Supplementary material for "New data for representing irrigated agriculture in economy-wide models"

Download zip file of materials

Contents

1. Organization

Table S.1 lists important files included as supplementary materials. Additionally, the parent directory contains Sets.gms and Sets.gdx with set and mapping definitions.

Table S.1. Important supplementary resources

Folder	Contents	File Name
1_SpatialData	Spatial data for crop areas, yields, output prices, and production value	
	Shapefiles for aggregation to water regions, countries, or GTAP regions	
	Code **(A)** and datasets (including **B**) to calculate prices for MIRCA2000 crops	**A. Prices.gms** **B. Prices\CropPrice Interpolation.xlsx**
	Code to regenerate price, output, and value spatial data	**GenerateRasters.py**
2_ProductionValue	Code to aggregate area, output, and value spatial data to desired regions (**A**) and resulting summary files **(B)**	**A. Aggregate.py** **B. Compiledarea_GTAP.csv, etc.**
	Code to create alternative value estimates by GTAP region and crop	**GTAPreporting.gms**
	Summary file with production values and evaluation metrics by GTAP region and crop	**ValueSummary.xlsx**
	Summary file with disaggregated crop prices and production shares by MIRCA2000 crop and GTAP region	**CropDisaggregation.xlsx**
3_SupplyCurves	Data for water region supply step functions	**WaterRegionSupply.xlsx**
	Code to perform water region cluster analysis	**ClusterData.gms**, **ClusterAnalysis.do**
	Scatterplots illustrating current cluster results	**Graphs**
	Code to create IRU irrigable land supply curves	**SupplyCurves.gms**

Notes: All raster datasets are at a 5 arcminute by 5 arcminute spatial resolution. Price spatial datasets are based on prices originally available at the country level (FAO, 2015c).

2. Running the code

This section provides step-by-step instructions to (1) calculate production value by crop and land type and (2) generate IRU irrigable land supply **curves. Copies of output files are already provided. All source data is include**d in the supplementary materials folder, but refer to Section 3 for a file naming key and instructions on replicating the download, if needed.

Note: Files were generated with Python 2.7, ArcGIS 10.5, Stata 11, and GAMS 24.3.3. These instructions assume the user has compatible installations of this software.

2.1 Generating production values

1) **Run Sets.gms** to create Sets.gdx with set definitions.
2) **Run GenerateRasters.py** in the 1_SpatialData folder. To first update the country-level price data used to create the price raster datasets, refer to Section 3.
3) **Run Aggregate.py** in the 2_ProductionValue folder to create .csv files with area, output, and value at the regional level by irrigated and rainfed crop. To change the aggregation region, update the "User-Specified Parameters" section at the top of the code file.
4) **Run GTAPreporting.gms** in the 2_ProductionValue folder to map the results to the GTAP regions and crops and to alternatively estimate values from output share.
5) **The final outputs are GTAPreporting.gdx**, **ValueSummary.xlsx**, and **CropDisaggregation.xlsx** in the 2_ProductionValue folder.[1]

2.2 Generating irrigable land supply curves

1) **Run ClusterData.gms** to calculate yield by water region. Update ClusterData.dta as explained in Section 3.6.
2) **Run ClusterAnalysis.do** in the 3_SupplyCurves folder to perform k-means clustering on the water regions based on their rainfed and irrigated yields. Irrigated response unit (IRU) clusters within an EPPA region are saved in ClusterData_Results.dta and are indicated by numbers 1 up to 4 in the

[1] The price140 parameter in the GDX file reports the "effective average price", i.e. the crop value divided by crop output within the GTAP region. This price may differ from the raster file price because of small misalignments of price and output spatial files along regional borders.

clus_id column.[2] A .txt file of water region/cluster/EPPA region mapping assignments is generated 3_SupplyCurves folder. Scatter plots illustrating results are created and saved in the \Graphs subfolder, which also includes an abbreviation key.

3) **Update and run Sets.gms** in the main folder. Specifically, update the 'mapwctr' parameter with the cluster assignments to map each water region to the correct cluster (b1 to b4 depending on the cluster analysis output).
4) **Run SupplyCurves.gms** in the 3_SupplyCurves folder to generate IRU (cluster) irrigable land supply curves. This code takes WaterRegionSupply.xlsx as an input.
5) **The final output is SupplyCurves.gdx** in the 3_SupplyCurves folder. The 'lam' parameter lists supply elasticities for each IRU, where quantity_1000 ha = (price_1 USD) ^ lam.

[2] K-means clustering begins with random starting clusters, so cluster assignments may change from run to run. The Congo water region in AFR region of EPPA and the Papau_Oceania water region in the ANZ region of EPPA are manually defined as their own clusters.

3. Downloading and updating datasets

This section explains how to replicate the download of the provided datasets. Follow the instructions below to ensure compatibility with code files.

3.1 Harvested areas

1) Download the "Annual Area Harvested" data at a 5-min resolution for each of the 26 crops and 2 land types (irrigated and rainfed as GeoTiff (.tif) files.
2) Rename each file according to the naming guide in **Table S.2** and save in 1_SpatialData\HarvestedAreas.

