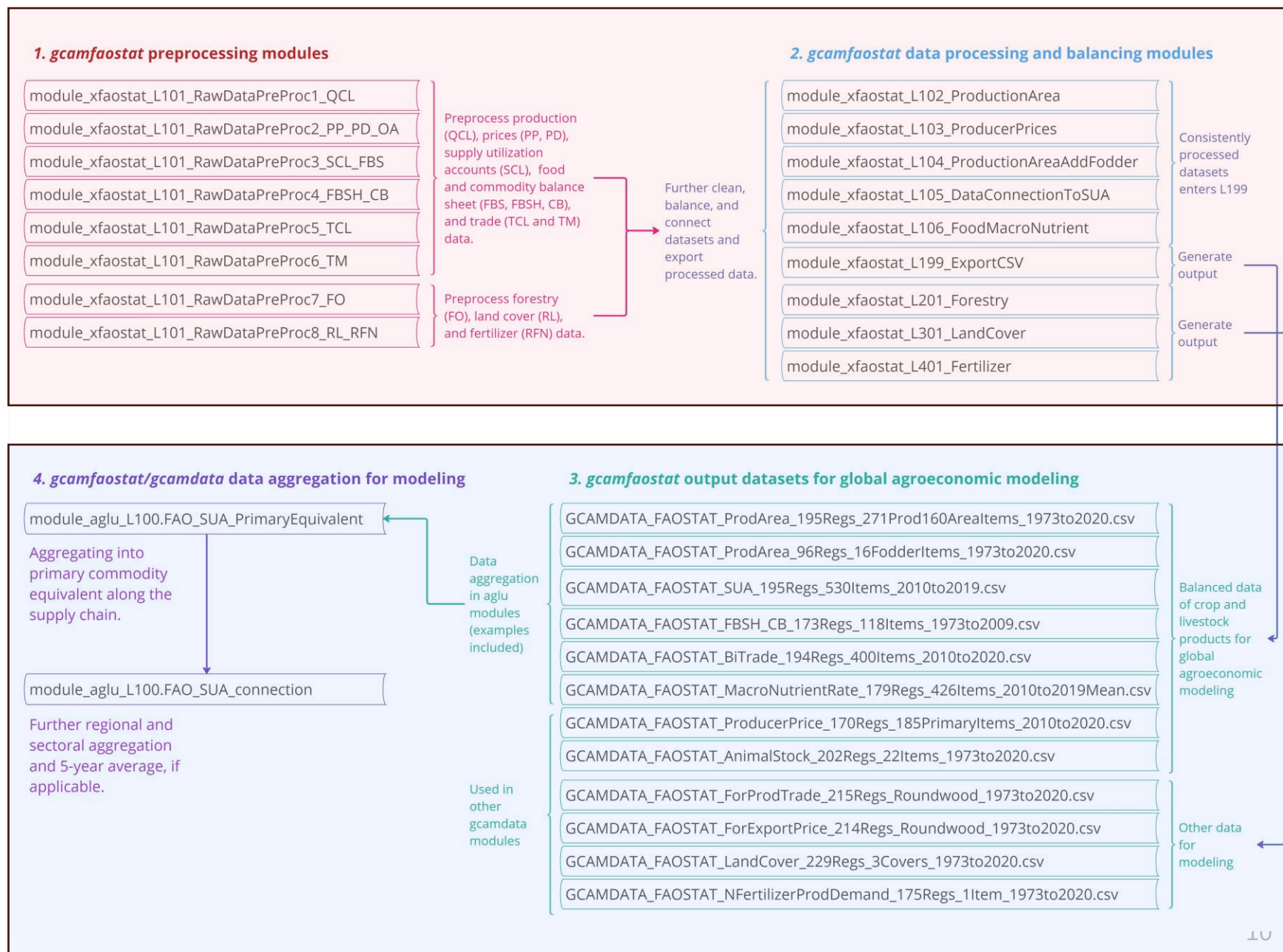
## 3. gcamfaostat output datasets for global agroeconomic modeling



# Data processing modules (functions)

- *driver\_drake()*
  - Runs data processing modules sequentially to generate intermediate data outputs and final output
  - Trace data flow

