Core Model Proposal #360: GCAM agriculture and land use (AgLU) data and method updates: connecting land hectares to food calories

Product: Global Change Analysis Model (GCAM)

Institution: Joint Global Change Research Institute (JGCRI)

Authors: Xin Zhao and Marshall Wise

Reviewers: Pralit Patel and Page Kyle

Date committed: 4/24/2023

IR document number: PNNL-34313

Related sector: Agriculture, Land, and Economy

Type of development: Data, Code, and Queries

Purpose: This Core Model Proposal revisits the processing of GCAM agriculture and land use (AgLU) data in gcamdata to (1) develop methods to automate AgLU data (mainly FAOSTAT) updates to facilitate future base year updates or other data adjustments, (2) compile the new FAO FBS (food balance sheet) and SUA (supply-utilization accounting) data to provide traceable flows from land-based primary production to end uses (food & non-food), (3) develop methods of generating "primary equivalent" to bridge the gap between primary agricultural supply and final consumption, (4) restructure and simplify gcamdata to incorporate the new data, and (5) update GCAM AgLU data and assumptions, where needed.

Contents

1. Introduction 3
1.1. Problems and motivations 3
1.2. Objectives 4
2. Description of changes 5
2.1. FAO data 5
Fig. 2.1 Snapshot of gcamdata *aglu/FAO* folder before the updates (the current *Master* branch) 5
Fig. 2.2 Snapshot of gcamdata *aglu/FAO* folder after the updates (the *AgLU-Update* branch) 6
Fig. 2.3 Mapping between FAO and GCAM food commodities (A) and count of FAO commodities per GCAM commodity (B) 6
Fig. 2.4 Mapping between FAO and GCAM primary (land-based) commodities (A) and count of FAO commodities per GCAM commodity (B) 7
2.2. IMAGE livestock data 7
2.3. USDA data 7
2.4. Minimum rental profit 7
2.5. Unmanaged land value 8
Fig. 2.5 Comparison of base year unmanaged land rental price for impacted basin 9
2.6. Managed pasture yield 9
2.7. Simplified data declaring structure in gcamdata 10
2.8. Module restructure and adding outline and assertion 10
2.9. Query updates 10
2.10. Overview of key changes in gcamdata 10
Table 1 Additional data and code changes made in gcamdata 11
Fig. 2.6 Flow chart for the primary equivalent processing in the *zchunk_LA100.FAO_SUA_PrimaryEquivalent module* 13
3. Impacts on the base data 14
Fig. 3.0 Connecting land hectares to food calories. 14
Fig. 3.1 Comparison of base year (2015) land cover allocation 15
Fig. 3.2 Comparison of land allocation by crop 16
Fig. 3.3 Comparing base year yield distribution over GCAM regions by commodity (A) and world mean yield (B) 17
Fig. 3.4 Comparing base year feed conversion rate distribution over GCAM regions by commodity (A) and world mean value (B) between AgLU-Update and Master. 18
Fig. 3.5 Supply-utilization-accounting (SUA) balance comparison 18
Fig. 3.6 Impact of the data update on the base year agricultural SUA balance. 20
Fig. 3.7 Comparison of base year world total Calorie consumption by commodity (A) and Calorie conversion rate distribution by commodity (B) between AgLU-Update and Master. 21
Fig. 3.8 Comparison of the base year regional caloric consumption per capita per day 22

Fig. 3.9 Comparison of regional fertilizer demand (by crop) and trade 23
Fig. 3.10 Comparison of global water withdrawal 24
Fig. 3.11 Comparison of base year producer price distribution by sector (A) and world mean prices (B) between AgLU-Update and Master. 25
Fig. 3.12 Comparison of base year NonCO2 GHG emissions by source between AgLU-Update and Master. 25
4. Shared policy assumption (SPA) GCAM validation runs 26
Fig. 4.1 (extension of Fig. 3.1B) Land cover difference 26
Fig. 4.2 (extension of Fig. 3.2B) Cropland area (by crop) difference 27
Fig. 4.3 Cropland area (by crop) change relative to the base year 28
Fig. 4.4 (extension of Fig. 3.6A) agricultural SUA balance difference 29
Fig. 4.5 (extension of Fig. 3.5B-Production) Comparison of Ag production projections 30
Fig. 4.6 (extension of Fig. 3.5B-Export / Import) Comparison of Ag trade projections 31
Fig. 4.8 (extension of Fig. 3.5B-Food) Comparison of Ag food demand projections 32
Fig. 4.9 (extension of Fig. 3.7A) Comparison of Ag food calorie projections 33
Fig. 4.10 (extension of Fig. 3.9 at the world level) Comparison of total fertilizer demand 33
Fig. 4.11 (extension of Fig. 3.10 at the world level) Comparison of total water withdrawal 34
Fig. 4.12 (extension of Fig. 3.11B) Comparison of world Ag price 34
Fig. 4.13 Comparison of land use change (LUC) carbon emission projections 35
Fig. 4.14 (extension of Fig. 3.12) Comparison of key emission projections 36
Fig. 4.15 Comparison of key climate variables and carbon price projections 37
5. Future work 38
6. Supplementary information (gcamdata-faostat) 39
6.1. Maintain agricultural supply-utilization balance in data and modeling 39
6.2. Sectoral aggregation along the processing chain using primary equivalent 40
Fig. 6.1 Illustration of the primary equivalent approach 40
6.3. Can gcamdata-faostat be included in gcamdata? 41
Acknowledgments 42
References 43

1. Introduction

Robust economic modeling relies on high-quality historical data and validated or tested behavior parameters. Historical data, especially data in the model base year, is extremely important for global economic equilibrium modeling because

(1) base data is a snapshot of the economy in the base year that represents the initial equilibrium of all modeled markets, which is used as a reference economy for future periods relative to which future socioeconomic changes, technological progress, policies, and other shocks are implemented,
(2) the model is calibrated to the base data using calibration parameters (e.g., share-weights in logit, multipliers in exponential demand functions, slopes in piece-wise supply functions, etc.), and
(3) the level of detail in base data also determined the modeling structure, e.g., spatial & sectoral resolution and interregional & cross-sectoral linkages, and future projections (output database in future periods) would have the same level of detail with the base data. That is, future projections and counterfactual experiments could be very sensitive to measurement errors and the representation of the base economy.

The gcamdata (an open-source R package) (Bond-Lamberty et al., 2019) has been used for processing raw data inputs into structured & balanced base data (in XMLs) for GCAM modeling in a transparent, reproducible, and flexible manner. The package provides an elegant and generalized structure for model and data development. However, data development is not a once and for all task, and continued efforts are needed to maintain and improve the processing since

(1) the raw data need to be maintained and updated regularly, e.g., for base year update (BYU),
(2) strong assumptions might have been used in aggregation, disaggregation, mapping, and averaging the data, and
(3) assumptions were made in places with no or low-quality data (e.g., interpolation or extrapolation to fill in missing values or uniform global price data) while better data becomes available.

1.1. Problems and motivations

This Core Model Proposal (CMP) attempts to examine and improve the processing of Agricultural and Land Use (AgLU) related data in GCAM/gcamdata. Here are some problems with the existing AgLU data and methods that motivated this CMP.

(1) The AgLU-related updates were not a focus in the last BYU. E.g., many FAO agro-food datasets were only updated to 2013 (FAO changed approaches thereafter) and extrapolated to 2017, and there was a bug that only a 3-year (2013 – 2015) average was used as opposed to a 5-year (2013 – 2017) average.
(2) GCAM agricultural commodity mappings are inconsistent. E.g., different mapping (FAO to GCAM) was used for agricultural production, food consumption, food calorie conversion coefficients, trade, and prices.
(3) Some of the existing data processing modules were one-direction, leading to unused or redundant processing and dependencies (e.g., net trade processing is no longer needed).

(4) Supply-utilization-accounting (SUA) data ignored the processing uses and used other uses (nonfood use in GCAM) to maintain SUA balance (i.e., market clearing conditions). The adjustments could likely affect future food & feed projections as processed food & feed were not traced. Note that the original AgLU data and method were documented in detail in Kyle et al. (2011). New data and methods have become available since then, along with some important recent modeling developments (e.g., trade, production technologies, parameter updates, and new primary equivalent method in this CMP), so there could be a need to update the GCAM AgLU document.
(5) Agricultural storage data, including opening and ending stocks (& interannual stock variation), was not traceable in the existing SUA balance. However, there is an ongoing task to incorporate agricultural storage responses in GCAM, which heavily rely on data representation.
(6) In addition, there were several other areas where the data assumptions should now be updated and improved, e.g.,
    a. Water content in livestock feed and meat was mostly ignored (leading to a misrepresentation of the feed crop input use).
    b. Unmanaged land value (from GTAP data in 2000) and livestock-related technology and feed conversion efficiency data (from IMAGE in 2005) are dated, so they could have discrepancies in extrapolation, mapping, and aggregation.
    c. Pasture yield assumptions ignore different land productivity & grazing intensity, leading to underestimated managed pasture land use.
    d. Fertilizer prices were inconsistent between gcamdata and GCAM, possibly leading to overestimated nonland costs and negative rental profits.

1.2. Objectives

This CMP aims to improve the GCAM AgLU data processing method to resolve the problems discussed above and to improve the traceability, consistency, and robustness of the AgLU data. More specifically, the objectives include:

(1) Develop a package (*gcamdata-faostat*) to download and process raw data directly from FAOSTAT for gcamdata and GCAM uses.
    a. The package generates all FAOSTAT-based raw data needed in gcamdata (i.e., aglu/FAO).
    b. The package compiles the SUA data to provide traceable flows from land-based primary production to trade & end uses (e.g., food, feed, loss, processed, storage variation, etc.) for 500+ agricultural commodities. The SUA data is connected to the Food Balance Sheet (FBS).
    c. This CMP, with a more consistent & traceable framework, will also simplify and expedite the ongoing and future AgLU BYU efforts.
(2) Develop methods of generating “primary equivalent” to bridge the gap between agricultural supply and final consumption and disaggregate food storage and loss data in GCAM. The approach permits tracing and aggregating physical flows along the vertical supply chain. The new data preserve balance across supply-utilization, space, time, and vertical processing sectors (primary equivalent).
(3) Update AgLU data raw input data to the latest available data. Restructure and simplify AgLU processing in gcamdata.

(4) Improve the robustness and documentation of the AgLU data processing & balancing.

2. Description of changes

2.1. FAO data

We develop a separate processing package, *gcamdata-faostat*, to generate and update all FAOSTAT csv files in the AgLU/FAO folder (see **Fig. 2.1** vs. **Fig. 2.2** for the changes). With the update, the FAO folder will only have 17 csv files (decreased from 35), including 8 supply-utilization accounting & prices files, 4 mapping files, and 4 files related to land cover, fertilizer, and forestry, and one unchanged/irrelevant file (FAO_ag_CROSIT.csv).

Fig. 2.1 Snapshot of gcamdata *aglu/FAO* folder before the updates (the current *Master* branch)

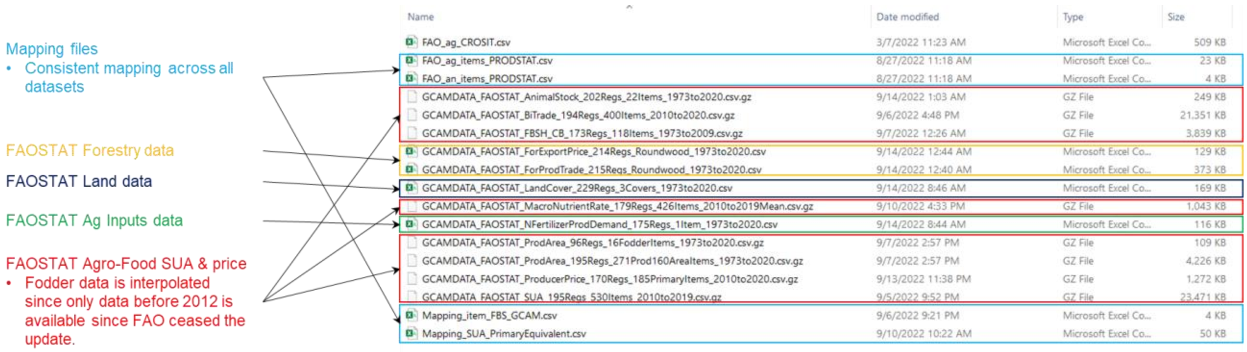


Fig. 2.2 Snapshot of gcamdata *aglu/FAO* folder after the updates (the *AgLU-Update* branch)

The main focus of the package is to provide data that connect the agricultural supply to utilization in a consistent and traceable manner. Major developments were made to generate the SUA files and associated mappings. More detail about *gcamdata-faostat* is provided in Supplementary Information (SI) in Section 6. With the new method, there will be one single mapping for all elements between FAO items (~530 items) and GCAM agricultural commodities (21 items). **Figs. 2.3 – 2.4** present the mappings between FAO and GCAM for food commodities and primary (land-based) commodities, respectively.

Note that the primary equivalent aggregation is included in gcamdata. It is generalized using a recursive function, and the function is controlled by a mapping file (*input/AgLU_FAO/Mapping_SUA_PrimaryEquivalent.csv*). It is able to disaggregate agricultural products into 71 primary commodities, including potatoes, tomatoes, cassava, etc. Note that intraregional trade is removed for commodities with available bilateral trade data at SUA level.

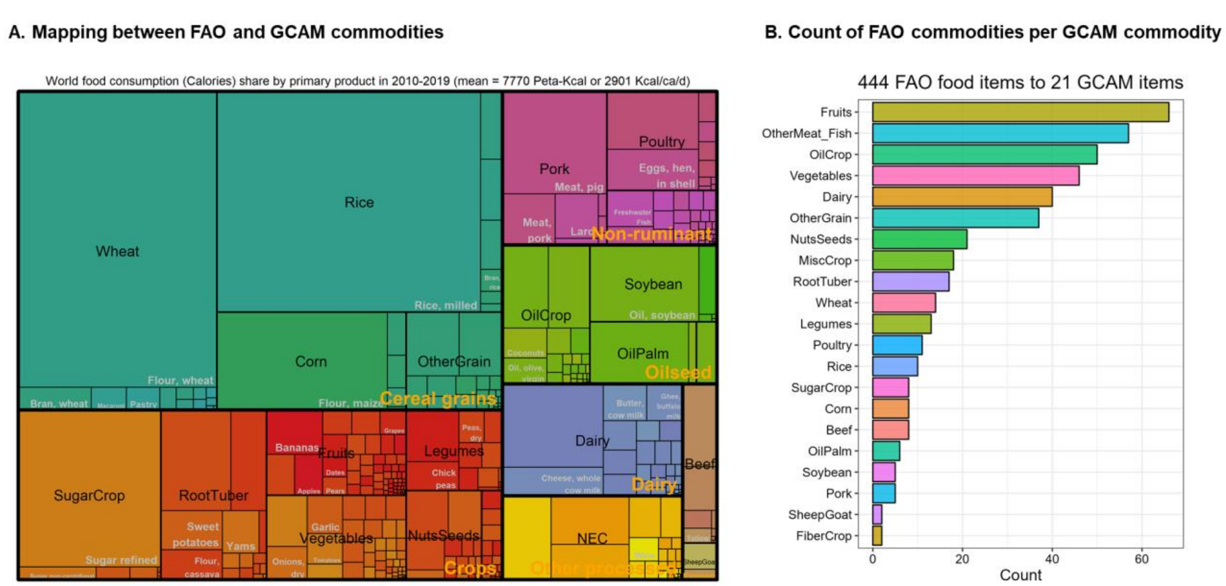


Fig. 2.3 Mapping between FAO and GCAM food commodities (A) and count of FAO commodities per GCAM commodity (B)

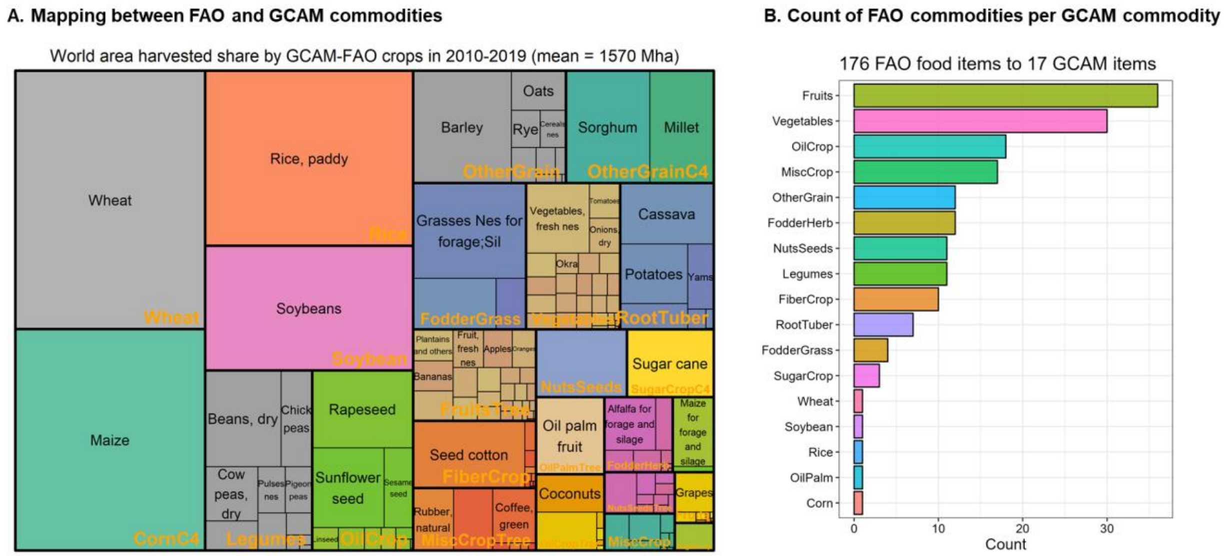


Fig. 2.4 Mapping between FAO and GCAM primary (land-based) commodities (A) and count of FAO commodities per GCAM commodity (B)

2.2. IMAGE livestock data

All data in *aglu/IMAGE* folder are updated to IMAGE v3.2 data (shared by Jonathan Doelman in Oct 2022). The source data were used directly, so the corresponding processing is included/updated in *module_aglu_LA100.IMAGE_downscale_ctry_y*. The IMAGE feed conversion rates data are on a dry matter basis, while the FAO feed crops are in wet tonnes. We modify the processing to prioritize the use of FAO data for feed crops (see *module_aglu_LA107.an_IMAGE_R_C_Sys_Fd_Y*). Similarly, IMAGE dry matter livestock product output was converted to wet tonnes for consistency (see *module_aglu_LA100.IMAGE_downscale_ctry_y*). These updates in data and assumption significantly improve the balance and traceability of the SUA in GCAM. Note that fodder crops data had relatively lower quality, so they should be improved (including water tracing) when better data is available.