3.2 Yields

1) Download the 58 files from year 2000 for crops 1 through 58.
2) Extract each file, rename according to the naming guide in Table S.2, and save in 1_SpatialData\Yields.

3.3 FAO prices

1) Navigate to the "Producer Prices - Annual" dataset in FAOSTAT.
2) Set the search parameters to include all food items and countries, the year 2000, and the element "Producer Price (USD/tonne)."
3) Download and save the file as FAOPrices.xlsx in 1_SpatialData\Prices.

3.4 FAO production

1) Navigate to the "Crops" dataset in FAOSTAT.
2) Set the search parameters to include all food items and countries, the year 2000, and the element "Production Quantity."
3) Download and save the file as FAOProduction.xlsx in 1_SpatialData\Prices.

3.5 Price spatial data

The price raster datasets are created from a country shapefile containing price data for the 26 crops. The steps below explain how to recreate the price shapefile. In the 1_SpatialData\Prices folder…

1) Run Prices.gms to create Prices.gdx. The GAMS script takes FAOproduction.xlsx and FAOprices.xlsx as inputs.
2) Copy the 'weightedPrice' parameter data from Prices.gdx to the *OriginalData* tab in CropPriceInterpolation.xlsx.
3) Review the instructions in the *README* tab of CropPriceInterpolation.xlsx, and update each crop tab as needed.
4) Save the updated *Summary* tab of CropPriceInterpolation.xlsx as countryprices.csv.

5) In an ArcMap project (refer to the Prices.mxd file), add the EPPA_Countries.shp file from the 1_SpatialData\SHP\EPPA folder and the countryprices.csv file.
6) Join the .csv file to the country shape file based on country name.
7) Export the country shapefile as Country_Prices.shp. This updated shapefile is used by GenerateRasters.py to convert the country-level data to the grid cell level.

3.6 ClusterData.dta

1) Run ClusterData.gms in the 3_SupplyCurves folder to update ClusterData.gdx.
2) Copy the yieldWaterReg parameter data into ClusterData.dta. Adjust the layout so that water regions and EPPA regions are row headers, and land types are column headers.
3) Save over the existing ClusterData.dta.

Table S.2. File naming guide

Crop	Harvested Area		Yield[a]	
	Irrigated	*Rainfed*	*Irrigated*	*Rainfed*
Wheat	area1.tif	area30.tif	yield1.ASC	yield30.ASC
Maize	area2.tif	area31.tif	yield2.ASC	yield31.ASC
Rice	area3.tif	area32.tif	yield3.ASC	yield32.ASC
Barley	area4.tif	area33.tif	yield4.ASC	yield33.ASC
Rye	area5.tif	area34.tif	yield5.ASC	yield34.ASC
Millet	area6.tif	area35.tif	yield6.ASC	yield35.ASC
Sorghum	area7.tif	area36.tif	yield7.ASC	yield36.ASC
Soybeans	area8.tif	area37.tif	yield8.ASC	yield37.ASC
Sunflower	area9.tif	area38.tif	yield9.ASC	yield38.ASC
Potatoes	area10.tif	area39.tif	yield10.ASC	yield39.ASC
Cassava	area11.tif	area40.tif	yield11.ASC	yield40.ASC
Sugarcane	area12.tif	area41.tif	yield12.ASC	yield41.ASC
Sugar beet	area13.tif	area42.tif	yield13.ASC	yield42.ASC
Oil palm	area14.tif	area43.tif	yield14.ASC	yield43.ASC
Rapeseed/canola	area15.tif	area44.tif	yield15.ASC	yield44.ASC
Groundnuts/Peanuts	area16.tif	area45.tif	yield16.ASC	yield45.ASC
Pulses	area17.tif	area46.tif	yield17.ASC	yield46.ASC
Citrus	area18.tif	area47.tif	yield18.ASC	yield47.ASC
Date palm	area19.tif	area48.tif	yield19.ASC	yield48.ASC
Grapes/vine	area20.tif	area49.tif	yield20.ASC	yield49.ASC
Cotton	area21.tif	area50.tif	yield21.ASC	yield50.ASC
Cocoa	area22.tif	area51.tif	yield22.ASC	yield51.ASC
Coffee	area23.tif	area52.tif	yield23.ASC	yield52.ASC
Others perennial	area24.tif	area53.tif	yield24.ASC	yield53.ASC
Others annual	area26.tif	area55.tif	yield26.ASC	yield55.ASC
Fodder grasses[b]	areaIrrigFodder.tif	areaRainFodder.tif	--	--
Maize, forage	--	--	yield27.ASC	yield27.ASC
Rye, forage	--	--	yield28.ASC	yield28.ASC
Sorghum, forage	--	--	yield29.ASC	yield29.ASC

Notes: [a] The "managed grassland/pasture" yield file (irrigated crop 25 and rainfed crop 54) does not need to be downloaded. [b] The harvested area dataset has one fodder crop while the yield dataset has three fodder crops.