## Common modules:



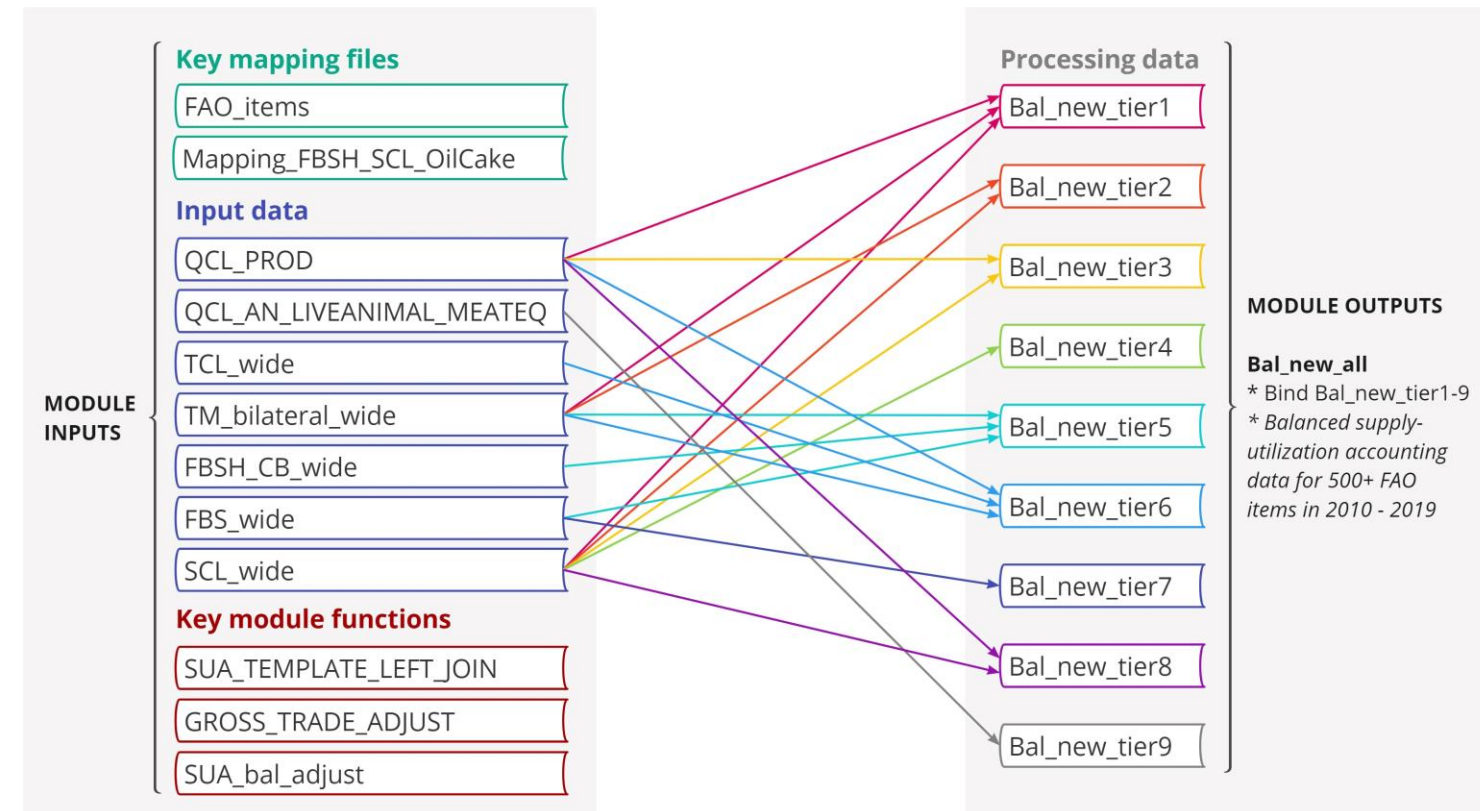
**GCAM-specific data/modules:**

- Aggregation
- Average

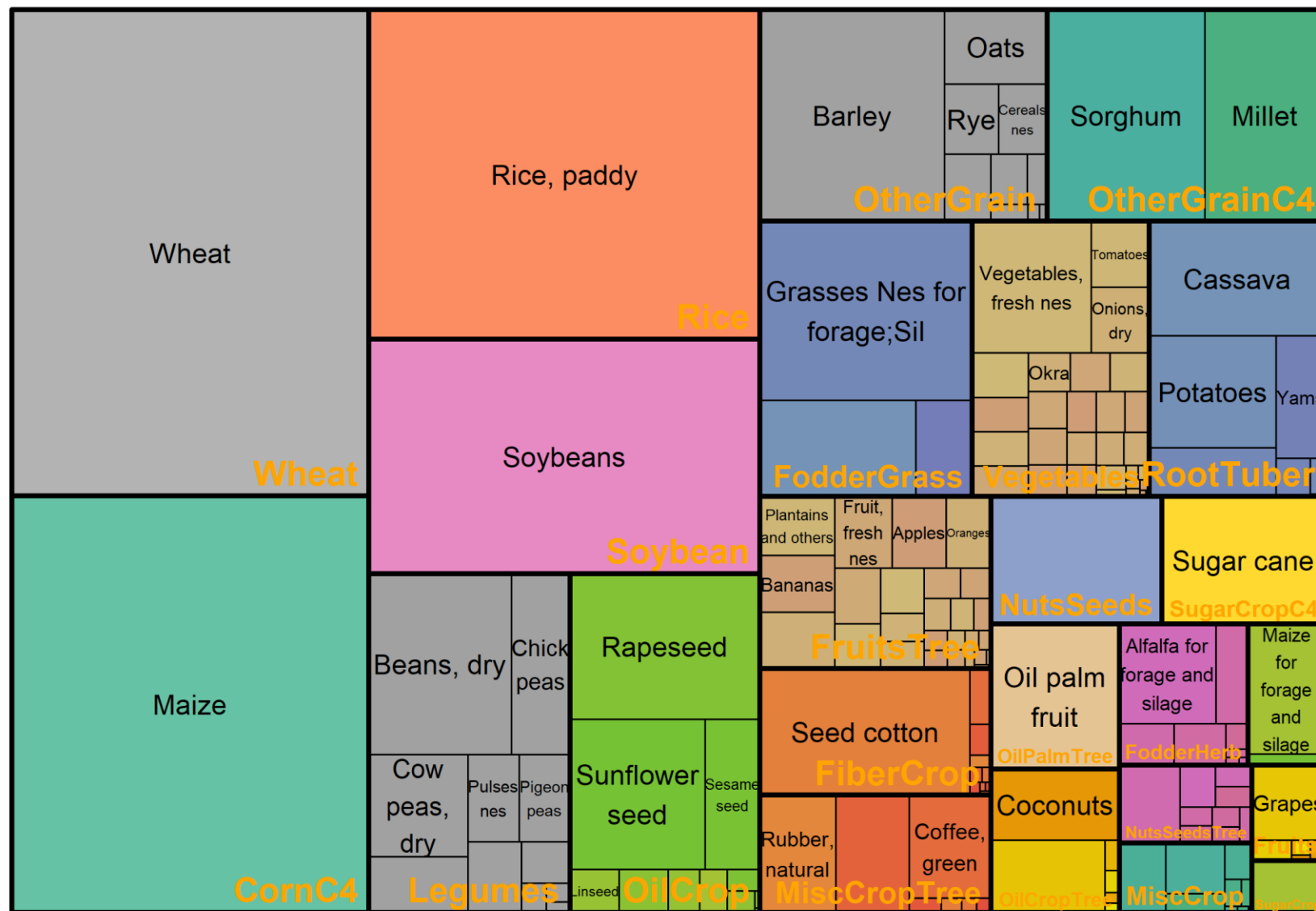
## gcamfaostat goals

- Encourage community collaboration and improve efficiency
  - Reduce the effort required for data processing.
- Facilitate FAOSTAT data use in global economic and multisector dynamic models
  - Ensure transparency, traceability, and reproducibility.
- Enhance the quality and accessibility of data for the global agroeconomic modeling community
  - Reduce model uncertainty
  - Foster more robust and harmonized outcomes in a collaborative, efficient, and open-source framework.