2.3. USDA data

A folder is created in *aglu/USDA* to include all AgLU data from USDA. The alfalfa prices (*aglu/USDA/USDA_Alfalfa_prices_USDt.csv*) are updated.

2.4. Minimum rental profit

Negative rental profit is one of the most common model errors in GCAM. It happens when agricultural input cost (including water, fertilizer, and others) is higher than producer price. Note that the fertilizer price processing in gcamdata was simplified, so regional fertilizer prices are calibrated in GCAM. As a result, the other cost (Nonland cost) could be overestimated in gcamdata for some regions, leading to relatively lower rental profit. There could be other reasons leading to negative profits, e.g., measurement errors in costs and the neglect of government payments to farmers.

Previously, a water subsidy was computed in gcamdata and added to rental profit. However, it was not as effective as intended since rental profits are calculated based on calibrated prices and costs in GCAM. In this CMP, we incorporate a minimum rental profit (minProfitMargin) in GCAM C++ code to ensure positive rental profits. In particular, we still use minProfitMargin generated from gcamdata to compute an implied subsidy (see *mImpliedSubsidy in ag_production_technology.cpp*). Note that the minProfitMargin is calculated in *module_aglu_L2052.ag_prodchange_cost_irr_mgmt* as a minimum value across all regions & crops, with the flexibility to differentiate further in future studies, e.g., decoupled payments to farmers (Chambers and Voica, 2017). And *L2052.AgCalMinProfitRate* is exported in *ag_cost_IRR_MGMT.xml*. Note that the old approach using water subsidy in gcamdata is also removed (from *zchunk_L2072.ag_water_irr_mgmt*).

There are 162 combinations of region x crop x technology being affected by the new minProfitMargin, mainly including SugarCrop in Australia_NZ, India, Brazil, and South Africa and MiscCrop in India and the Middle East. Note that a quick bugfix on revising the India Nutmeg fertilizer application rate was also included in this CMP to improve the fertilizer input-out coefficient, which also helped the negative profit issue.

2.5. Unmanaged land value

About 5 basins (2 in SEA) had missing values for unmanaged land rental profit, and a few others had very poor/low values (e.g., Hong basin in SEA). gcamdata filled in missing with a global minimum (close to Hong's value). In this CMP, we made changes to use the median to fill in missing and added a higher minimum threshold (see *module_aglu_L221.land_input_1*).

The land use change results in the "outlier" regions would be significantly improved with the update (**Fig. 2.5**). For example, for Hong basin, the unmanaged land price is 200+ times higher than the value used before. With the new value, it is about 350 2015$ per ha, representing a moderate/reasonable fertile land rental price in the US. Previously, the value was too low, and any land carbon policy would have a significantly large impact.

Note that the source of the data for unmanaged land value is 2000 GTAP data, which is dated. Future work is needed to examine these values more carefully and test more values that are more consistent with GCAM data (e.g., using values calculated based on GCAM rental profit of managed land).

Region	Basin	New 1975$ per bm2	Old 1975$ per bm2	Ratio (New / Old)
Africa_Southern	NileR	707824	41079	17
Africa_Southern	RiftValley	707824	41079	17
Southeast Asia	XunJiang	10454986	41079	255
Southeast Asia	SChinaSea	10454986	41079	255
Taiwan	ChinaCst	6679889	41079	163
Taiwan	Taiwan	6679889	41079	163
Brazil	SAmerCstNE	1362502	40957	33
Southeast Asia	Hong	10454986	44104	237
South Asia	HamuMashR	2610948	74004	35
Argentina	LaPuna	4560933	172194	26
Africa_Eastern	CongoR	707824	673455	1.1

Fig. 2.5 Comparison of base year unmanaged land rental price for impacted basin between *AgLU-Update* and *Master* branches.

2.6. Managed pasture yield

The productivity of pasture was determined in module_aglu_LB121.Carbon_LT by the carbon yield. For some reason, a uniform carbon yield is used for pasture, i.e., 0.6 kg C per m2. However, pasture yield could be very different across regions given different land productivity and grazing intensity (not all pasture is grazed). In this CMP, we update the module to differentiate pasture yield using vegetation carbon and consider a global grazing intensity value. As a result, the global pasture yield decreased significantly from 13.5 to 3.2 dry matter tonnes per ha, and more managed pasture area was used in the base year. The assumptions here should be revisited and studied carefully later.

Note that the gross pasture C yield per year is calculated as vegetation C density / mature age. However, it could underestimate the annual pasture yield since grazing is likely in high yield age. So we use a 40 percentile value across basins (~1.4t/ha) as the minimum gross C pasture annual yield. Then the grazed pasture crop yield is calculated as gross pasture C yield X grazing intensity / Cellulosic C content (0.45). Global grazing intensity is tuned to have a global average pasture dry matter yield of ~3 tonnes per ha. More details are provided in the module_aglu_LB121.Carbon_LT of gcamdata. Note that there are recent literature estimates of global grazing intensity, e.g., Wolf et al. (2021). As a result, the global pasture yield decreased significantly from 13.5 to 3.2 dry matter tonnes per ha, and more managed pasture area was used in the base year. The data requires more examination together with the vegetation carbon data. And the assumptions here should be revisited and studied carefully later.

2.7. Simplified data declaring structure in gcamdata

MODULE_INPUTS and MODULE_OUTPUTS are defined at the very beginning of a chunk as characters. They will be called when needed.

a) The following code is used to load data

```
lapply(MODULE_INPUTS, function(d){
   nm <- tail(strsplit(d, "/")[[1]], n = 1)
   assign(nm, get_data(all_data, d, strip_attributes = T), envir = parent.env(environment()))  })
```

Note that the about code is improved and formalized in a function:
get_data_list(all_data, MODULE_INPUTS, strip_attributes = TRUE)

b) Return data is updated to *return_data(MODULE_OUTPUTS)*

c) *return_data()* in Utils-data.R is modified to make this possible.

This change is included in most AgLU-related R chunks that are relevant to this CMP. One advantage of *get_data_list()* is we do not need to duplicate dataset name uses in both *Module_Inputs* and *get_data*. When we add new data or update the data dataset name in *Module_Inputs*, we won't need to change *get_data_list()*. For the same reason, *Module_Outputs* is used in *return_data()* now. With this change, the code is shortened, and the data update will be less error-prone.

2.8. Module restructure and adding outline and assertion

Many changes were made to simplify the processing of AgLU data in gcamdata. Where applicable, outlines and more detailed comments are added in the key data processing chunks (e.g., see *module_aglu_LA100.FAO_SUA_PrimaryEquivalent* and *module_aglu_LA100.FAO_SUA_connection*. In addition, where applicable, assertions are added to ensure data balance, e.g., *see module_aglu_LA101.ag_FAO_R_C_Y* and *module_aglu_LB109.ag_an_ALL_R_C_Y*.

2.9. Query updates

The following queries are updated or added in the default model interface queries:

- "ag tech implied subsidy" is added under crop production.
- "ag trade" related queries are added under AgLU.
- "residue biomass production" is updated.
- "demand balances by crop commodity" is fixed.
- "gas prices by sector" is fixed.

2.10. Overview of key changes in gcamdata

Other key data and code changes made in gcamdata are summarized **in Table 1**. Note that the data files shown in **Figs. 2.2** (data processed by *gcamdata-faostat*) are not included in **Table 1** as the file name is self-explanatory. Files removed from **Fig. 2.1** are also not included in **Table 1**. For most of the R script updates, the outline and detailed descriptions were added if missing. The original plan for the *gcamdata-faostat* was more ambitious, e.g., providing an input-output view in both quantity and value forms. However, agricultural cost data were not updated in this CMP, and there are data limitations in the feed inputs and land rental profits calculation in gcamdata (as some prices are calibrated in GCAM). After all, the changes in the base data because of the new data and method are already not small. **Section 3** discusses these changes in base data, and **Section 4** discusses the impacts on the validation runs.

Note that *zchunk_LA100.FAO_SUA_PrimaryEquivalent.R* includes the new primary equivalent aggregation (along the processing supply chain) method. The functions are written in a flexible way so that more agricultural commodities and regions can be disaggregated consistently. **Fig. 2.6** shows a flow chart for the primary equivalent processing.

Table 1 Additional data and code changes made in gcamdata

Data file or R chunk	Changes made
aglu/FAO/FAO_ag_items_PRODSTAT.csv	Refined primary crop production mapping between FAO and GCAM, including a price_item column as indicators.
aglu/FAO/FAO_an_items_PRODSTAT.csv	Refined primary livestock sector production mapping between FAO and GCAM, including a price_item column as indicators.
aglu/FAO/Mapping_item_FBS_GCAM.csv	Mapping file for connecting FAO FBS and SUA dataset before 2010.
aglu/FAO/Mapping_SUA_PrimaryEquivalent.csv	Mapping file used for primary equivalent aggregation.
aglu/A_recent_feed_modifications	Not needed anymore.
aglu/USDA	Add a USDA folder to include USDA csv data.
aglu/USDA/USDA_Alfalfa_prices_USDt.csv	Update alfalfa prices.
aglu/Mekonnen_Hoekstra_Rep47_A2.csv	Update to the new FAO mapping.
aglu/Various_ag_resbio_data.csv	Update item and include item_code for FAO crops.
aglu/AGLU_ctry.csv	Update FAO country names
common/FAO_GDP_Deflators.csv	data update
water/FAO_an_items_Stocks.csv	Update item and include item_code for FAO items.
constants.R	aglu.FALLOW_YEARS and aglu.MODEL_PRICE_YEARS are now updated to 2013:2017 (from 2008:2012 and 2008:2016, respectively) Add gcam.REAL_PRICE_BASE_YEAR <- 1975 : This is only used for AgLU prices now. Add aglu.MODEL_MEAN_PERIOD_LENGTH <- 5 : used for averaging AgLU data. This can be changed to 1 year to imply no averaged data used.

module-helpers.R	Add 3 functions: FAO_AREA_DISAGGREGATE_HIST_DISSOLUTION: Disaggregated data for the historical period for of a dissolved region. FAO_AREA_DISAGGREGATE_HIST_DISSOLUTION_ALL: FAO_AREA_DISAGGREGATE_HIST_DISSOLUTION for all regions. Moving_average: calculate the moving average of any year (e.g., AgLU is five year)
pipeline-helpers.R	extend gdp_deflator to include 2020 and 2021.
zchunk_batch_ag_cost_IRR_MGMT_xml.R zchunk_batch_ag_Fert_IRR_MGMT_xml.R	Use L2062 cost directly in zchunk_batch_ag_cost_IRR_MGMT_xml, moved from zchunk_batch_ag_Fert_IRR_MGMT_xml. The L2062 cost data was adjusted by the fertilizer costs. The original L2052 costs were overwritten in configuration anyways so they are removed to save space.
zchunk_batch_ag_For_Past_bio_base_IRR_MGMT_xml.R	Remove L2012.AgHAtoCL_irr_mgmt from exporting as it is not used. Harvest frequency should be revisited.
zchunk_L133.water_demand_livestock.R	Use FAO item code in the updated animal stock data.
zchunk_LA100.regional_ag_an_for_prices.R	Process price data from source for consistency, use consistent mappings and reduce hard-coded assumptions and extrapolations.
zchunk_LB132.ag_an_For_Prices_USA_C_2005.R	Merged into zchunk_LA100.regional_ag_an_for_prices.R
zchunk_LB1321.regional_ag_prices.R	Merged into zchunk_LA100.regional_ag_an_for_prices.R
zchunk_LA100.0_LDS_preprocessing.R	Adjustment for small yield crops in small region: GTAP_crop == "FrgProdNES", GLU %in% c("GLU049", "GLU021"), iso == "pol"
zchunk_LA105.an_FAO_R_C_Y.R	Chunk removed. The processing was merged into zchunk_LA100.FAO_downscale_ctry.R if needed.
zchunk_LA106.ag_an_NetExp_FAO_R_C_Y.R	Chunk removed. The processing was merged into zchunk_LA100.FAO_downscale_ctry.R if needed.
zchunk_LB1091.ag_GrossTrade.R	Chunk removed. The processing was merged into zchunk_LA100.FAO_downscale_ctry.R if needed.
zchunk_LA101.ag_FAO_R_C_Y.R	Removed the food processing part and simplified the production and area downscaling part.
zchunk_LA108.ag_Feed_R_C_Y.R	Pasture feed consumption would be set to zero if FodderGrass were too large. This was adjusted by adding a minimum pasture share (10%) over Pasture_FodderGrass to avoid zero pasture adjustments. This fixes the EU solution/calibration issues.
zchunk_LB109.ag_an_ALL_R_C_Y.R	Rebalance supply-utilization after feed adjustments in gcamdata based on IMAGE IO coefficients.
zchunk_LB110.For_FAO_R_Y.R	Include gross trade in forest balance here to reduce processing dependency.
zchunk_L240.ag_trade.R	remove L1091.GrossTrade_Mt_R_C_Y
zchunk_L202.an_input.R	remove L1091.GrossTrade_Mt_R_C_Y and L132.ag_an_For_Prices
zchunk_L2052.ag_prodchange_cost_irr_mgmt.R	Replace 132 prices with 1321 prices
zchunk_LA100.FAO_SUA_PrimaryEquivalent.R	This chunk compiles balanced supply utilization data in primary equivalent in GCAM region and commodities. A method to generate a primary equivalent is created for the new FAOSTAT supply utilization data (2010 to 2019). The new SUA balance is connected to the old one (before 2010). Production and harvested area data with FAO region and item for primary production are provided. For FAO food items, macronutrient values are calculated at SUA item level. Data processing was consistent across scales. Note that GCAM regions and commodities in aggregation mapping can be changed in corresponding mappings. Note that intraregional trade is removed in this chunk.

zchunk_aglu_LA100.FAO_SUA_connection.R	Further process and aggregate SUA data for GCAM use. Calculate the 5-year average.
zchunk_aglu_LA100.FAO_preprocessing_OtherData.R	Get FAO data ready for forestry, fertilizer, animal stock, and land cover. Calculate the 5-year average.