4. Key to water region supply data

The WaterRegionSupply.xlsx file in 3_SupplyCurves contains the supply step function data. The workbook includes three sheets:

- **Price** – the cost in USD/ha of an upgrade
- **Quantity** – the additional irrigated land in ha gained from an upgrade
- **Technology** – the upgrade adopted

On all sheets, the rows list the 126 water regions. Columns B through P represent up to 15 possible upgrades (5 irrigation and 10 storage); however, in no basin are all 15 upgrades adopted. For each river basin, the column where the price becomes 99999 USD/ha and the quantity becomes 0 ha indicates that expansion possibilities have been exhausted.

On the technology tab, the values 1 through 15 indicate the type of each upgrade. Many basins begin with some existing storage or irrigation. Technologies that fall below existing capacity are tacked on at the end of the row, beginning at the point where the expansion limit is reached. These extra technologies and their ordering should be disregarded. **Table S.3** supplies a key to the technology numbering. **Table S.4** defines the water region naming scheme.

Table S.3. Technology key for irrigation and storage

Upgrade	Technology Number
Irrigation Upgrades	
Canal Lining	1
Flood	2
Furrow	3
Low-efficiency sprinkler	4
High-efficiency sprinkler	5
Storage Upgrades	
Storage #1	6
Storage #2	7
Storage #3	8
Storage #4	9
Storage #5	10
Storage #6	11
Storage #7	12
Storage #8	13
Storage #9	14
Storage #10	15

Table S.4. Water region definitions

Short Name	Full Name	Short Name	Full Name	Short Name	Full Name
AMA	Amazon	IWA	Iberia_West_Atla	PHI	Philippines
AMD	Amudarja	IEC	India_East_Coast	ROW	ROW
AMR	Amur	INE	Indonesia_East	RWI	Red_Winnipeg
ARA	Arabian_Peninsul	INW	Indonesia_West	RHI	Rhine
ARK	Arkansas	IND	Indus	RHO	Rhone
BAL	Baltic	IRE	Ireland	RIC	Rio_Colorado
BLA	Black_Sea	ITA	Italy	RIG	Rio_Grande
BOR	Borneo	JAP	Japan	SEA	SE_Asia_Coast
BRT	Brahmaputra	KAL	Kalahari	SAH	Sahara
BRR	Brahmari	KRI	Krishna	SAY	Sahyada
BRI	Britain	LBA	Lake_Balkhash	SAL	Salada_Tierra
CAL	California	LCB	Lake_Chad_Basin	SAN	San_Francisco
CAN	Canada_Arctic_At	LAJ	Langcang_Jiang	SCA	Scandinavia
CAR	Carribean	LIM	Limpopo	SEI	Seine
CAV	Cauvery	LBO	Loire_Bordeaux	SEN	Senegal
CAF	Central_African_	LMO	Lower_Mongolia	SON	Songhua
CAM	Central_America	LUN	Luni	SAC	South_African_Co
CAU	Central_Australi	MAD	Madagascar	SKP	South_Korea_Peni
CCA	Central_Canada_S	MAT	Mahi_Tapti	SAF	Southeast_Africa
CHJ	Chang_Jiang	MEK	Mekong	SEU	Southeast_US
CHC	Chile_Coast	MIM	Middle_Mexico	SRL	Sri_Lanka
CHO	Chotanagpui	MIS	Mississippi	SYD	Syrdarja
COL	Colorado	MOU	Missouri	TMM	Thai_Myan_Malay
COB	Columbia	MAU	Murray_Australia	TIE	Tierra
CON	Congo	NZE	New_Zealand	TIG	Tigris_Euphrates
CUB	Cuba	NIG	Niger	TOC	Toc
DAN	Danube	NIL	Nile	USN	US_Northeast
DNI	Dnieper	NAC	North_African_Co	UME	Upper_Mexico
EAC	East_African_Coa	NER	North_Euro_Russi	UMO	Upper_Mongolia
EGH	Easten_Ghats	NKP	North_Korea_Peni	URA	Ural
EAU	Eastern_Australi	NSA	North_South_Amer	URU	Uruguay
EME	Eastern_Med	NEB	Northeast_Brazil	VOG	Volga
ELB	Elbe	NWA	Northwest_Africa	VOT	Volta
GAN	Ganges	NWS	Northwest_South_	WAC	West_African_Coa

GOD	Godavari	OB	Ob	WAI	Western_Asia_Ira
GBA	Great_Basin	ODE	Oder	WAU	Western_Australi
GLA	Great_Lakes	OHI	Ohio	WGM	Western_Gulf_Mex
HAI	Hail_He	ORA	Orange	YEN	Yenisey
HOA	Horn_of_Africa	ORI	Orinoco	YHE	Yili_He
HUL	Hual_He	PAO	Papau_Oceania	YUC	Yucatan
HUN	Huang_He	PAR	Parana	ZAM	Zambezi
IEM	Iberia_East_Med	PEC	Peru_Coastal	ZHJ	Zhu_Jiang

Notes: Definitions also available in 1_SpatialData\SHP\RiverBasins.dbf.