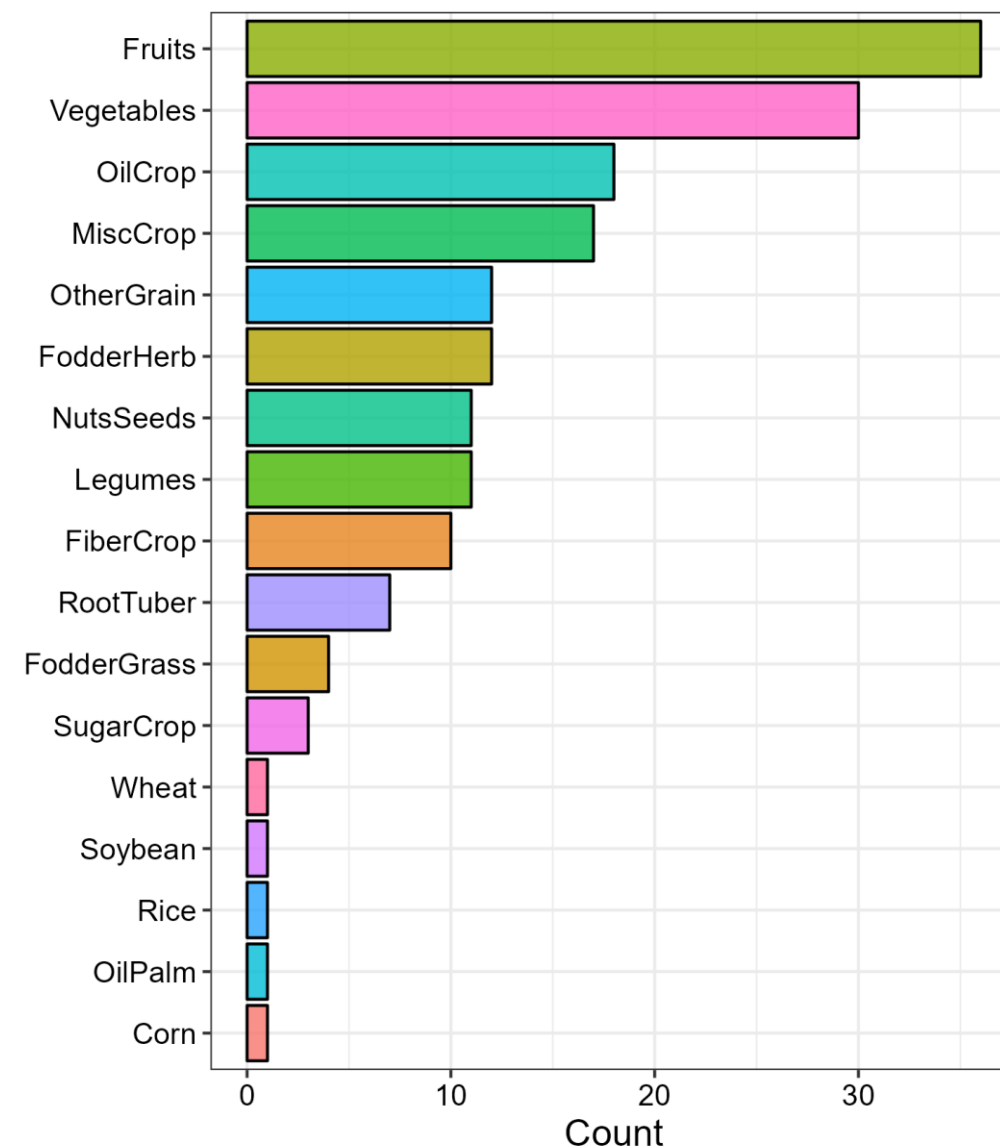
- Harmonize various FAOSTAT datasets to generate a consistent set of agricultural supply and utilization accounts (SUA) data
  - Over 500 commodities
  - With different quality/assumptions
- In comparison, FBS (food balance sheet) had ~100 aggregated commodities
  - Lack of connections to bilateral trade and stocks
  - With no detailed documentation
- There will be another common module for the new primary equivalent approach
  - Aggregation across region/sector
  - Aggregation vertically along supply chains



World area harvested share by GCAM-FAO crops in 2010-2019 (mean = 1570 Mha)