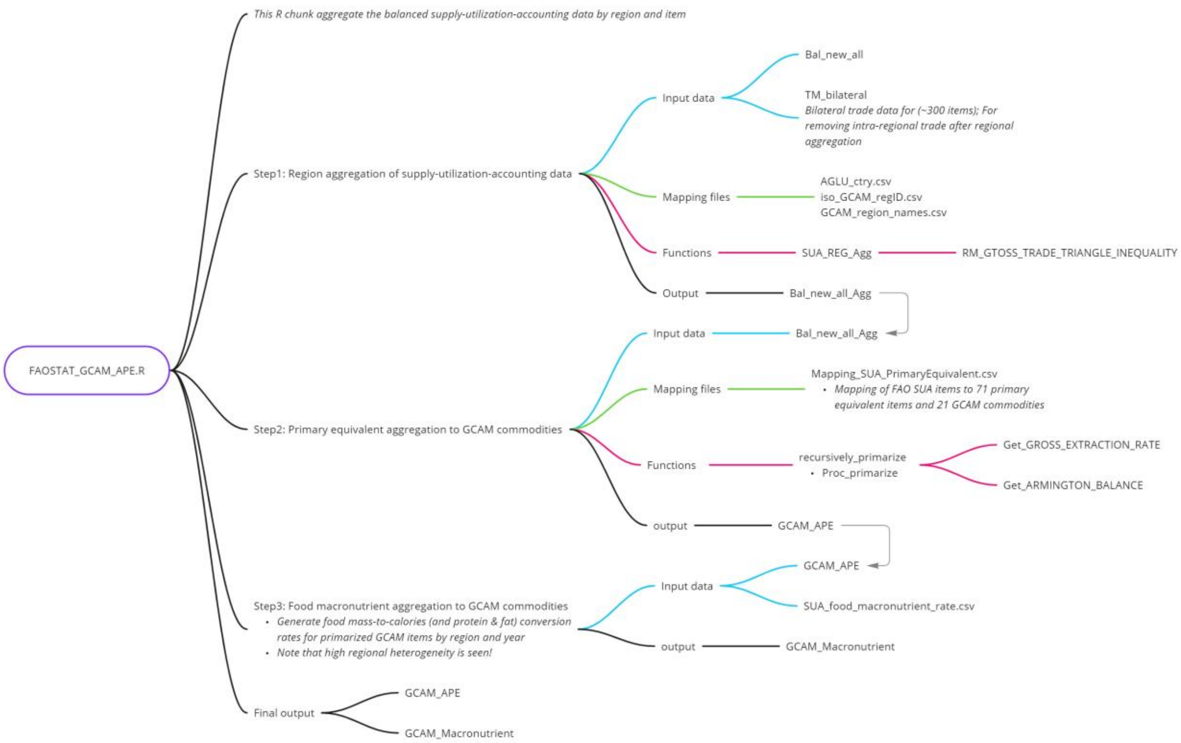


Fig. 2.6 Flow chart for the primary equivalent processing in the *zchunk_LA100.FAO_SUA_PrimaryEquivalent module*

3. Impacts on the base data

Fig. 3.0 shows the connection between land hectares to food calories. GCAM and its base data represent the land balance. Cropland (harvested areas), managed pasture & forest, and other inputs (e.g., fertilizer and water) are used for agricultural production. Note that GCAM does not explicitly model harvested area currently. Instead, the corresponding cropland cover is used, so the total crop area equals to total cropland cover (not including other arable land). In other words, crop harvest frequency is assumed to be fixed at the initial value. This can be improved in future studies given the importance of cropping intensity responses (Zhu et al., 2022). Furthermore, with the new primary equivalent approach, processed food and feed are represented in their primary equivalent. The supply-utilization balance is improved. Feed conversion rates and calorie conversion rates are computed based on data in the aggregation procedure. Thus, the food calories are consistent with the widely used FAOSTAT data. Agricultural storage data are currently aggregated into other use, but can be easily disaggregated in future work. In addition to the physical volume flows, there could be value (prices) and emission flows from land to food and other end uses.

This section compares the AgLU base year data between the new branch (*AgLU-Update*) and the GCAM master branch. The corresponding result figures for the key areas are referenced in **Fig. 3.0**. The figures are mostly self-explanatory, but explanations are also added in the captions.

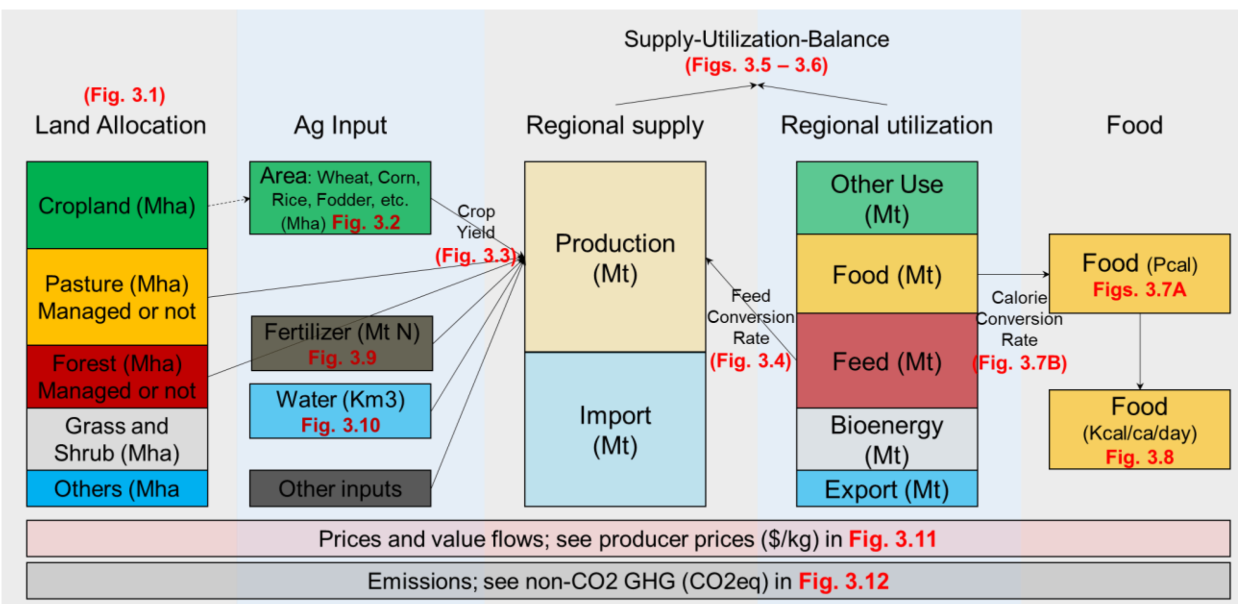


Fig. 3.0 Connecting land hectares to food calories.

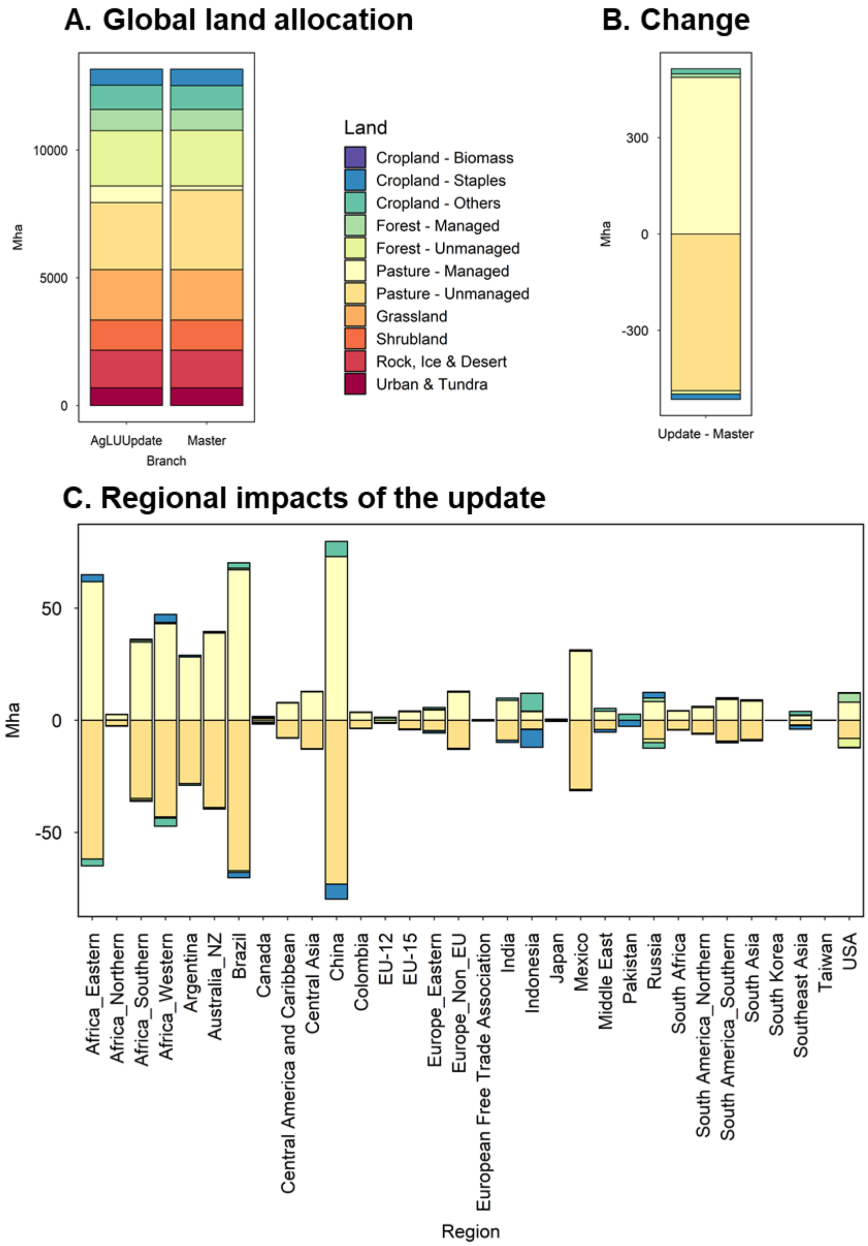


Fig. 3.1 Comparison of base year (2015) land cover allocation between AgLU-Update and Master branches. With the data updates, the main impact is a higher share of managed pasture in the total pasture area. The world managed pasture increased by ~500 million hectares (Mha), from 160 to 650 Mha, almost entirely from unmanaged pasture. The change was mainly driven by lower pasture yield, higher livestock product output in the base year, and livestock feed conversion rates related updates. Note that the total pasture area (both managed and unmanaged) is ~3275 Mha. So the managed pasture share increased from ~5% to ~20% with the updates. Given the importance of the managed pasture area (i.e., directly linked to livestock production), the assumptions on pasture yield and grazing intensity should be further examined. In addition, there are also relatively small changes in cropland areas (see **Fig. 3.2** for details).

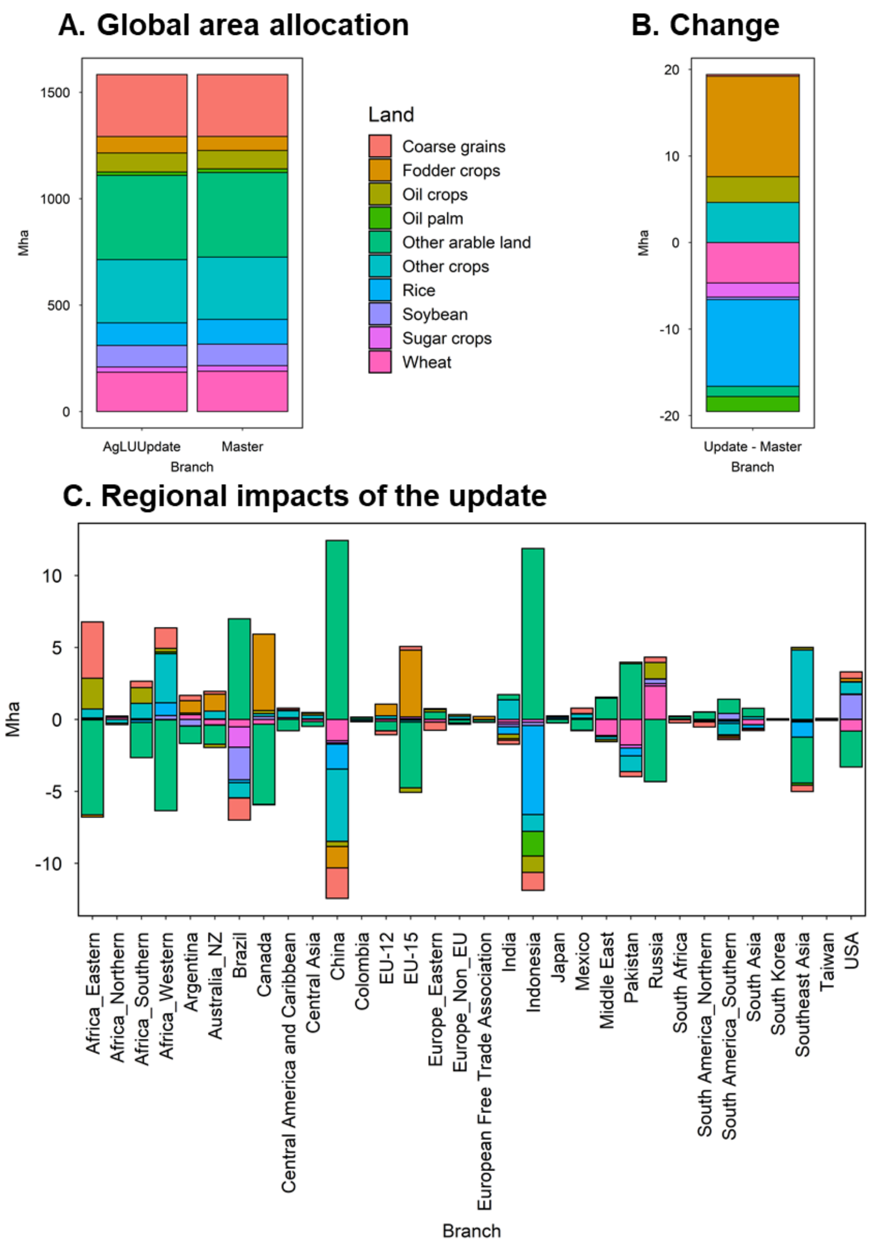
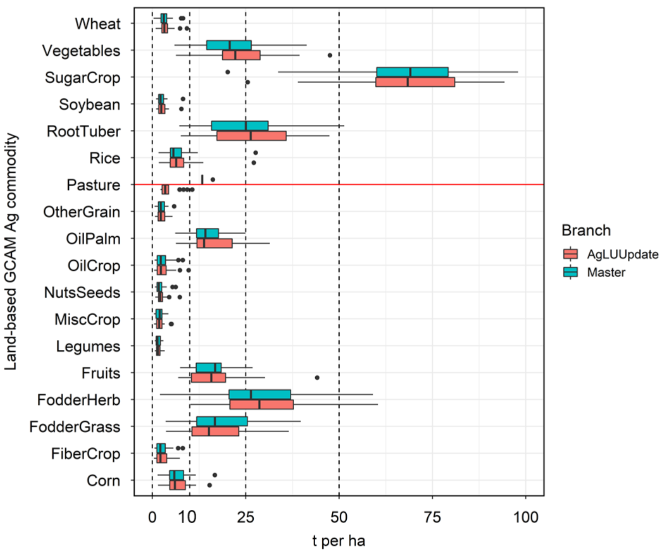


Fig. 3.2 Comparison of land allocation by crop between AgLU-Update and Master branches. Other arable land had relatively larger changes in the new FAO data. Note that FAO only provides harvested area for crops while other arable land is considered an unused cropland. Globally, other arable land is about 400 Mha (with very small world total change due to the update). Since harvested area for crop is mapped/scaled to land cover for the crop, changes in other arable land can also affect the mapping/scaling process at the water basin level. With the update, there would be larger regional other arable land available mainly in China, India and Australia_NZ and smaller other arable land area in African regions, EU-15, Russia, etc.

A. Regional yield distribution by crop

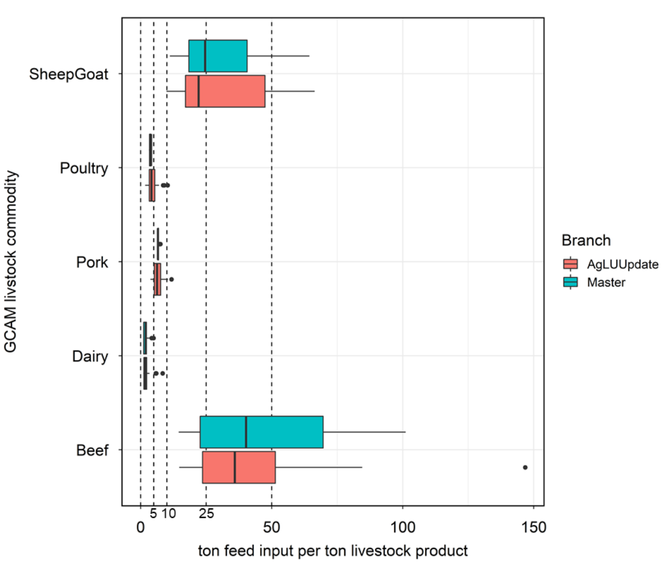


B. World mean crop yield comparison

Sector	AgLUUpdate (t/ha)	Master (t/ha)	Change
Pasture	3.2	13.5	-76%
FodderGrass	17.5	22.7	-23%
Fruits	15.6	16.3	-4%
FodderHerb	29.7	30.6	-3%
OtherGrain	2.2	2.3	-2%
NutsSeeds	1.9	2.0	-1%
OilCrop	2.4	2.4	-1%
RootTuber	17.7	17.2	3%
Legumes	1.2	1.1	3%
Wheat	4.0	3.9	4%
FiberCrop	2.8	2.7	4%
Corn	7.0	6.7	5%
Soybean	3.2	3.0	7%
SugarCrop	87.2	81.3	7%
Rice	6.9	6.4	8%
Vegetables	28.4	24.1	17%
MiscCrop	1.5	1.3	18%
OilPalm	21.6	16.0	35%

Fig. 3.3 Comparing base year yield distribution over GCAM regions by commodity (A) and world mean yield (B) between AgLU-Update and Master. Pasture yield has the largest change after using regional vegetation carbon yield information and considering grazing intensity (a world average of 13.5 dry matter ton/ha used previously was likely overestimated as the yield was larger than most feed crops). FodderGrass yield also decreased (-23%) due to updated data and improved extrapolation assumptions. More importantly, the yield of Vegetables, MiscCrop, and OilPalm increased (by 17 – 35%), reflecting data and mapping improvements. Crop yields are generally higher for food crops with the updates.