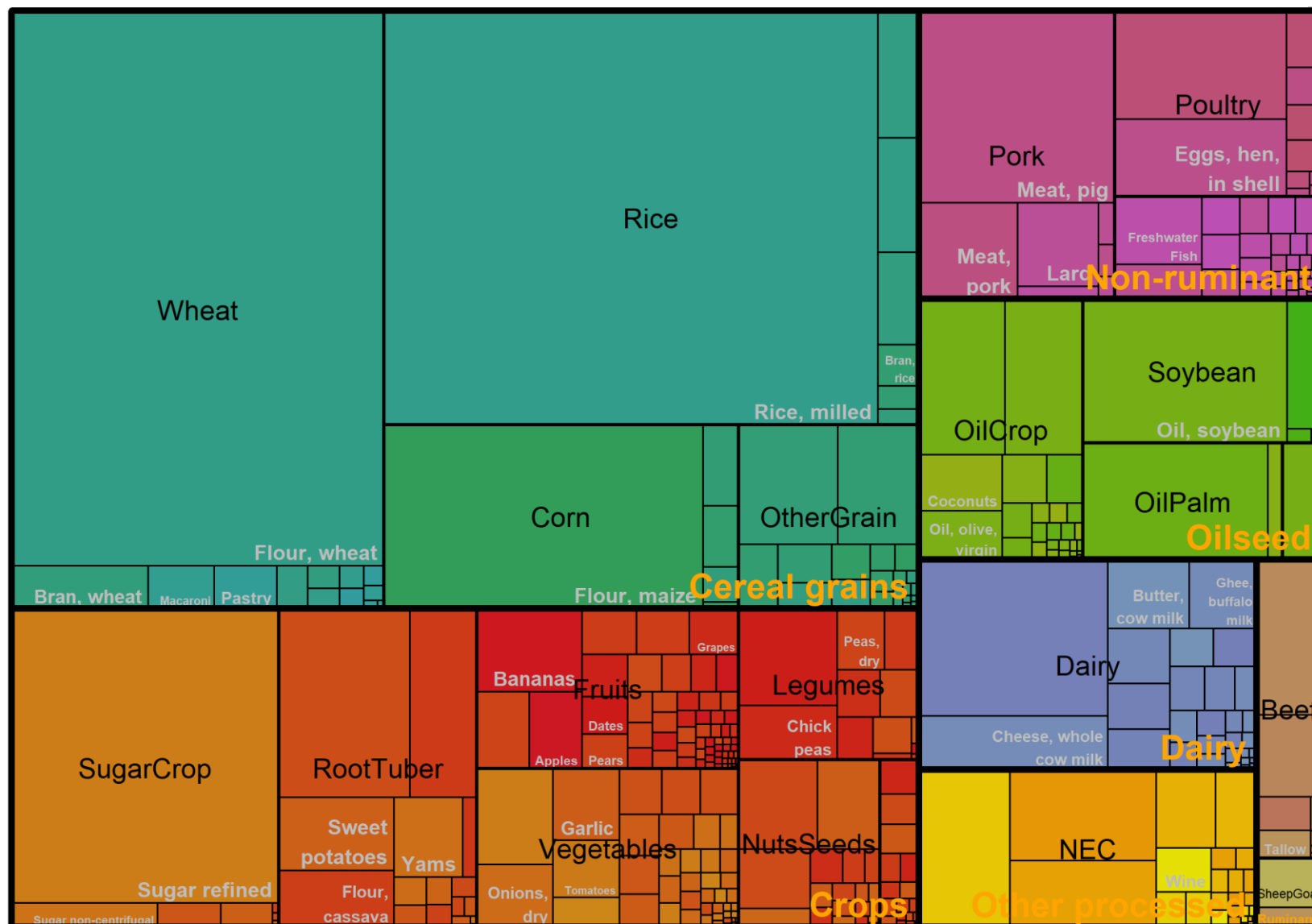


## 176 FAO food items to 17 GCAM items

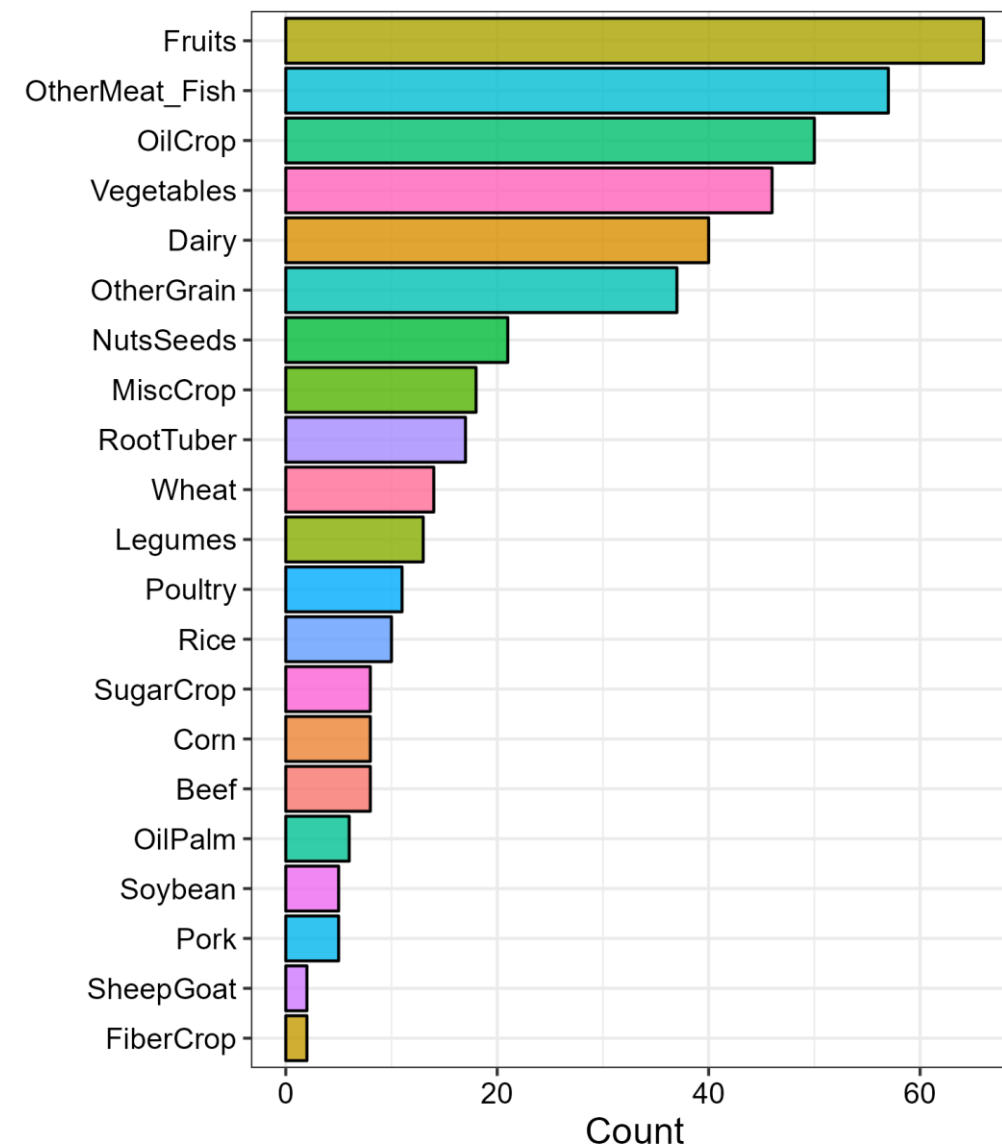


# Food Calories: Mapping 444 FAO food items to 21 GCAM items

World food consumption:  
7770 Peta-Kcal per year or ~2900 Kcal/ca./day



444 FAO food items to 21 GCAM items



# Challenges in connecting land to food

- Primary land-based output is usually further processed
- Models do not include the full processing chain
- Supply utilization accounting:



# Challenges in connecting land to food

- Primary land-based output is usually further processed
- Models do not include the full processing chain
- Supply utilization accounting:

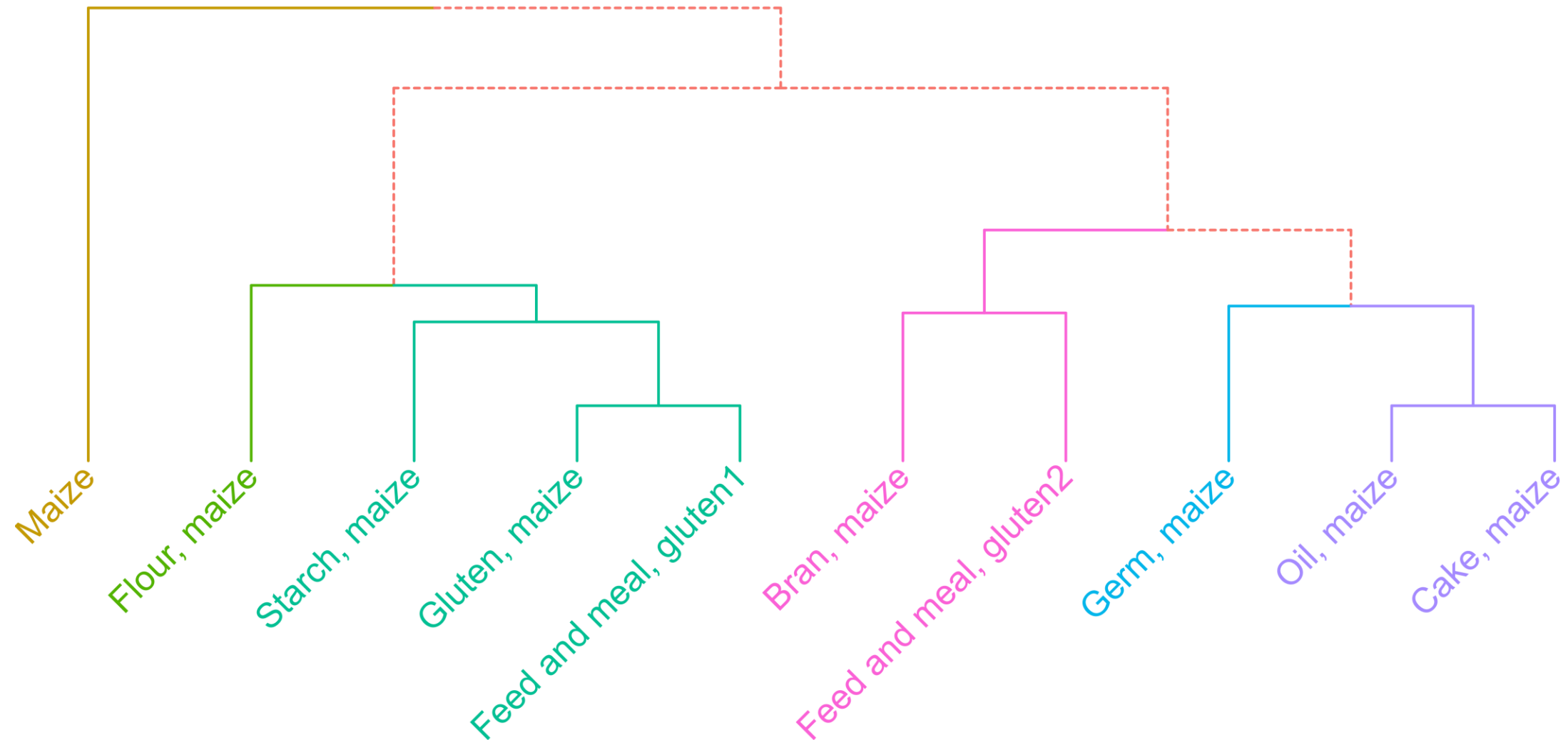


# Challenges in connecting land to food

- Primary land-based output is usually further processed
- Models do not include the full processing chain
- Supply utilization accounting:



# Maize example: Showing the depth of agri-food sectors

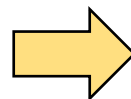


- Supply utilization account (SUA) balances for all commodities

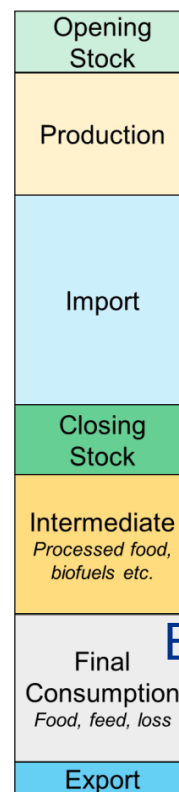
## Vertical aggregation: Primary Commodity Equivalent



Extraction  
rate



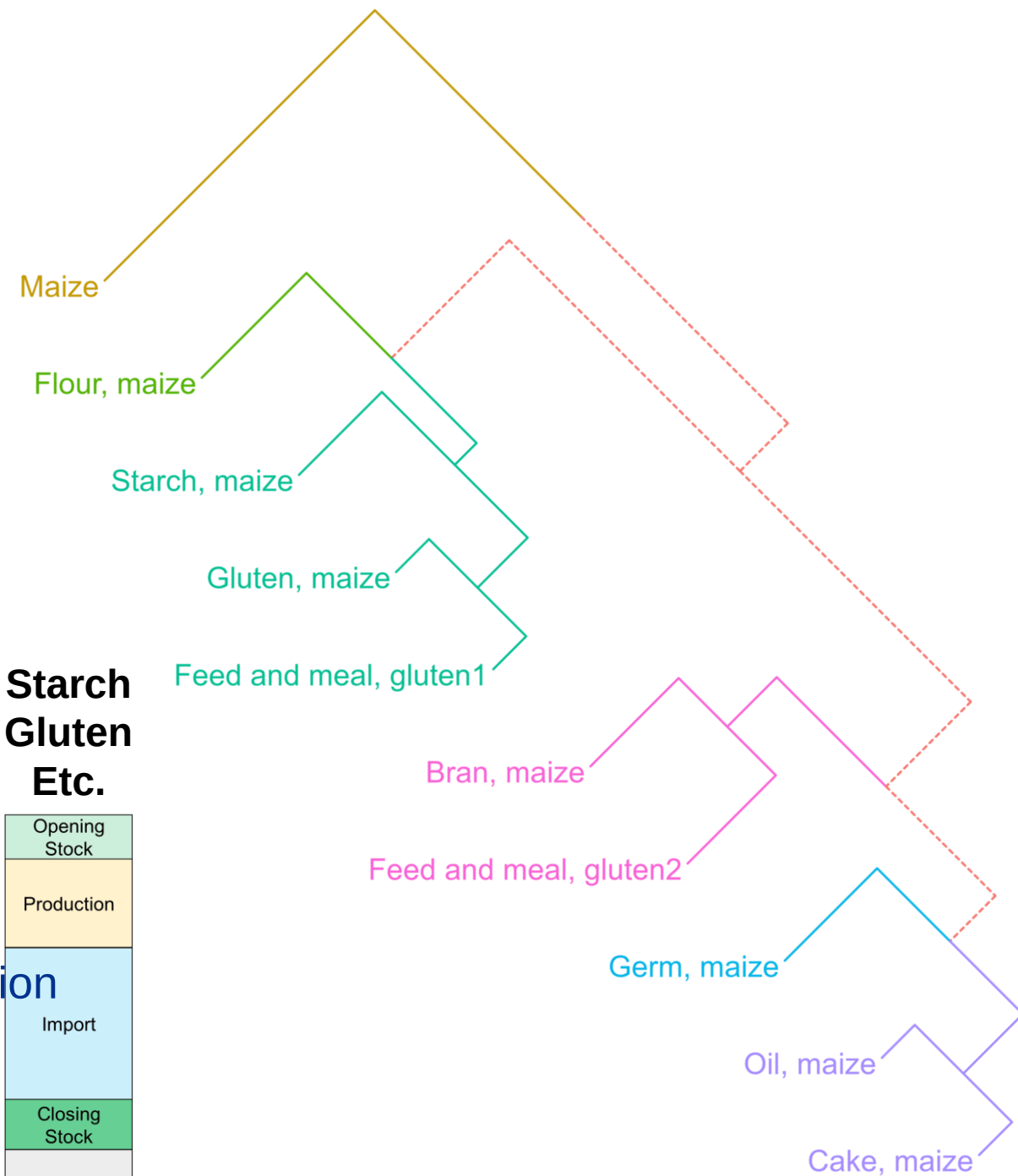
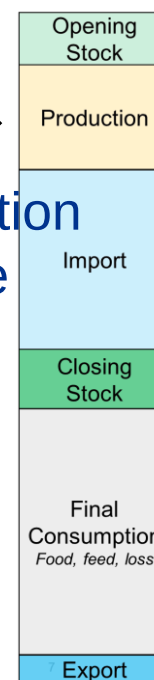
**Maize**



Extraction  
rate

**Flour  
Germ  
Bran**

**Starch  
Gluten  
Etc.**



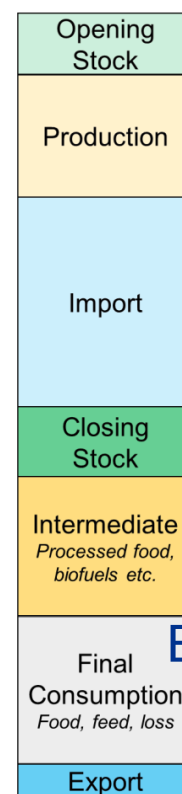
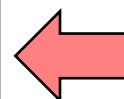
## Vertical aggregation: Primary Commodity Equivalent

- Consistency:
  - Space
  - Time
  - Supply - Utilization
- Preserve:
  - Yield
  - Macronutrients