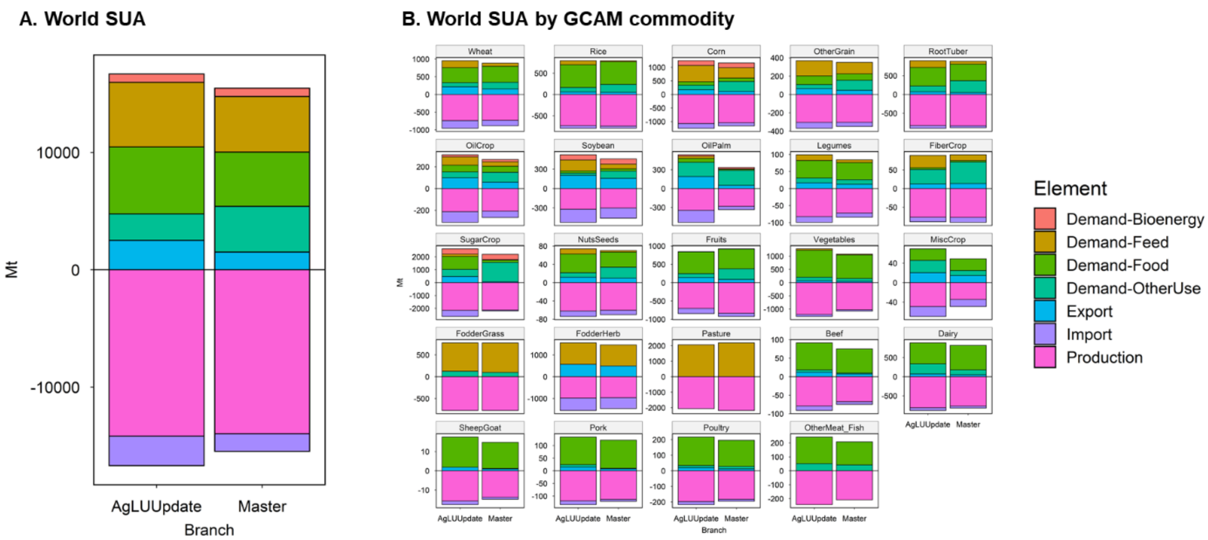
A. Feed conversion efficiency distribution



B. World feed conversion efficiency comparison

Sector	AgLUUpdate (t/t)	Master (t/t)	Change
Beef	39.8	43.4	-8%
Dairy	1.6	1.7	-2%
Pork	6.0	6.6	-9%
Poultry	4.5	3.5	25%
SheepGoat	20.2	22.7	-11%

Fig. 3.4 Comparing base year feed conversion rate distribution over GCAM regions by commodity (A) and world mean value (B) between AgLU-Update and Master. The feed conversion rate in Master had a unit of dry matter tonnes per dry matter tonne. This is changed in the updated branch to be more consistent with FAO data (mostly in wet tonnes). The global average feed conversion rates did not change significantly with the update, though all livestock sectors except Poultry became more productive with the updates.



C. Changes in other use (NonFoodDemand) at the world level

element	sector	Master (Mt)	AgLUUpdate (Mt)	Diff (Mt)	Diff (%)	unit
Demand-OtherUse	Dairy	123	260	137	111%	Mt
Demand-OtherUse	FodderGrass	99	126	28	28%	Mt
Demand-OtherUse	Vegetables	106	132	27	25%	Mt
Demand-OtherUse	MiscCrop	11	25	15	141%	Mt
Demand-OtherUse	OtherMeat_Fish	42	51	9	22%	Mt
Demand-OtherUse	Pork	2	10	8	513%	Mt
Demand-OtherUse	Beef	3	6	3	133%	Mt
Demand-OtherUse	Legumes	14	14	0	2%	Mt
Demand-OtherUse	SheepGoat	0	0	0	-100%	Mt
Demand-OtherUse	Poultry	16	15	-1	-5%	Mt
Demand-OtherUse	OilPalm	239	224	-14	-6%	Mt
Demand-OtherUse	NutsSeeds	24	10	-14	-60%	Mt
Demand-OtherUse	FiberCrop	58	39	-18	-32%	Mt
Demand-OtherUse	OilCrop	90	55	-35	-39%	Mt
Demand-OtherUse	OtherGrain	108	44	-64	-59%	Mt
Demand-OtherUse	Wheat	191	117	-74	-39%	Mt
Demand-OtherUse	Soybean	112	36	-76	-68%	Mt
Demand-OtherUse	Rice	188	105	-83	-44%	Mt
Demand-OtherUse	RootTuber	315	152	-162	-52%	Mt
Demand-OtherUse	Fruits	287	106	-182	-63%	Mt
Demand-OtherUse	Corn	362	172	-190	-53%	Mt
Demand-OtherUse	SugarCrop	1508	543	-965	-64%	Mt
	Total	3896	2242	-1654	-42%	

Fig. 3.5 Supply-utilization-accounting (SUA) balance comparison (*AgLU-Update* vs. *Master*) for an aggregated agricultural commodity (A) and GCAM commodities (B) at the world level in the base year (2015). Changes in other use (NonFoodDemand) is shown in (C). The new SUA balance, relying on the new method of primary equivalent aggregation, indicates higher agricultural trade due to the improved mapping and the inclusion of secondary trade (needed to maintain balanced), e.g., vegetable oil trade is represented in primary oil crop trade. More

importantly, the Demand-OtherUse decreased dramatically due to the improvements in tracing processed food and feed demand and considering water content in feed inputs in the new methods. Thus, utilizations for food and feed are larger. The remaining other use is mainly demanded by, e.g., fiber crops (cotton), MiscCrop (rubber), OilPalm (industrial (non-biofuel) use of palm oil), and utilization for seed, loss, storage variation, or remaining processing of other agricultural products. As a result, fundamental changes in SUA balance are seen for most agricultural commodities (see Fig. 3.6 for more details). Impacts on OilPalm, SugarCrop, and OilCrop are particularly larger since relatively higher shares of the commodities are used for processing and feed. In addition, the production of livestock products is also higher in the updated data. Note that the future demand growth in "other use" (NonFoodDemand in GCAM) is only driven by population (no price and income elasticity) in GCAM. More food/feed use and less "other use" as generally suggested by the new data could lead to a different future total demand if future demand growth is different between food/feed and "other use". Note that some commodities (e.g., Dairy, Veg, etc.) also had higher other use, implied by our data (e.g., loss, storage diff, non-biofuels industrial use, remaining processed use).

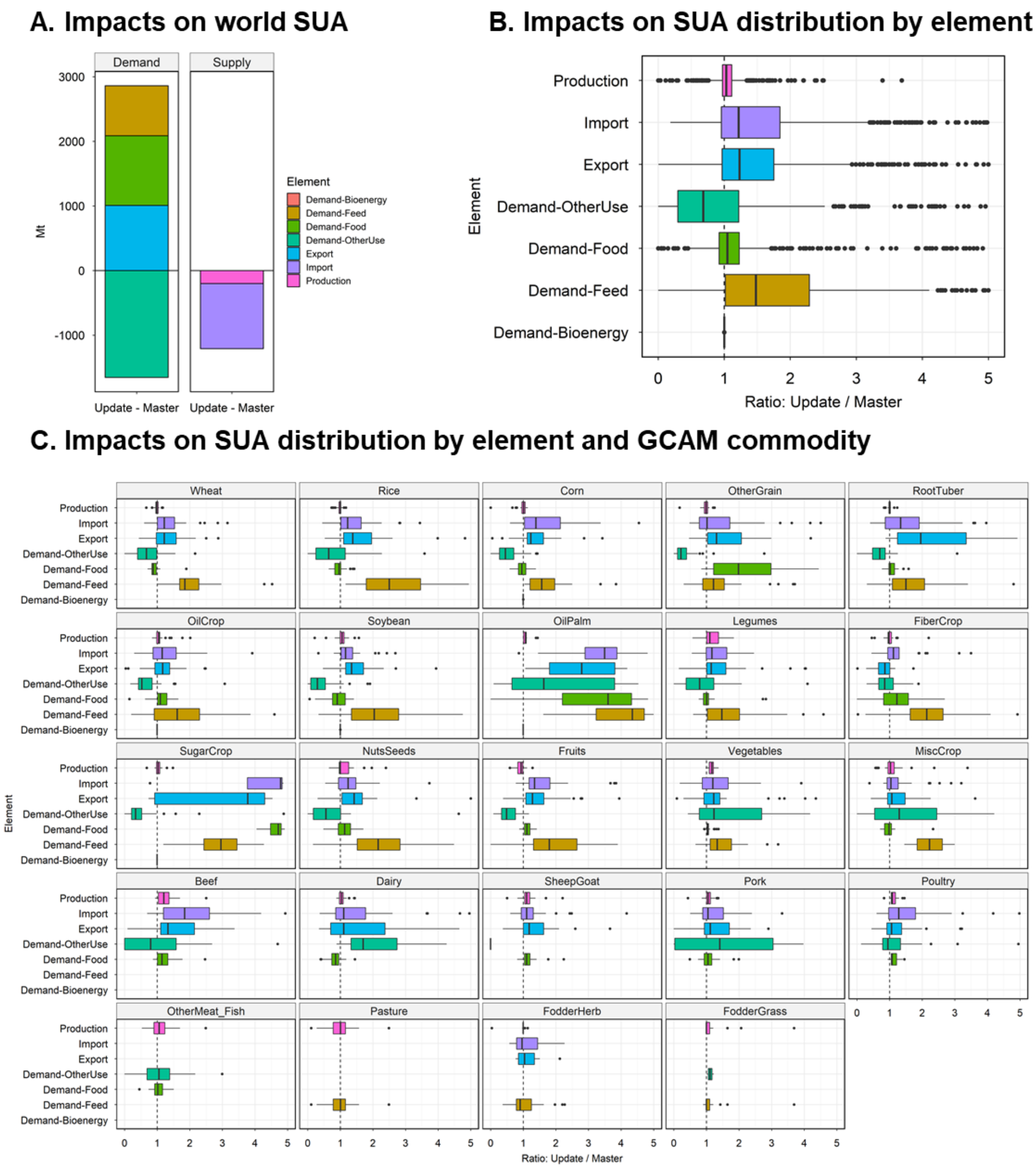


Fig. 3.6 Impact of the data update on the base year agricultural SUA balance. The subfigures include the difference (Update – Master) for all agricultural commodities (A), distributions of the ratio (Update / Master) across commodity x region combinations by SUA elements (B), and the distributions or the ratio across GCAM regions by commodity (C).

A. World total Calorie consumption by commodity

Sector	AgLUUpdate (Pcal)	Master (Pcal)	Change
OtherMeat_Fish	151	196	-23%
SheepGoat	29	36	-18%
FiberCrop	29	35	-17%
OilCrop	276	307	-10%
Soybean	225	250	-10%
Corn	427	465	-8%
MiscCrop	42	44	-5%
Rice	1432	1496	-4%
Wheat	1455	1513	-4%
Pork	367	363	1%
OilPalm	153	149	3%
Legumes	178	170	4%
SugarCrop	566	542	4%
Fruits	273	255	7%
RootTuber	412	369	12%
Beef	113	97	16%
Dairy	460	394	17%
Poultry	251	215	17%
NutsSeeds	152	130	17%
OtherGrain	302	238	27%
Vegetables	271	201	35%

B. Calorie conversion rate distribution by commodity

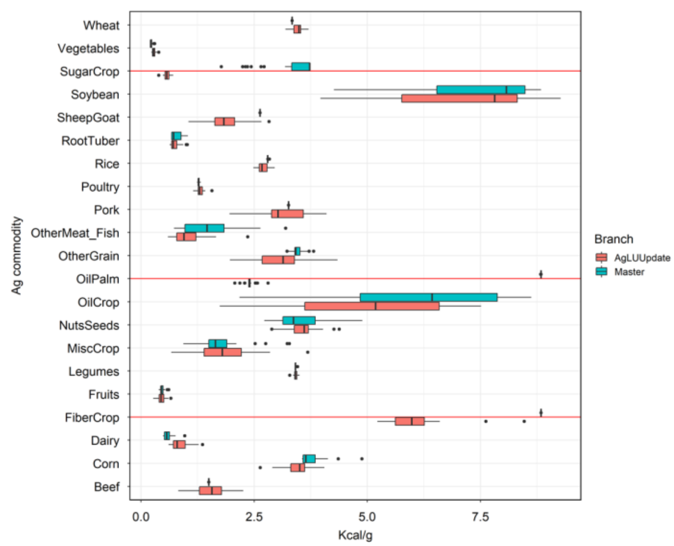


Fig. 3.7 Comparison of base year world total Calorie consumption by commodity (A) and Calorie conversion rate distribution by commodity (B) between AgLU-Update and Master.

The total world food calorie consumption is 7564 Pcal in the updated branch, which is about 100 Pcal larger than the Master (7466 Pcal). The sectoral difference could be striking. The improved data show -23% to +35% changes in food calorie available across food commodities. Note that a few sectors (e.g., alcoholic beverage, infant food, etc., accounting for ~5% of calorie consumption) are not include in the data due to inconsistent mappings. They are assumed to be exogenous in the new method.

A comparison of regional caloric consumption per capita per day is presented in **Fig. 3.8**. The calorie conversion rates are also calculated in a way consistent with the new SUA method because the food consumption of a commodity includes both primary and processed food consumptions which have different calorie contents. Notably, the calorie conversion rate for SugarCrop, OilPalm, and FiberCrop (mainly cottonseed oil) become considerably lower with the update since their processed food consumption is represented in primary equivalent.

Region	Staples (Kcal/ca/day)			NonStaples (Kcal/ca/day)			Total Food Demand (Kcal/ca/day)		
	AgLUUpdate	Master	Change	AgLUUpdate	Master	Change	AgLUUpdate	Master	Change
South America_Northern	1.1	1.5	-28%	1.4	1.5	-12%	2.5	3.1	-20%
Canada	1.1	1.2	-13%	2.0	2.4	-15%	3.1	3.6	-14%
USA	1.1	1.2	-8%	2.1	2.5	-15%	3.1	3.6	-13%
EU-12	1.4	1.6	-12%	1.7	2.0	-12%	3.1	3.5	-12%
EU-15	1.2	1.3	-8%	2.1	2.3	-12%	3.2	3.6	-11%
Europe_Eastern	1.5	1.5	-3%	1.5	1.7	-14%	2.9	3.2	-9%
European Free Trade Association	1.0	1.1	-6%	2.1	2.3	-9%	3.1	3.4	-8%
Europe_Non_EU	1.6	1.8	-12%	1.8	1.8	-1%	3.4	3.6	-7%
Russia	1.5	1.6	-4%	1.7	1.9	-9%	3.2	3.4	-6%
Pakistan	1.2	1.3	-3%	1.0	1.1	-9%	2.2	2.4	-6%
Japan	1.2	1.2	-3%	1.2	1.3	-8%	2.4	2.6	-6%
Mexico	1.4	1.6	-10%	1.6	1.6	1%	3.0	3.2	-5%
Africa_Northern	2.1	2.2	-6%	1.2	1.2	0%	3.3	3.4	-4%
Australia_NZ	1.0	0.9	4%	2.2	2.3	-7%	3.1	3.3	-3%
Brazil	1.1	1.3	-12%	2.0	2.0	2%	3.2	3.3	-3%
Southeast Asia	1.5	1.7	-10%	1.1	1.0	13%	2.6	2.7	-1%
Indonesia	2.0	2.1	-6%	0.9	0.8	9%	2.9	2.9	-1%
South Africa	1.6	1.7	-6%	1.2	1.1	7%	2.8	2.8	-1%
Central America and Caribbean	1.2	1.2	-2%	1.4	1.4	2%	2.6	2.6	0%
South America_Southern	1.3	1.4	-3%	1.3	1.3	5%	2.6	2.6	1%
India	1.4	1.5	-4%	1.0	0.9	9%	2.5	2.4	1%
South Korea	1.4	1.5	-9%	1.7	1.5	12%	3.1	3.0	2%
Central Asia	1.5	1.6	-7%	1.4	1.2	21%	2.9	2.8	5%
Colombia	1.1	1.2	-12%	1.8	1.5	20%	2.9	2.7	6%
Argentina	1.2	1.2	4%	1.9	1.7	9%	3.1	2.9	7%
Africa_Southern	1.4	1.4	-1%	0.8	0.6	24%	2.2	2.1	7%
China	1.7	1.7	4%	1.5	1.3	18%	3.2	2.9	10%
South Asia	1.9	1.8	5%	0.6	0.5	29%	2.5	2.3	10%
Africa_Eastern	1.4	1.3	8%	0.8	0.7	15%	2.2	2.0	11%
Taiwan	1.0	0.6	61%	1.9	2.0	-4%	2.9	2.6	12%
Africa_Western	1.7	1.5	13%	0.8	0.7	16%	2.5	2.2	14%
Middle East	1.5	1.3	16%	1.3	1.1	19%	2.8	2.4	18%

Fig. 3.8 Comparison of the base year regional caloric consumption per capita per day

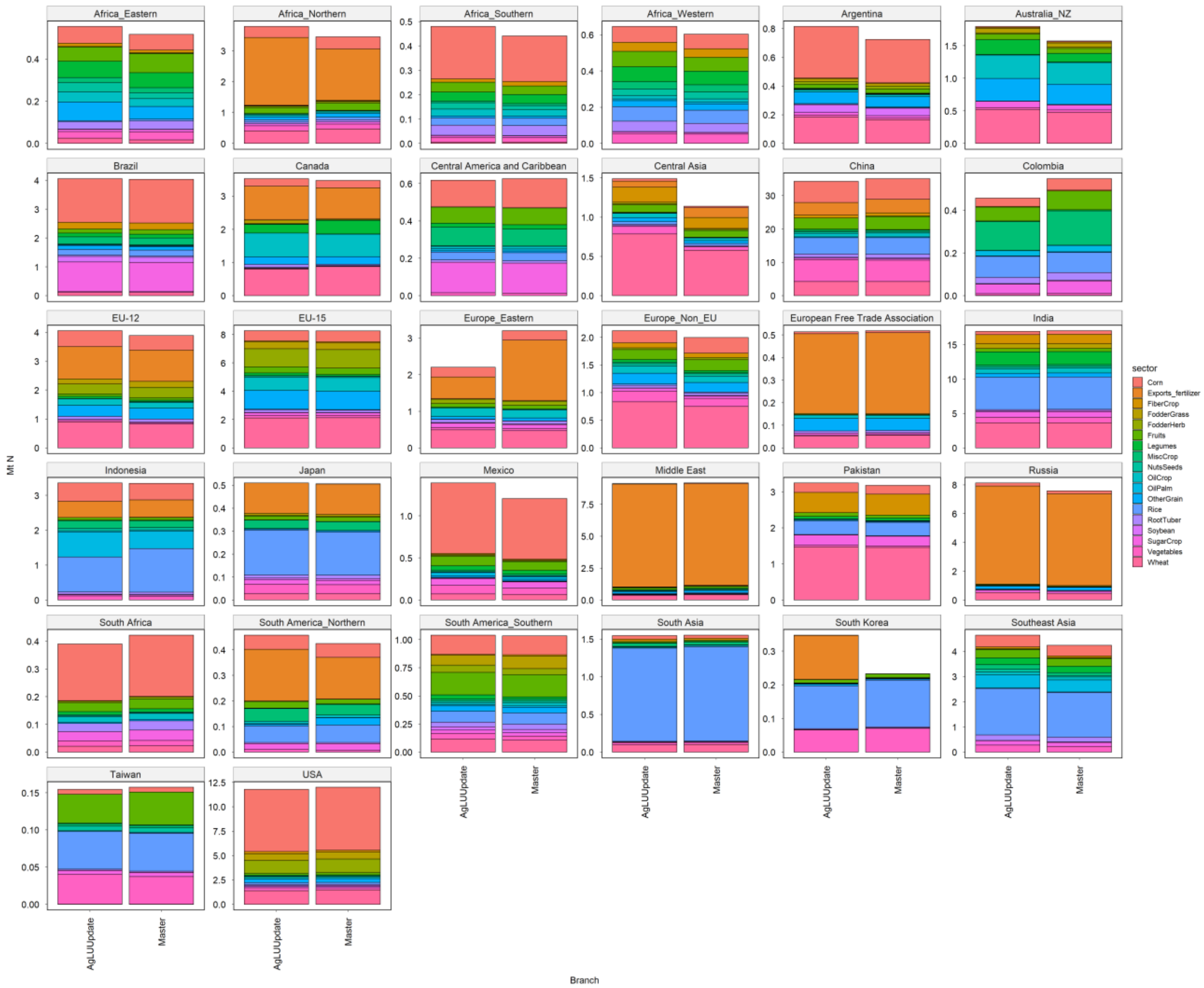


Fig. 3.9 Comparison of regional fertilizer demand (by crop) and trade between AgLU-Update and Master. Overall, the fertilizer demand did not change significantly due to the data update in the major producing or consuming regions, e.g., Russia, USA, China, India, etc. Trade modeling of N fertilizer can be incorporated in GCAM in future studies.

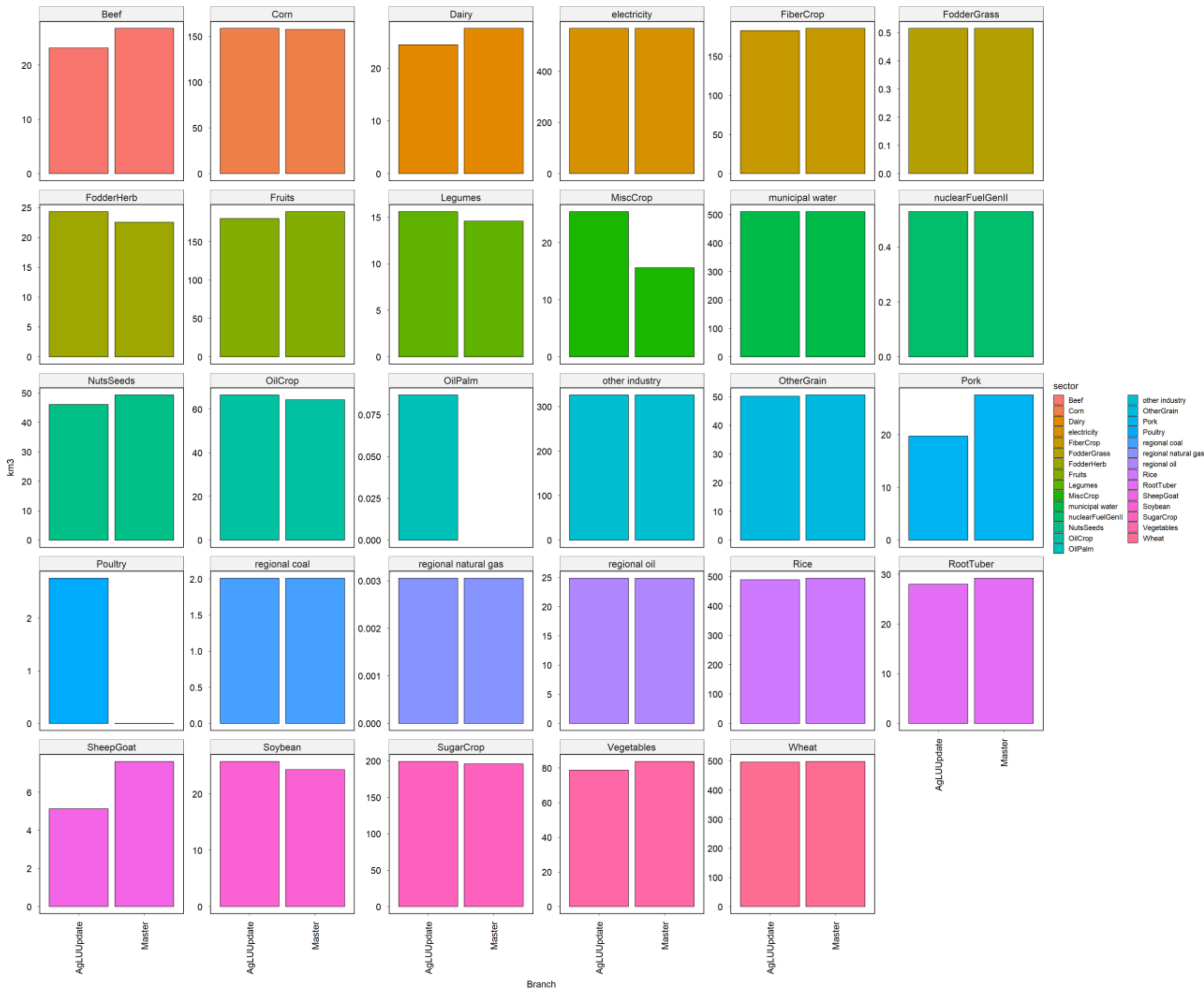
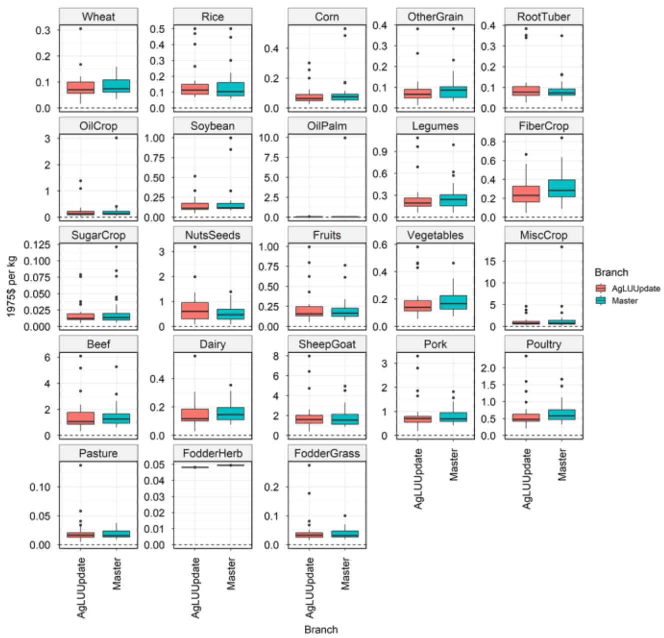


Fig. 3.10 Comparison of global water withdrawal by sector between AgLU-Update and Master. Livestock sectors have lower water use in the updated branch due to the animal stock data and mapping update, except for Poultry (no water use likely a mapping bug before). The updated mapping file also improved the water use for OilPalm and MiscCrop.

A. Distribution over GCAM regions by sector



B. Global mean prices by sector

Sector	AgLUUpdate (1975$/kg)	Master (1975$/kg)	Change
Dairy	0.105	0.152	-31%
Legumes	0.161	0.231	-30%
OilPalm	0.032	0.046	-30%
FiberCrop	0.245	0.297	-18%
Poultry	0.528	0.623	-15%
SugarCrop	0.014	0.016	-15%
Beef	1.107	1.290	-14%
MiscCrop	0.607	0.696	-13%
Wheat	0.066	0.075	-12%
OtherGrain	0.057	0.064	-11%
OilCrop	0.131	0.144	-9%
Rice	0.129	0.139	-8%
Corn	0.064	0.067	-5%
Pork	0.724	0.754	-4%
FodderHerb	0.048	0.049	-3%
Soybean	0.105	0.104	0%
SheepGoat	1.914	1.870	2%
Fruits	0.182	0.178	3%
Pasture	0.018	0.017	8%
FodderGrass	0.039	0.036	9%
RootTuber	0.079	0.072	10%
Vegetables	0.172	0.155	11%
NutsSeeds	0.438	0.373	18%

Fig. 3.11 Comparison of base year producer price distribution by sector (A) and world mean prices (B) between AgLU-Update and Master. With the data update, the global mean producer price could change by -31% to +18% across the agricultural sectors. Less extreme outliers are seen at the regional level (e.g., MiscCrop and OilPalm).

Sector	AgLUUpdate (Gt CO2eq)	Master (Gt CO2eq)
CH4 (energy)	2.37	2.37
CH4_AGR (Ag Production)	3.30	3.30
CH4_AWB (Ag waste burning)	0.06	0.05
CH4_UnMGMTLand (fire)	0.32	0.30
N2O (energy)	0.70	0.70
N2O_AGR (Ag Production)	2.16	2.16
N2O_AWB (Ag waste burning)	0.01	0.01
N2O_UnMGMTLand (fire)	0.27	0.26
Other NonCO2 GHGs	1.01	1.01
Other NonCO2 GHGs_UnMGMTLand (fire)	0.00	0.00

Fig. 3.12 Comparison of base year NonCO2 GHG emissions by source between AgLU-Update and Master. The emissions were mostly not affected by the updates. This would imply a smaller emission intensity if the production increased due to the update.

4. Shared policy assumption (SPA) GCAM validation runs

Following the GCAM CMP convention, we provide GCAM projection results (comparing *AgLU-Update* and master branches) from reference & RCP 2.6 scenarios across shared socioeconomic pathways (GCAM core & SSP1-5 assumptions; not including SSP3-RCP2p6). In the following, we present the key global results in the following figures. Note that most SPA results (**Figs. 4.1 – 4.15**) are an extension of base year results (i.e., figures shown in **Section 3**) to future periods. Again, the figures are mostly self-explanatory, but key changes and insights are discussed the captions.

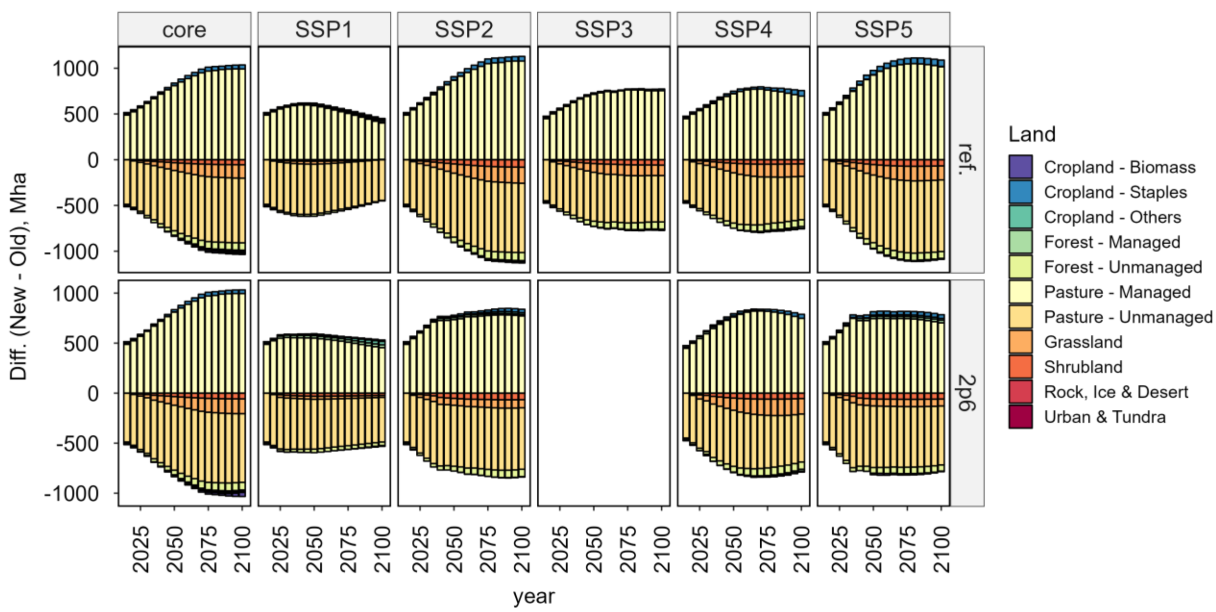


Fig. 4.1 (extension of Fig. 3.1B) Land cover difference between the *AgLU-Update* and *Master* branches. As expected, the impacts on the future projections are mainly due to initial data change, i.e., more managed pasture from unmanaged pasture. However, managed pasture would have a relatively greater future demand due to the data update, leading to a future decrease in natural land. This is mainly driven by the higher future ruminant livestock product consumption (in the baseline driven by population & GDP growth) and lower pasture yield. Thus, a lower initial pasture yield (fixed in future periods) leads to even higher future pasture demand. Changes in cropland areas are relatively smaller compared with pasture changes (see **Fig. 4.2**).

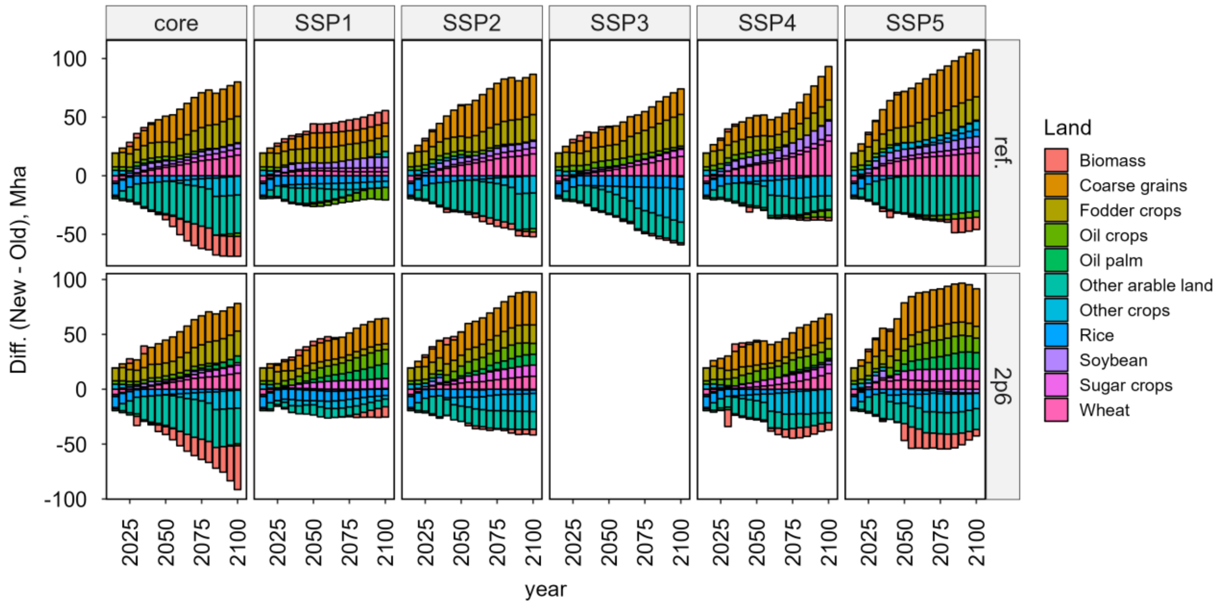


Fig. 4.2 (extension of Fig. 3.2B) Cropland area (by crop) difference between the *AgLU-Update* and *Master* branches. Similar to higher future pasture demand, there are overall higher feed (e.g., fodder crops, coarse grains, etc.) demand driven by the higher future livestock products production. Other arable land and biomass are main land sources to meet higher future land demand for feed crops. Note that there are spikes in biomass land in some scenarios (e.g., SSP4-2p6) when examining the area difference between the branches. But there are no spikes in the land use change results (**Fig. 4.3**).