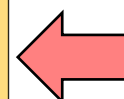
Maize

Extraction  
rate

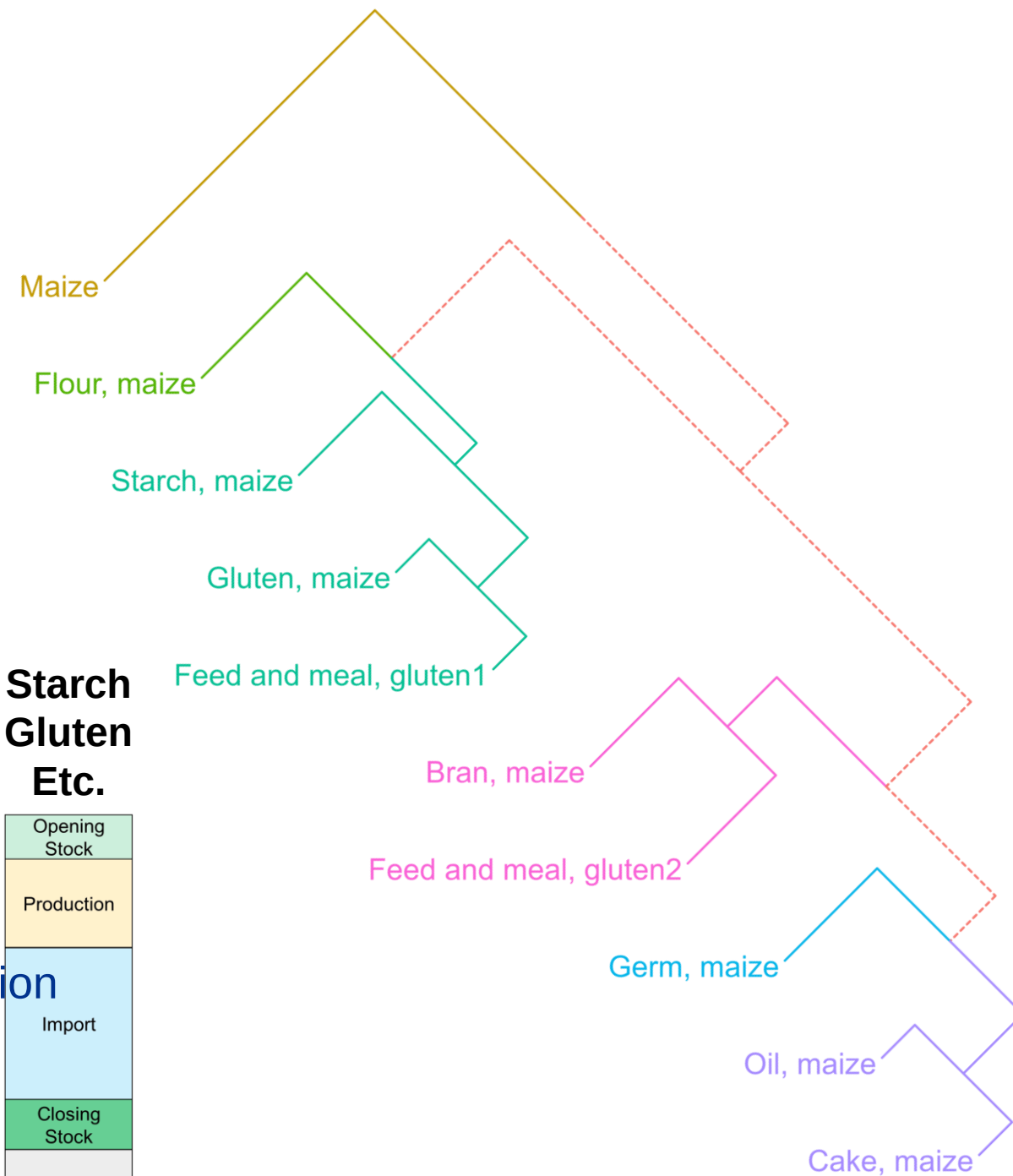
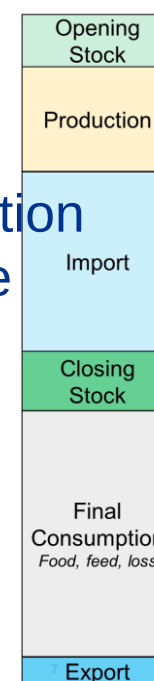


Flour  
Germ  
Bran

Extraction  
rate

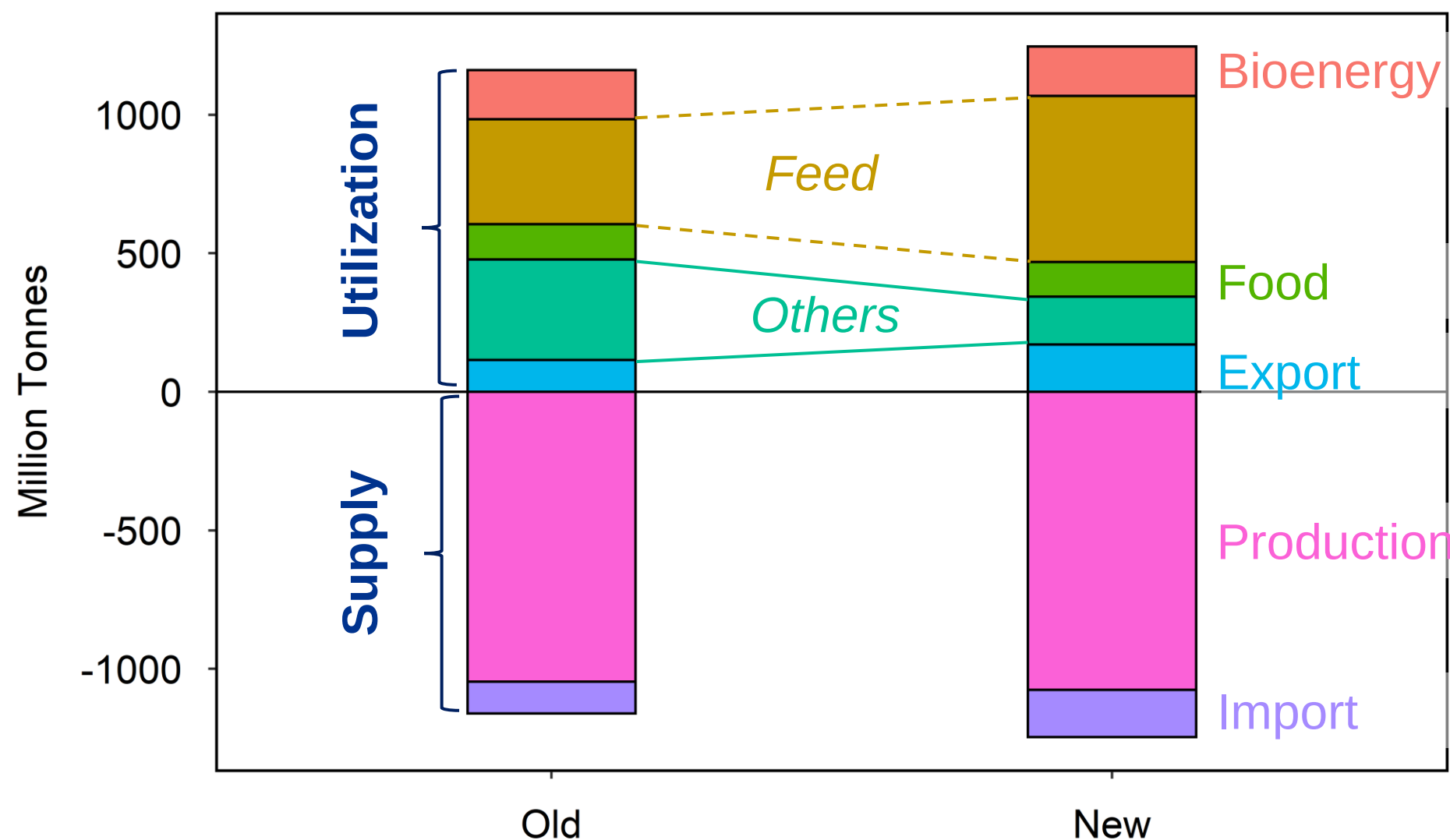


Starch  
Gluten  
Etc.