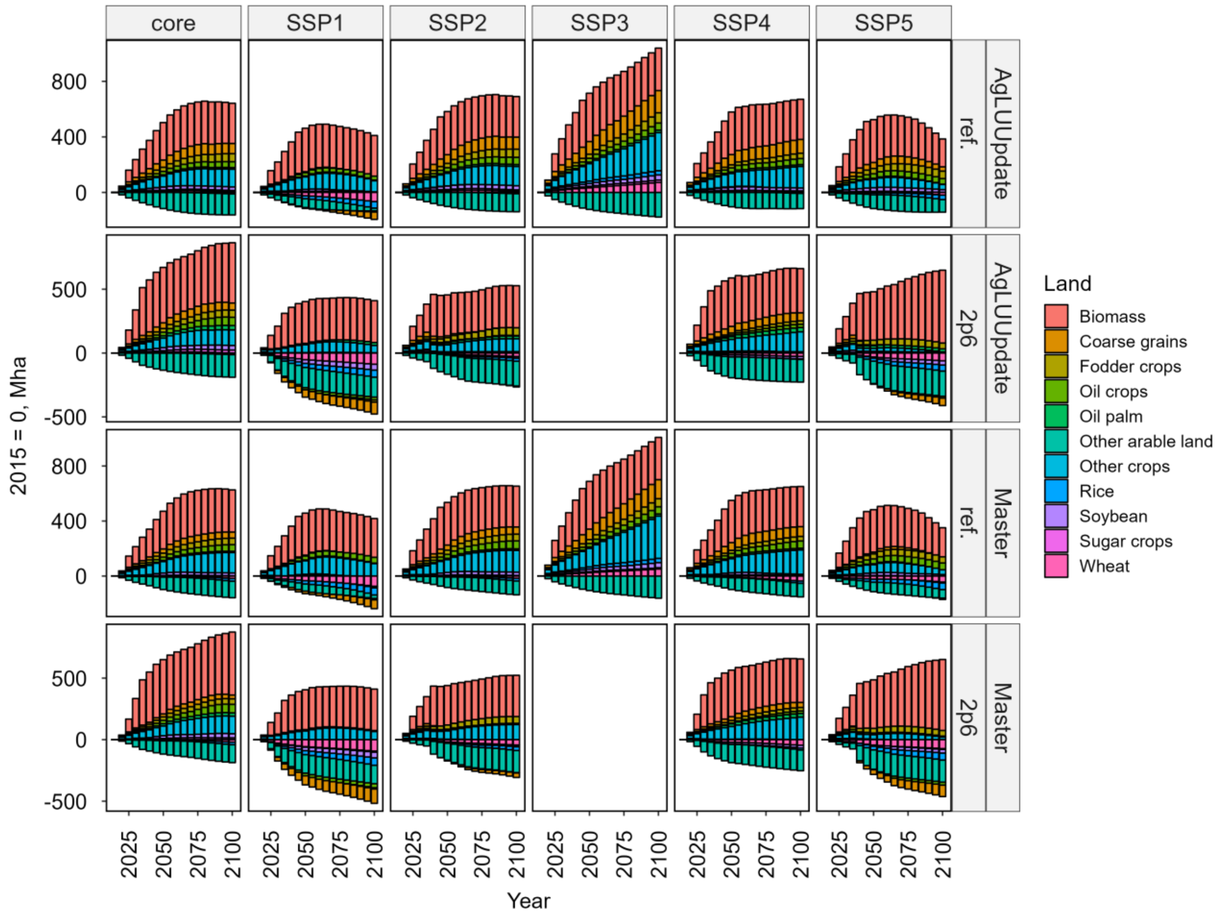


Fig. 4.3 Cropland area (by crop) change relative to the base year in the *AgLU-Update* and *Master* branches across SPA runs.

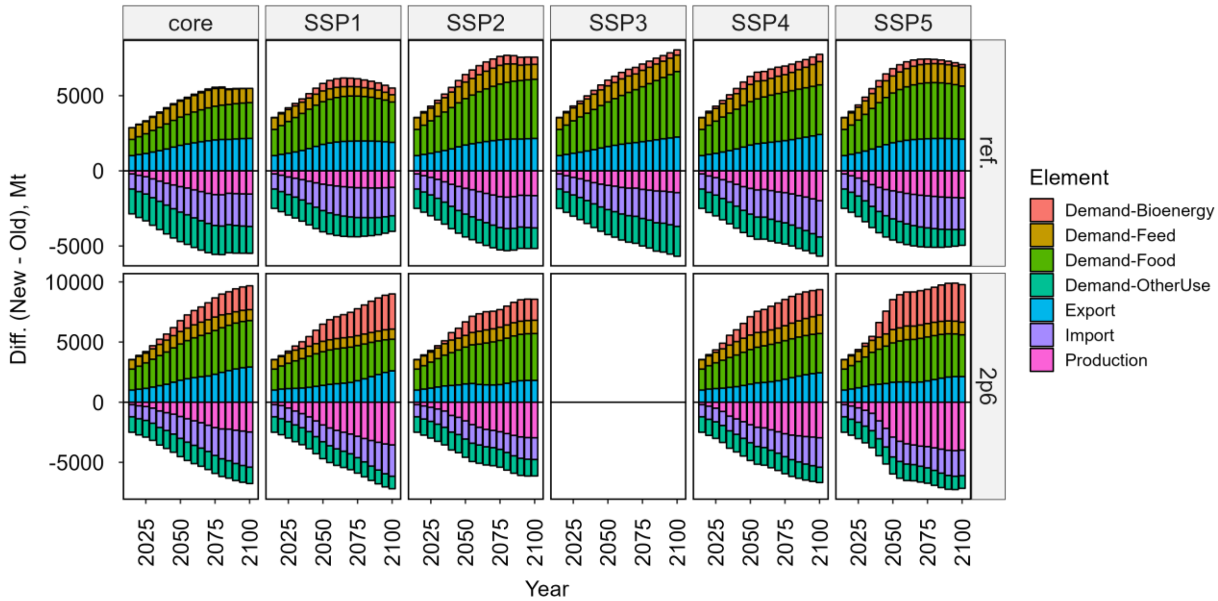


Fig. 4.4 (extension of Fig. 3.6A) agricultural SUA balance difference between the *AgLU-Update* and *Master* branches for an aggregated commodity. At the global aggregated scale, the initial data differences are magnified over time. Agricultural production becomes increasing larger over time across all scenarios driven by the overall higher initial crop yields and feed conversion rates. The higher agricultural supply encourages higher overall demand for food, feed, and first-generation bioenergy. Trade (world export = world import) also expands more with the higher global agricultural productivity. **Figs. 4.5 – 4.9** present more detailed results by sector and element.

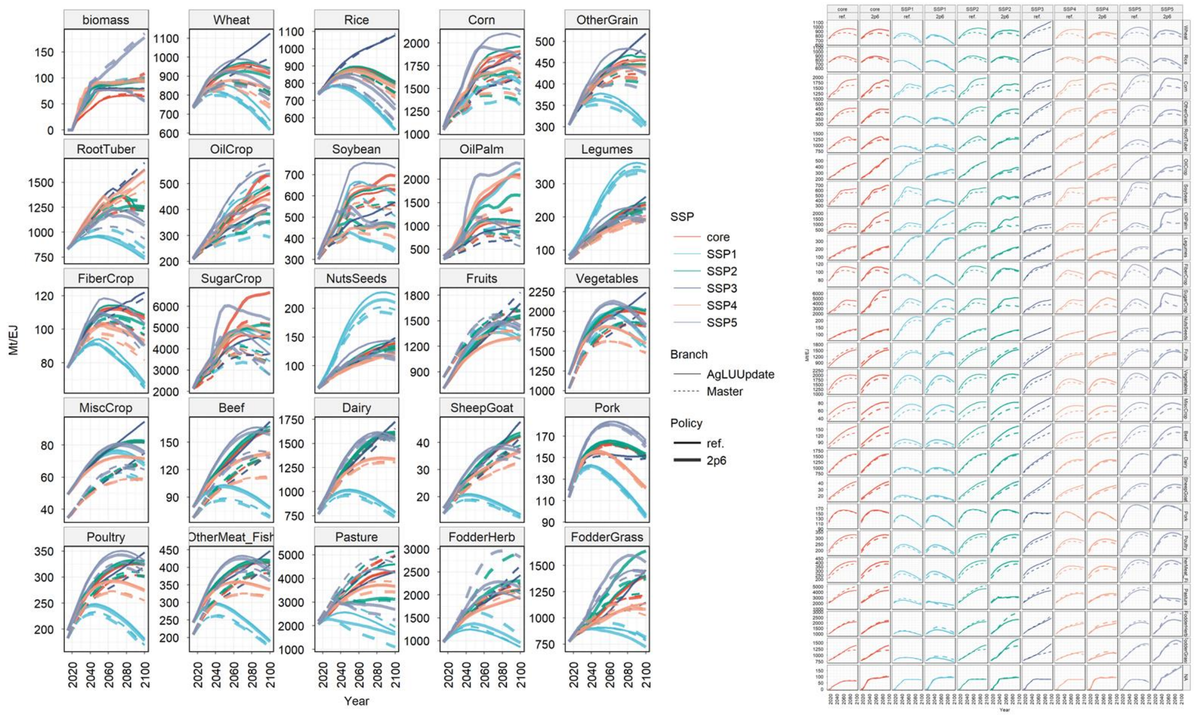


Fig. 4.5 (extension of Fig. 3.5B-Production) Comparison of Ag production projections by sectors across SPA runs. Note that the two figures (left and right) presented here have different facet formats but the same data. Initial data change is the main factor explaining the shifts in future projections, particularly for Legumes, Fruits, Vegetables, MiscCrop, Dairy, Pork, etc. Also, consistent with the land use change results, there is higher future production of meat products and crops with feed uses (e.g., Wheat, Corn, OtherGrain, OilPalm, etc.). Note that for most crops, the "other use" became much smaller in the new/improved SUA data. The future demand growth in "other use" or NonFoodDemand is only driven by population (no price and income elasticity) in GCAM. In contrast, food/feed has a relatively larger future demand growth due to income/GDP growth. Thus, the shifting of "other use" to food/feed use implied by our data is leading to higher future total demand (and thus production). This is particularly the case for palm and sugar crops as they had a larger portion of "other use" before.

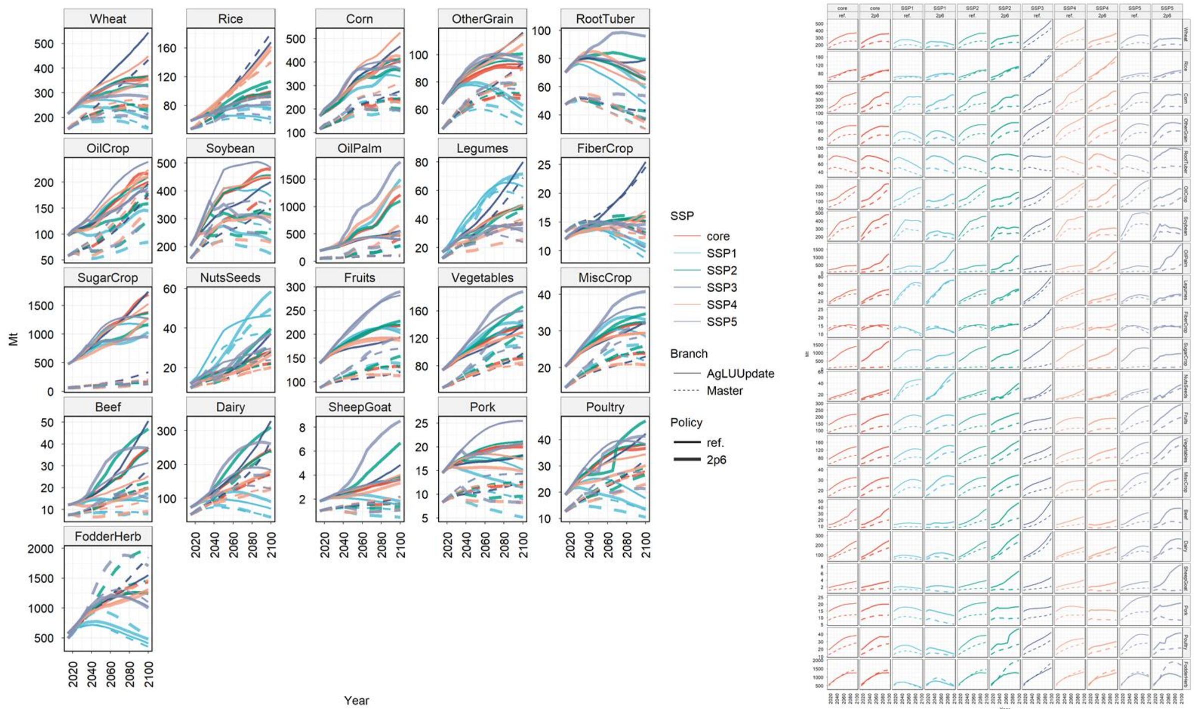


Fig. 4.6 (extension of Fig. 3.5B-Export / Import) Comparison of Ag trade projections by sectors across SPA runs. Note that the Armington framework is used for all commodities here except FodderHerb (integrated world market). In most cases, future trade shifts with the initial data change. But for commodities with large secondary product trade consistently considered in the new method, e.g., raw/refined sugar (SugarCrop) and OilPalm (palm oil), more significant future changes are seen due to the higher total future demand (see Fig. 4.5 caption).

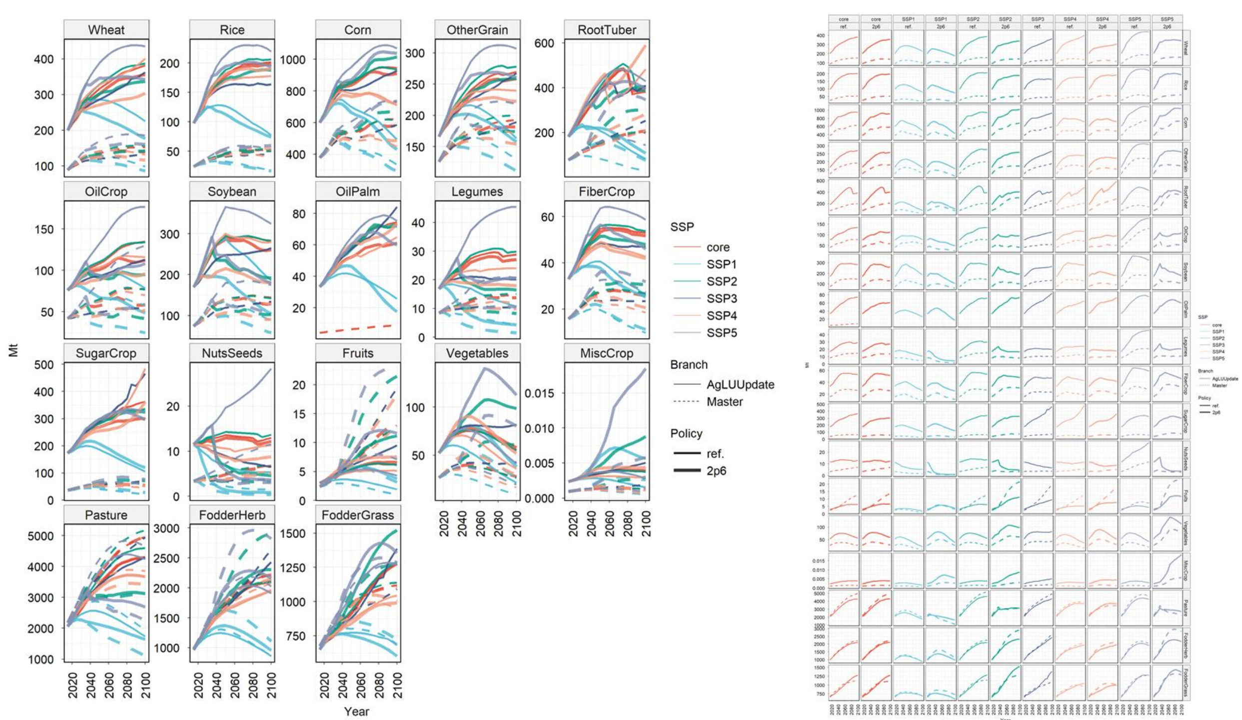
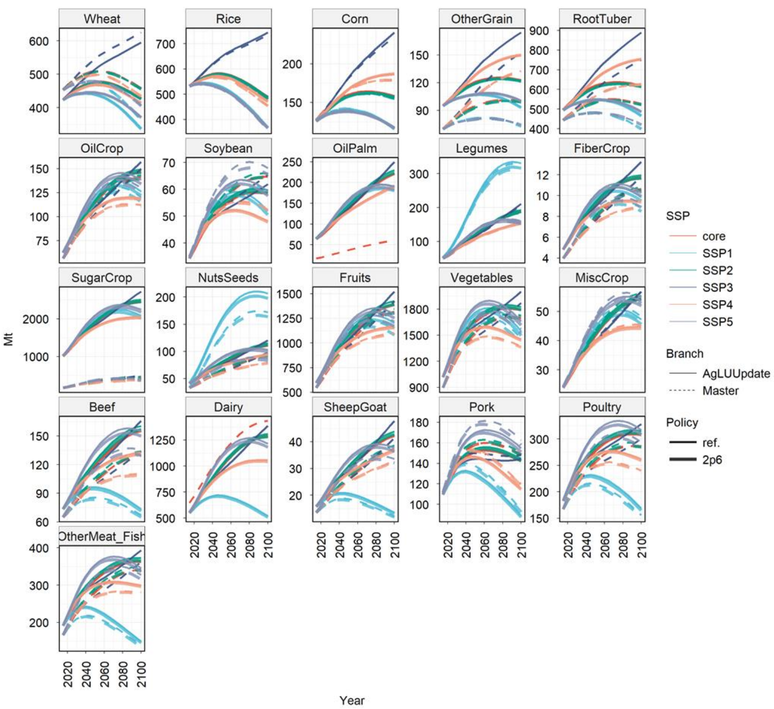


Fig. 4.7 (extension of Fig. 3.5B-Feed) Comparison of Ag feedstuff demand projections by sectors across SPA runs. The new data imply a relatively higher feed use of food crops and related products, and the pattern is also magnified in future periods. For OilPalm, palm kernel meal is now included in the source data.