## Implications: Corn supply utilization balance

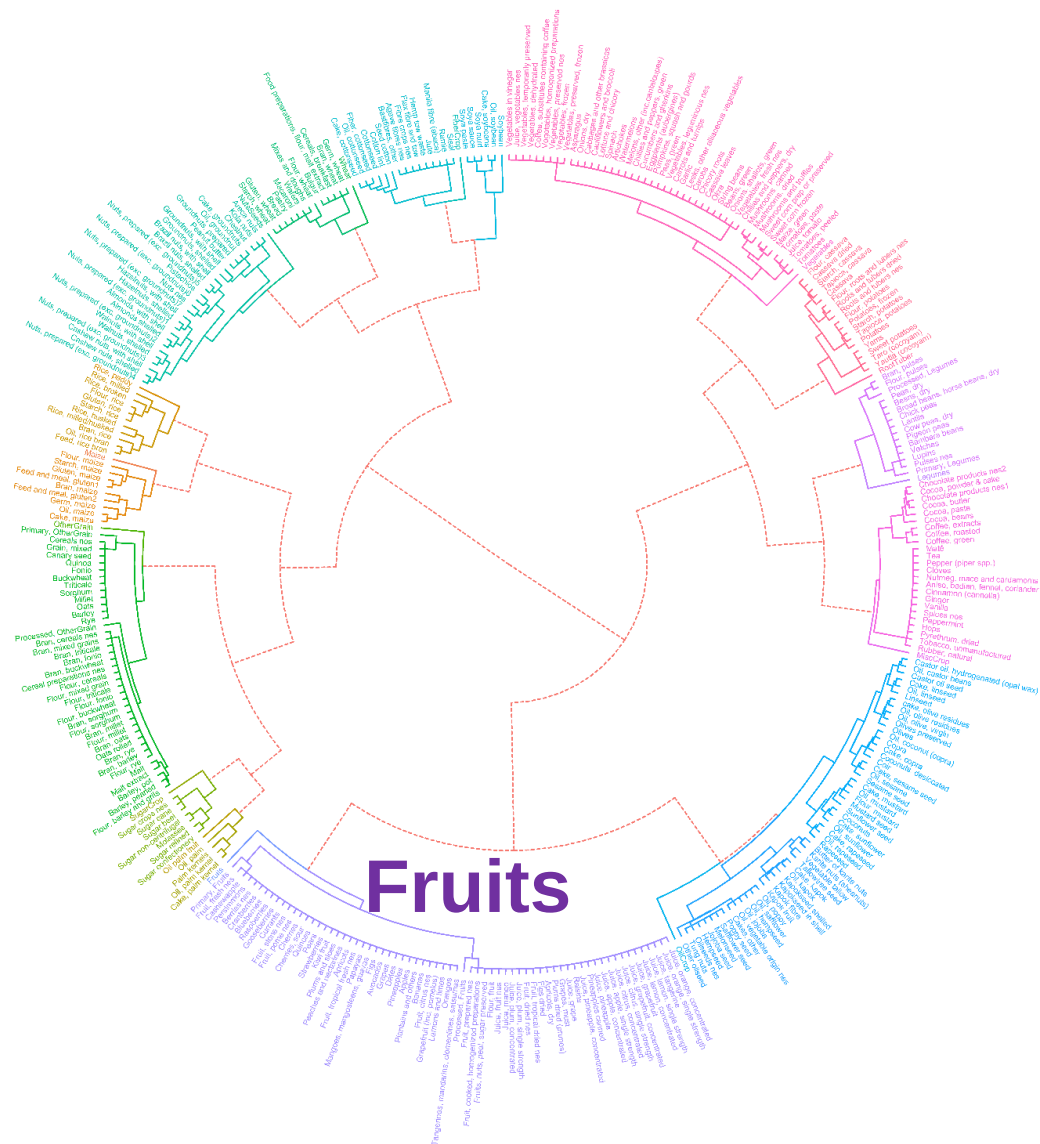
- GCAM base data: Corn SUA



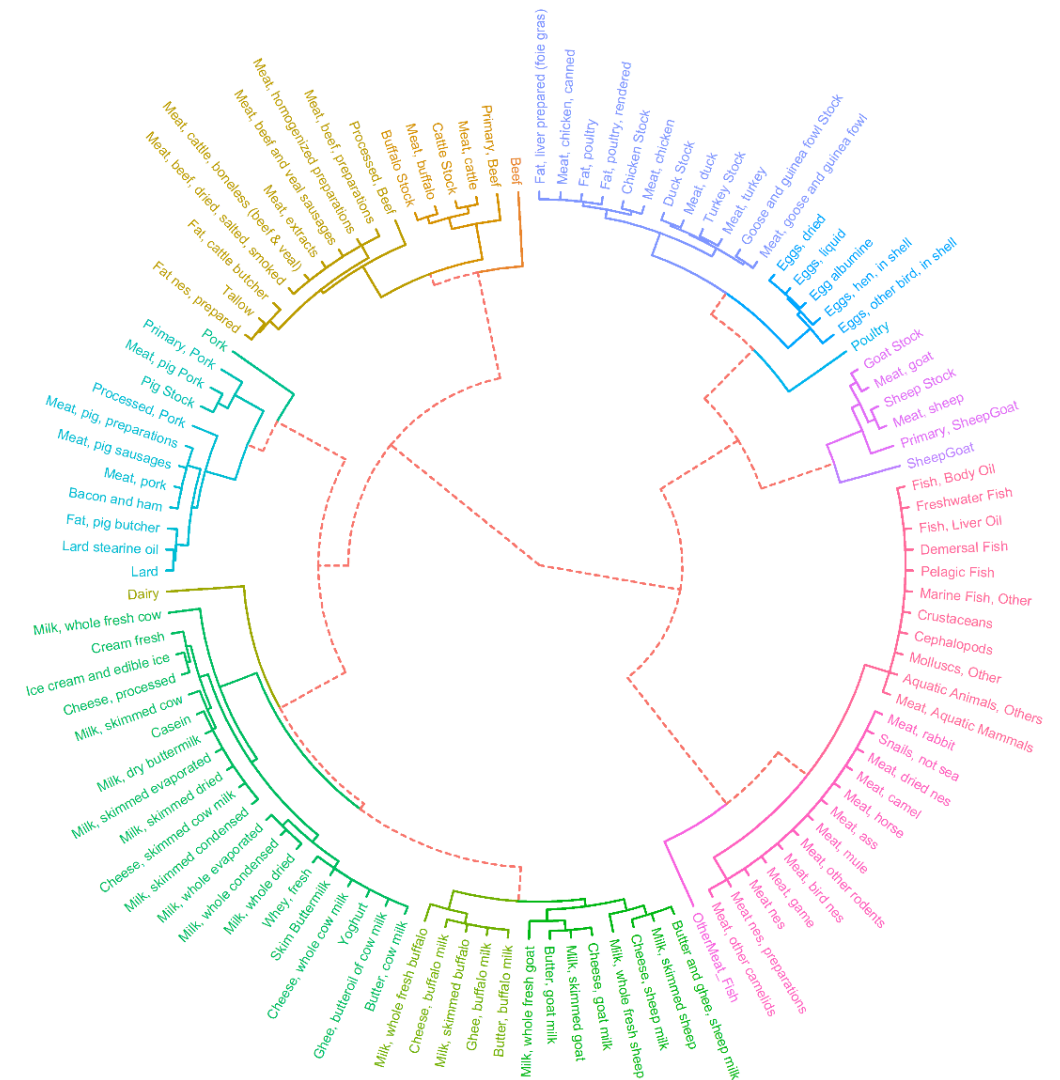
- Feed-use is larger (+60%)
- Other-use is much smaller (-50%)
- Similar patterns for other commodities
- Direct food: 165 kcal/ca/day

## Nested mapping structure

### 17 GCAM crop commodities



### 6 GCAM livestock commodities



# Summary of additional gcamfaostat features/functions

- Balanced Supply Utilization Accounting (SUA) dataset of over 500 agricultural and food commodities
- A nested commodity mapping to understand the relationship across all the commodities.
- A Primary Equivalent function that allows aggregating agriculture and food commodities along the supply chain.
- The approach is transparent, reproducible, consistent, and flexible.