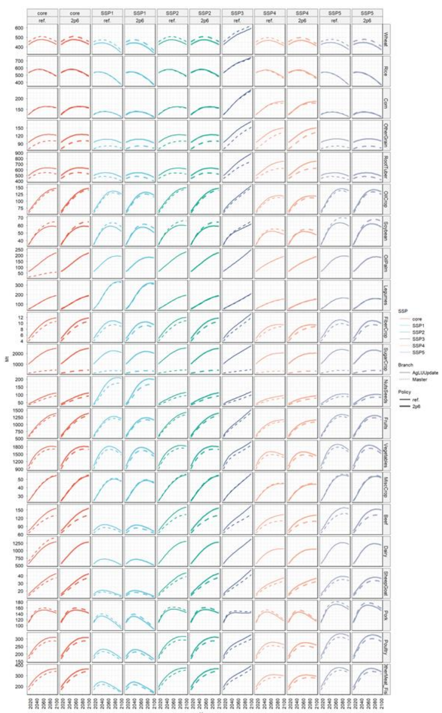


Fig. 4.8 (extension of Fig. 3.5B-Food) Comparison of Ag food demand projections by sectors across SPA runs. Note that the food consumption is in primary equivalent, so the unit is the same as the primary product. E.g., palm oil and raw sugar are represented using oil palm fruit and sugar crops, respectively. Higher food consumption in most cases could reflect this new method. But the primary equivalent approach does not affect the total calorie accounting (**Fig. 4.9**).

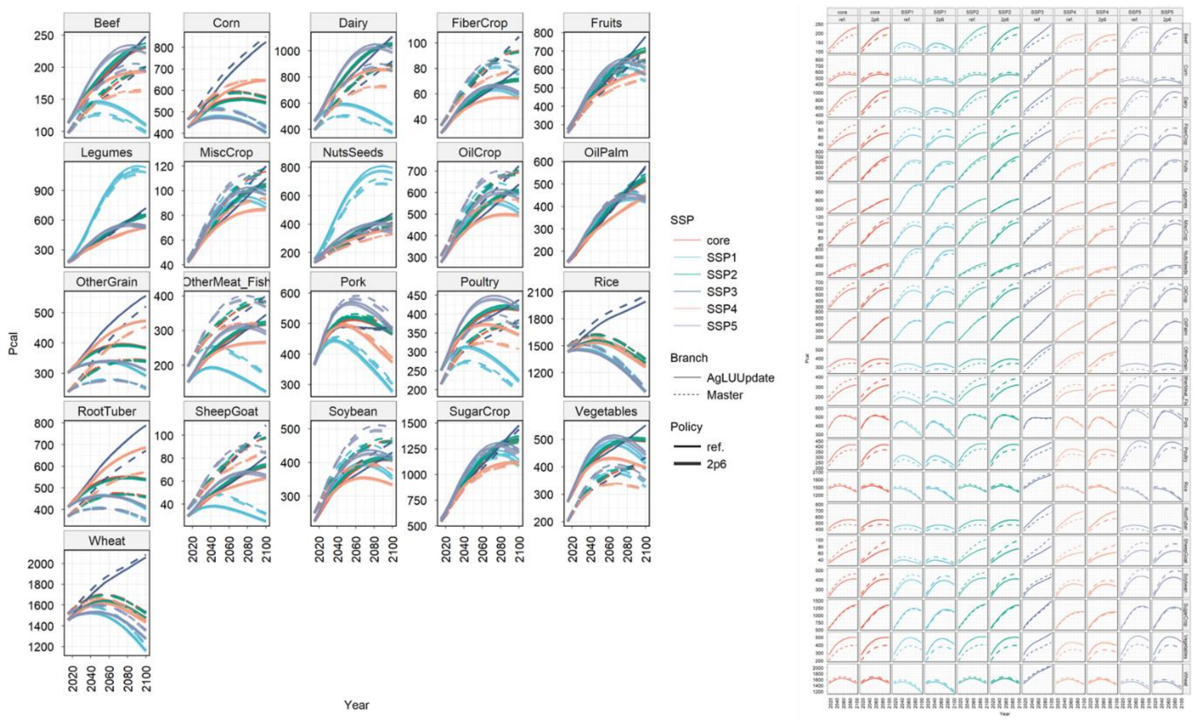


Fig. 4.9 (extension of Fig. 3.7A) Comparison of Ag food calorie projections by sectors across SPA runs. The changes in food calorie projections generally reflect the improved calorie data from FAO.

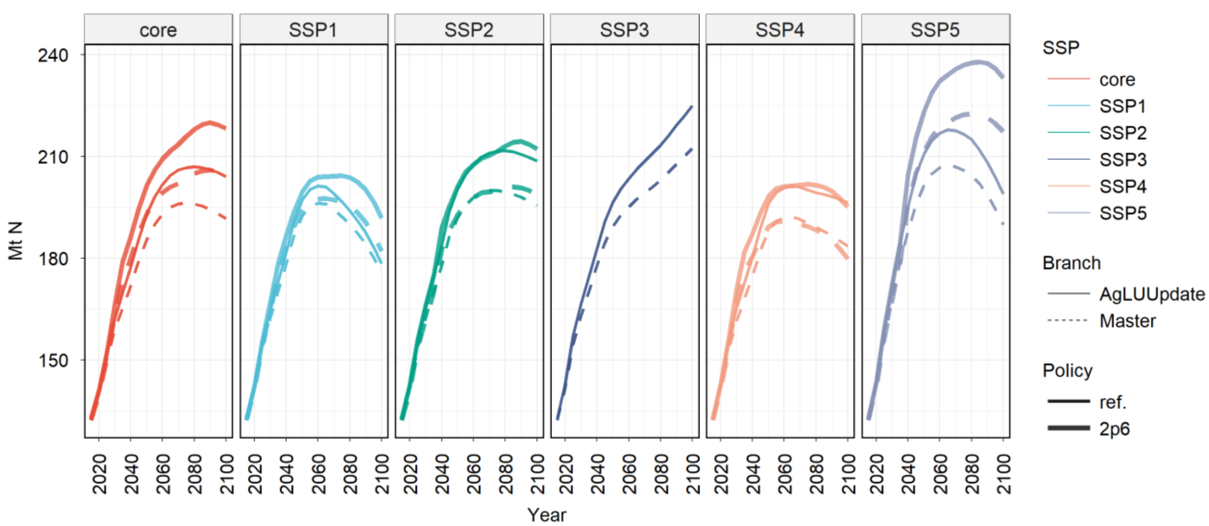


Fig. 4.10 (extension of Fig. 3.9 at the world level) Comparison of total fertilizer demand across SPA runs. The higher overall future fertilizer demand is driven by the higher crop production.

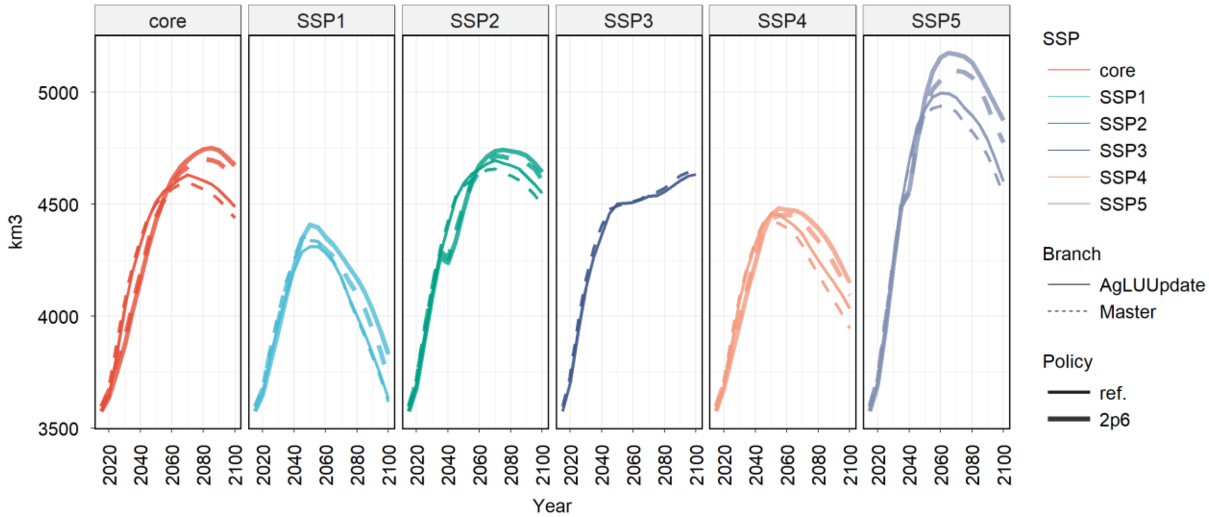


Fig. 4.11 (extension of Fig. 3.10 at the world level) Comparison of total water withdrawal across SPA runs. With the new data, the livestock water use intensity is relatively lower. However, the total water withdrawal is generally higher, driven by the higher future agricultural production.

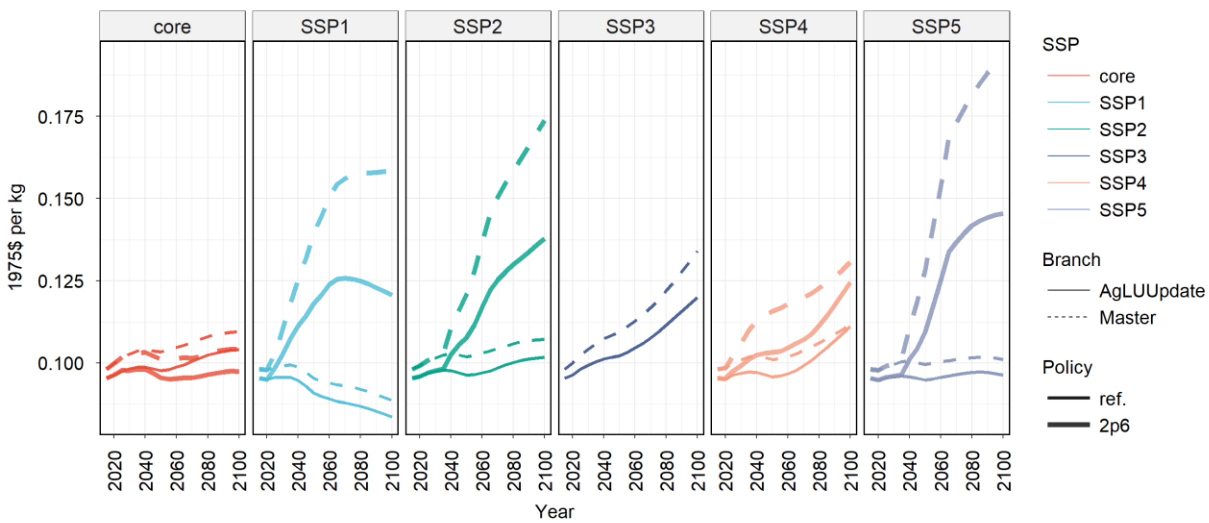


Fig. 4.12 (extension of Fig. 3.11B) Comparison of world Ag price (including crops and livestock sectors) projections by sectors across SPA runs. The agricultural prices are lower at the global average value in the base data. The future price changes are also more moderate, particularly under RCP2p6 scenario, because of the relatively higher overall productivity. For reference scenarios, the future price changes were mainly from the initial shift. The impacts from the higher future demand due to lower "other use" was mostly compensated by the higher overall productivity. However, in RCP2p6 with land carbon pricing, the higher productivity (and other data changes, e.g., more trade) helped more in alleviating extreme price changes.

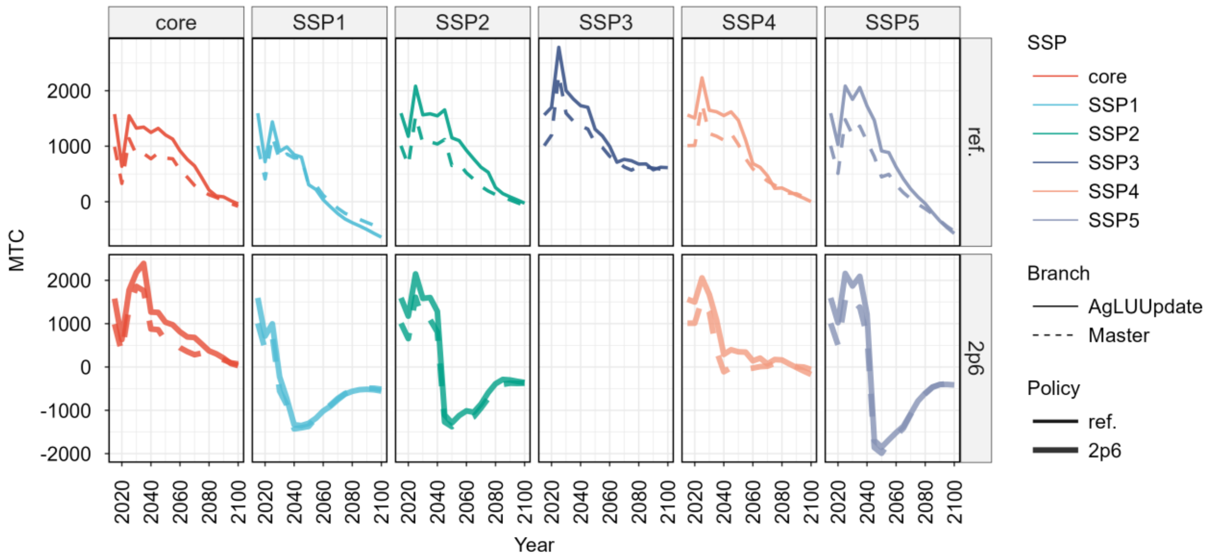


Fig. 4.13 Comparison of land use change (LUC) carbon emission projections by sectors across SPA runs. With the updates, LUC emissions are higher in the earlier periods due to the higher agricultural land demand.

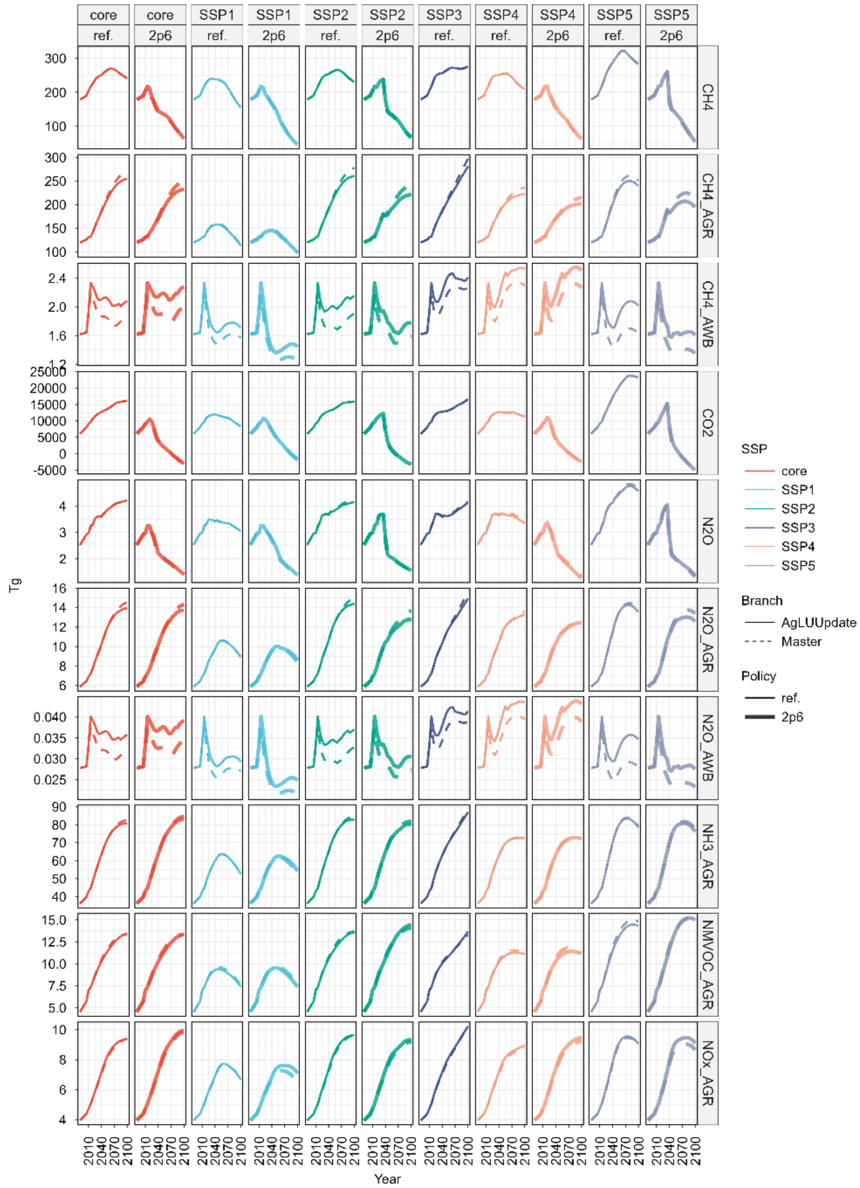


Fig. 4.14 (extension of Fig. 3.12) Comparison of key emission projections across SPA runs. There were some important implications for non-CO2 emissions. For CH4_AGR and N2O_ARG, the total emissions were the same initially, while the implied emission intensity per Ag production became lower because of the higher production initially. As a result, future CH4_AGR and N2O_ARG emissions are relatively lower even with higher Ag and beef production. But for the "_AWB" (waster burning) emissions, the increase was driven by the high future crop production.