## Why is the new data approach important?

- Bridges the gap between agricultural inputs (e.g., land) and final consumptions (e.g., food)
- Enhances data quality and scientific robustness in multisector global economic modeling
- Enables further disaggregation of additional regions, agricultural commodities, and processing sectors

## Why is the new data approach important?

- Bridges the gap between agricultural inputs (e.g., land) and final consumptions (e.g., food)
- Enhances data quality and scientific robustness in multisector global economic modeling
- Enables further disaggregation of additional regions, agricultural commodities, and processing sectors
- Accounts for storage and trade in secondary products
- Food is more than food calories
  - Macronutrients (protein and fat)
  - May include micronutrients in future work
- Improves the tracing of monetary and physical flows along supply chains
  - Value-added (labor & capital)
  - Water, nutrients, emissions, etc.

## Next steps and contribution

- More developments
  - Revise assumptions in low-quality data zones
  - Incorporate more data areas into the framework
- Data development is never a once-and-for-all task
  - Continued efforts are needed to sustain and improve the processing procedures
- Promoting broader applications
- We welcome and value community contributions to *gcamfaostat*
  - Decentralized development by users



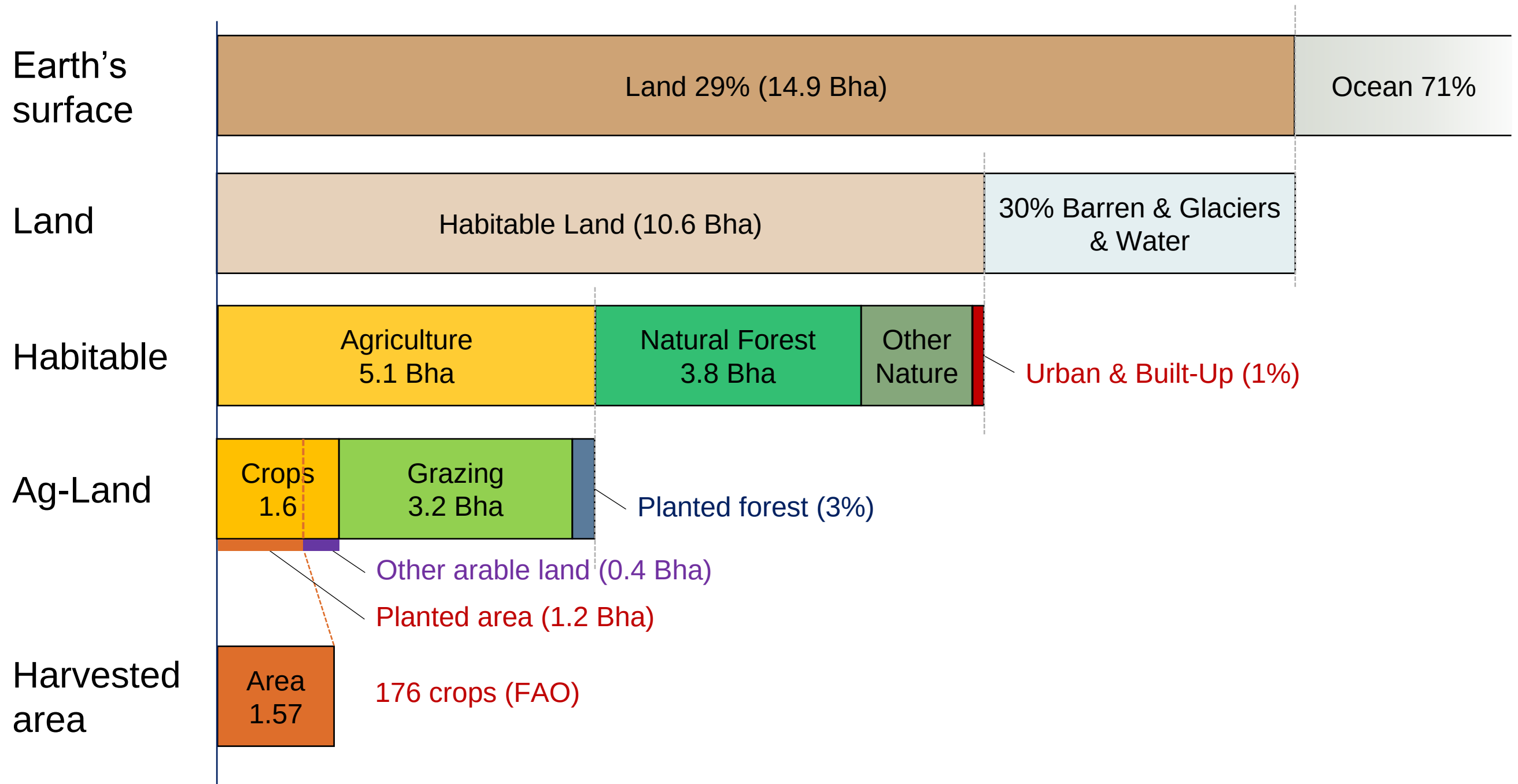
GLOBAL CHANGE INTERSECTORAL MODELING SYSTEM

We acknowledge the support from the U.S. Department of Energy, Office of Science, as part of research in the Multi-Sector Dynamics, Earth and Environmental System Modeling Program

**Thank you**



# GCAM represents all land



# Balancing data across all dimensions

## Balance across supply – utilization (market clearing)

- $Opening\ stock_{reg,t} + Production_{reg,t} + Import_{reg,t} = Closing\ stock_{reg,t} + Consumption_{reg,t} + Export_{Reg,t}$

## Balance across space

- $\sum_{reg} Import_{reg} = \sum_r Export_{reg}$
- $\sum_{reg} Net\ trade_{reg}$

## Balance in primary production input to output

- $Crop\ yield = \frac{primary\ output}{harvested\ area}$
- $Feed\ conversion\ rate = \frac{feed\ input}{livestock\ product\ output}$

## Balance across time

- $Opening\ stock_t = Closing\ stock_{t-1}$
- $Opening\ stock_t + Stock\ variation_t = Closing\ stock_t$

## Balance along supply chains

- $extraction\ rate = \frac{senondary\ production}{primary\ processed\ used}$

## Balance in food availability and macronutrients

- *Tracing for calories, protein, and fat in food tonnes*