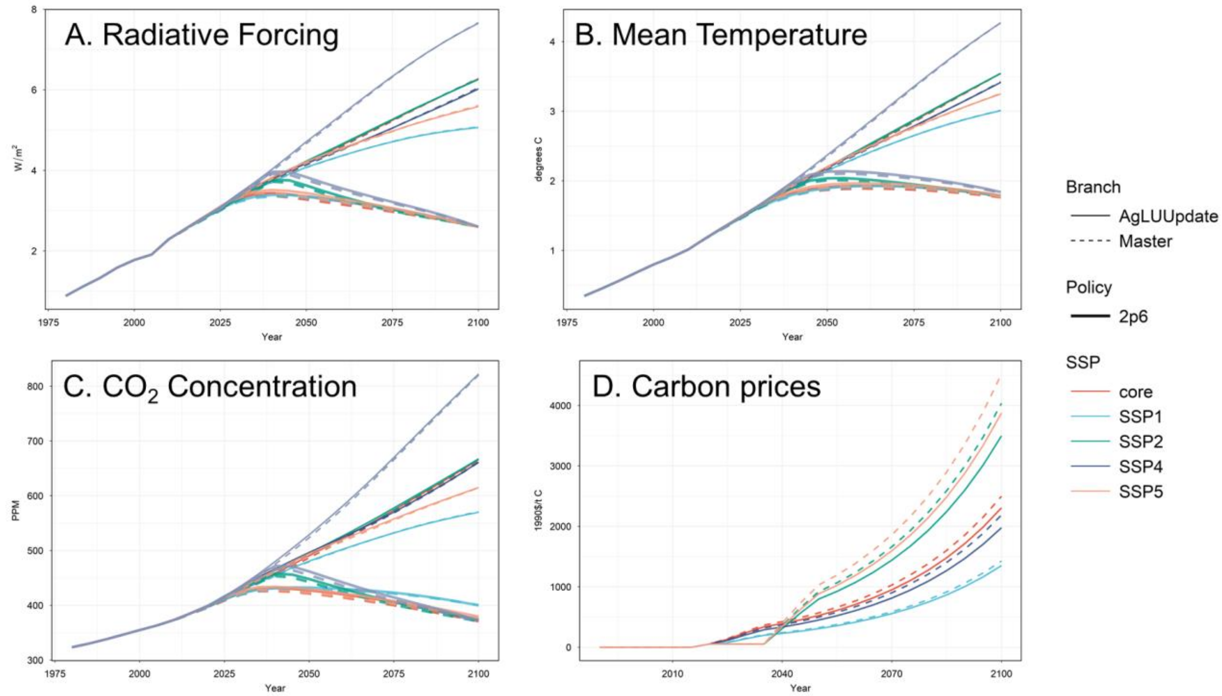


Fig. 4.15 Comparison of key climate variables and carbon price projections across SPA runs. Overall, the Ag data updates lead to slightly lower climate outcomes (forcing, temperature, and concentration), mainly due to the lower future nonCO2 emissions and higher productivity. So C prices become slightly lower, implying relatively smaller mitigation efforts are needed when targeting 2.6.

5. Future work

In an early CMP focused on forest trade and demand, we discussed the importance of the base data and the need for a more consistent, transparent, and traceable method to process FAO data for global economic modeling. These can be finally crossed from the To-do list in that CMP (other future tasks list there could also be important for future development).

This CMP is closely related to a few past and ongoing GCAM AgLU studies or CMPs, e.g., land allocation (Wise et al., 2014), food demand (Edmonds et al., 2017), trade (Zhao et al., 2022, 2021), crop-remapping, BYU, hindcast, GCAM-macro, and agricultural storage developments, all of which rely on quality data. There are aspects that could be improved in future work:

- Future changes in livestock stocks will have a considerable impact on future global economic and environmental projections. So revisiting fodder crops, feed sources, and livestock product IO data could improve our understanding. We are still extrapolating fodder crop data from FAO after 2011. The quality was low even before 2011. Also, livestock carcass yield and feed conversion rate are assumed to be fixed currently. These efficiency factors likely will improve in the future periods with technological progress, which should be studied more carefully.

- gcamdata downscaling right now scale harvested area to land cover. The uncertainty there could be large. Future changes in harvest frequency could be important as well.

    - We had a chance to future improve the data processing to fix the scaling process that allowed the multi-cropping on tree cropland. This fix is merged in this CMP though the impact was not reflected in the results above.

    - In short, the tree crops had a lower yield, and their land area matched FAO harvested area now since they cannot multi-cropping. For non-TreeCrops, we recomputed the crop harvest frequency at basin levels to scale harvest area to cropland cover (and yield is adjusted accordingly). The overall impacts on energy systems and emissions are negligible. The impacts on AgLU results are also generally moderate and reasonable. The initial data change leads to a recalibration of our parameters and future changes are small. The 2015 LUC emissions increased by ~7% but were smaller in later years (mostly <2% or around 10 MtC).

- Ag cost data is not a focus of this update because we did not use FAO data. But the cost data will be revisited to separate labor and capital. AgLU IO table can be generated after that.

- Possible double counting of residual biomass supply. GCAM currently has two separate residual biomass supply sources: (1) a supple curve for livestock feed from residues and (2) a residual supply linked to crop production for bioenergy use. The accounting should be examined later. Residual biomass supply (for bioenergy) assumptions should also be examined and updated since the currently used ones are dated (Gregg and Smith, 2010).

- Improving pasture productivity and grazing intensity data.

- Historical data should not be changing across scenarios unless there is a good reason. Currently, land data could change, e.g., the share of managed pasture and managed forest, when different land carbon data (e.g., Q3 or median from SoilGrid) is used, mainly because of the linkage between pasture & forest productivity and carbon density. This linkage should be broken in future updates!

- The land supply parameters, including both unmanaged land value and land logit exponents, should be further studied and updated. The implied land supply elasticities have changed considerably after the land protection CMP, leading to more sensitive land use changes in response to prices (compared to literature values).

- In the Ag storage work, we will separate storage and also possibly secondary Ag trade.

- gcamdata-faostat will be published as a separate package so a broader community can use more harmonized data.

6. Supplementary information (gcamdata-faostat)

6.1. Maintain agricultural supply-utilization balance in data and modeling

a. Balance across supply – utilization
   - $Opening\ stock_{reg,\ t} + Production_{reg,\,t} + Import_{reg,\,t} = Closing\ stock_{reg,\,t} + Consumption_{reg,t} + Export_{Reg,\,t}$

b. Balance across space
   - $\sum_{reg} Import_{reg} = \sum_{r} Export_{reg}$
   - $\sum_{reg} Net\ trade_{reg}$

c. Balance across time
   - $Opening\ stock_{t} = Closing\ stock_{t-1}$
   - $Opening\ stock_{t} + Stock\ variation_{t}\ = Closing\ stock_{t}$

d. Balance across vertical processing sectors
   - Processed use of primary consumption, when applied to an extraction rate, becomes the production of the secondary product.

e. Data quality

gcamdata-faostat includes functions to clean and process the raw data and balance the data in all key dimensions.

6.2. Sectoral aggregation along the processing chain using primary equivalent

Fig. 6.1 illustrates the SUA balance along the processing chain of Maize and products and the processing structure. For example, Maize is first processed into flour, germ, and bran. Flour, if processed, is used to produce starch (and then gluten & meal). Maize germ, if processed, produces oil and cake. There is an SUA balance for all primary and processed products along the supply chain. We improve the FAO method of primary equivalent to represent processed commodities using primary equivalent with the consideration of extraction rate, coproduction, and trade and storage balance.

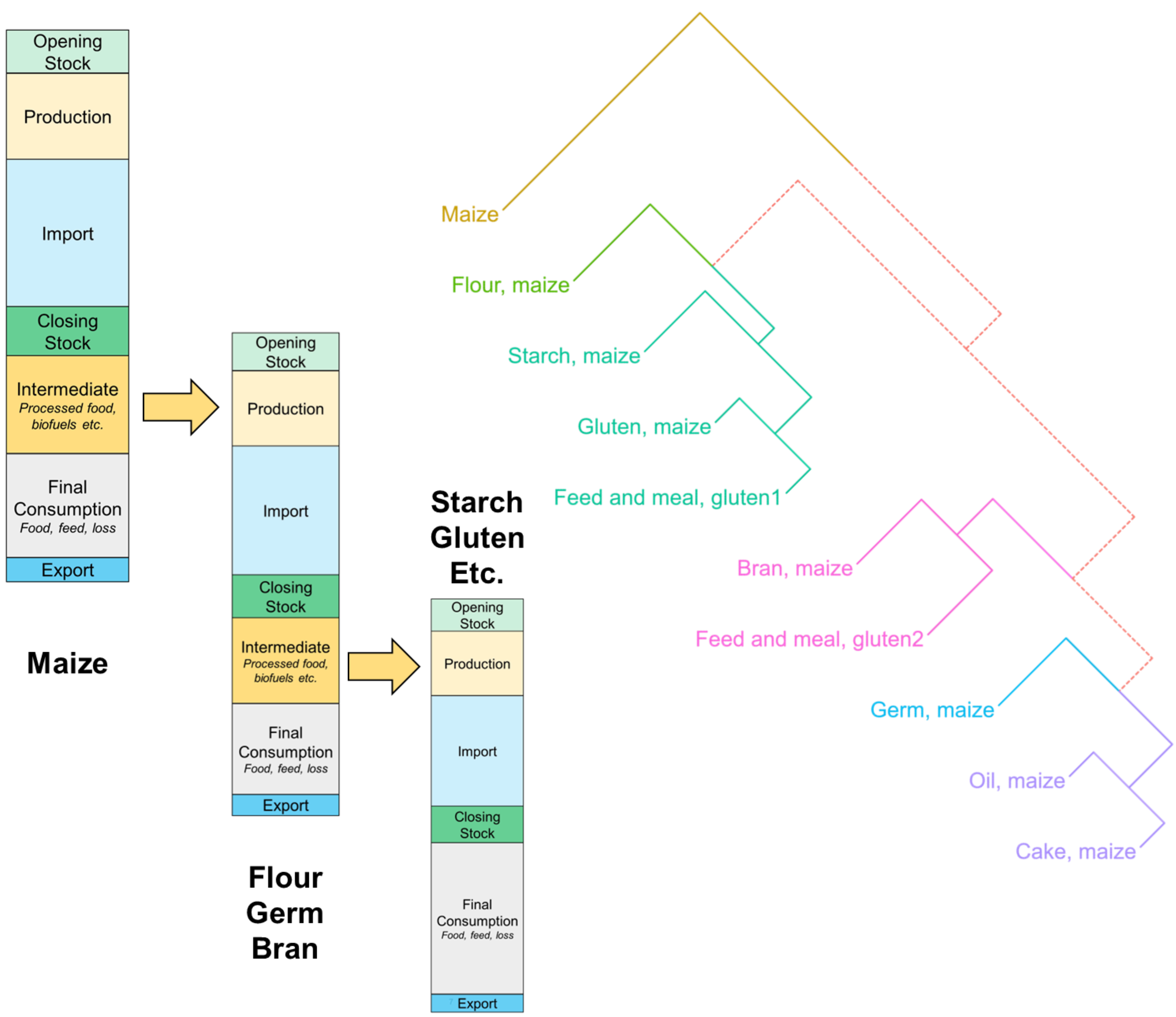


Fig. 6.1 Illustration of the primary equivalent approach for deriving agricultural supply-utilization accounting (SUA) balance, including opening and closing stock using an example of Maize). For example, Maize is first processed into flour, germ, and bran, and flour, if processed,

is used to produce starch (and then gluten & meal). The approach represents processed products in their primary product to bridge the gap between land use and food or bioenergy consumption.

6.3. Can gcamdata-faostat be included in gcamdata?

- The source data used are all publicly available. But it won't make sense to include all the raw data and the processing in the gcamdata framework, given the size of the project (1.8 GB) and additional processing time.
- The repo will be made open-access, e.g., github/JGCRI.
- The long-term vision of the project can be set as an R package for processing and generating agricultural data for global economic modeling. It can include USDA data as well. And it provides functions for generating model-specific files (e.g., GCAM or other models).

Acknowledgments

We appreciate Pralit Patel, Page Kyle, Stephanie Waldhoff, Kanishka Narayan, Abigail Snyder, Ying Zhang, Niazi, Hassan, Steven Smith, and other GCAM team members for their valuable comments, suggestions, and support. Additionally, we extend our gratitude to Jonathan Doelman (PBL Netherlands Environmental Assessment Agency) for generously sharing the livestock sector data in IMAGE v3.2.

References

Bond-Lamberty, B., Dorheim, K., Cui, R., Horowitz, R., Snyder, A., Calvin, K., Feng, L., Hoesly, R., Horing, J., Kyle, G.P., Link, R., Patel, P., Roney, C., Staniszewski, A., Turner, S., Chen, M., Feijoo, F., Hartin, C., Hejazi, M., Iyer, G., Kim, S., Liu, Y., Lynch, C., McJeon, H., Smith, S., Waldhoff, S., Wise, M., Clarke, L., 2019. gcamdata: An R Package for Preparation, Synthesis, and Tracking of Input Data for the GCAM Integrated Human-Earth Systems Model. Journal of Open Research Software 7. https://doi.org/10.5334/jors.232

Chambers, R.G., Voica, D.C., 2017. "Decoupled" Farm Program Payments are Really Decoupled: The Theory. American Journal of Agricultural Economics 99, 773–782. https://doi.org/10.1093/ajae/aaw044

Edmonds, J.A., Link, R., Waldhoff, S.T., Cui, R., 2017. A global food demand model for the assessment of complex human-earth systems. Clim. Change Econ. 08, 1750012. https://doi.org/10.1142/S2010007817500129

Gregg, J.S., Smith, S.J., 2010. Global and regional potential for bioenergy from agricultural and forestry residue biomass. Mitig Adapt Strateg Glob Change 15, 241–262. https://doi.org/10.1007/s11027-010-9215-4

Kyle, G.P., Luckow, P., Calvin, K.V., Emanuel, W.R., Nathan, M., Zhou, Y., 2011. GCAM 3.0 Agriculture and Land Use: Data Sources and Methods (No. PNNL-21025). Pacific Northwest National Lab. (PNNL), Richland, WA (United States). https://doi.org/10.2172/1036082

Wise, M., Calvin, K., Kyle, G.P., Luckow, P., Edmonds, J., 2014. ECONOMIC AND PHYSICAL MODELING OF LAND USE IN GCAM 3.0 AND AN APPLICATION TO AGRICULTURAL PRODUCTIVITY, LAND, AND TERRESTRIAL CARBON. Climate Change Economics 5, 1–22.

Wolf, J., Chen, M., Asrar, G.R., 2021. Global Rangeland Primary Production and Its Consumption by Livestock in 2000–2010. Remote Sensing 13, 3430. https://doi.org/10.3390/rs13173430

Zhao, X., Calvin, K.V., Wise, M.A., Iyer, G., 2021. The role of global agricultural market integration in multiregional economic modeling: Using hindcast experiments to validate an Armington model. Economic Analysis and Policy 72, 1–17. https://doi.org/10.1016/j.eap.2021.07.007

Zhao, X., Wise, M.A., Waldhoff, S.T., Kyle, G.P., Huster, J.E., Ramig, C.W., Rafelski, L.E., Patel, P.L., Calvin, K.V., 2022. The impact of agricultural trade approaches on global economic modeling. Global Environmental Change 73, 102413. https://doi.org/10.1016/j.gloenvcha.2021.102413

Zhu, P., Burney, J., Chang, J., Jin, Z., Mueller, N.D., Xin, Q., Xu, J., Yu, L., Makowski, D., Ciais, P., 2022. Warming reduces global agricultural production by decreasing cropping frequency and yields. Nat. Clim. Chang. 1–8. https://doi.org/10.1038/s41558-022-01492-5