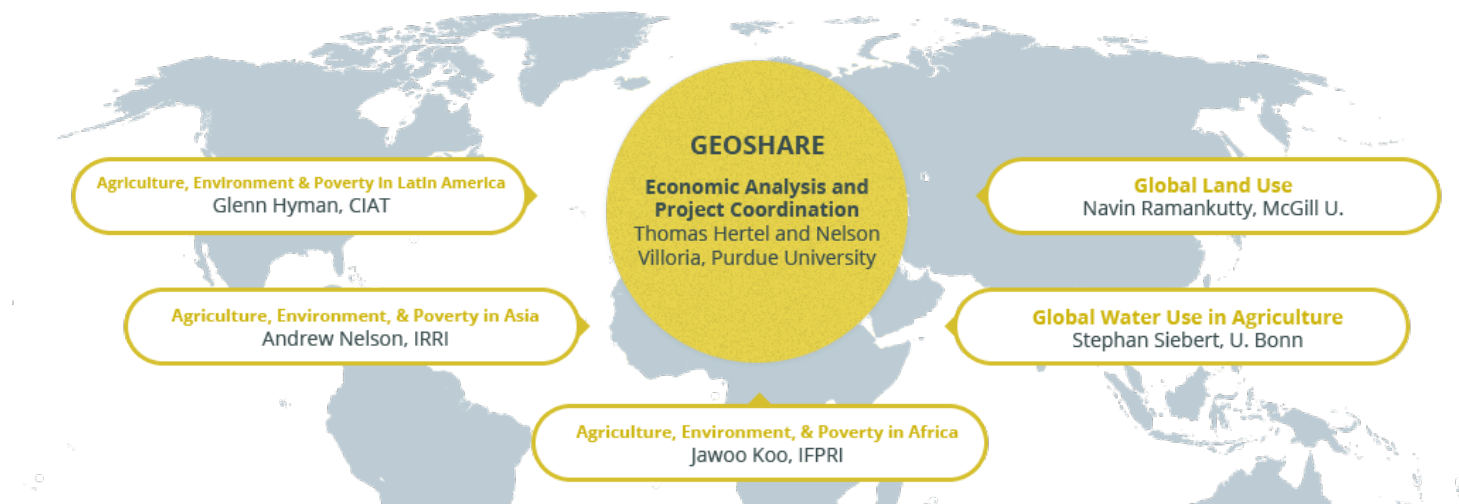


# AgMIP Data Aggregator: Leveraging Globus Online and HUBzero To Make Global Spatial Data Accessible

Lan Zhao, Carol X. Song, Nelson B. Villoria,  
Hongjun Choi, Steve M. Clark  
Purdue University

# What is GeoShare?

- **Geospatial Open Source Hosting of Agriculture, Resource & Environmental Data for Discovery and Decision Making**
- **Develop and maintain a freely available, global, spatially explicit database on agriculture, natural resources, and the environment accompanied by analysis tools and training programs for **scientists, decision makers, and development practitioners****



# The AgMIP Data Initiative



- **Agricultural Modeling Intercomparison and Improvement Project**
- To improve models used for analysis of climate impacts on agriculture
- Fast-track multi-model multi-sector climate impact assessment (part of ISI-MIP/AgMIP):
  - 7 Crop Models
  - 5 Global General Circulation Models
  - 5 Representative Climate Pathways (RCPs)
  - CO<sub>2</sub> Fertilization and irrigation
  - 12 Crops
  - All the combinations ran from 1971 to 2099
- + 36,000 0.5<sup>0</sup>x0.5<sup>0</sup> grids with crop yields
- `isimipgo#us-archive`

# The Challenges

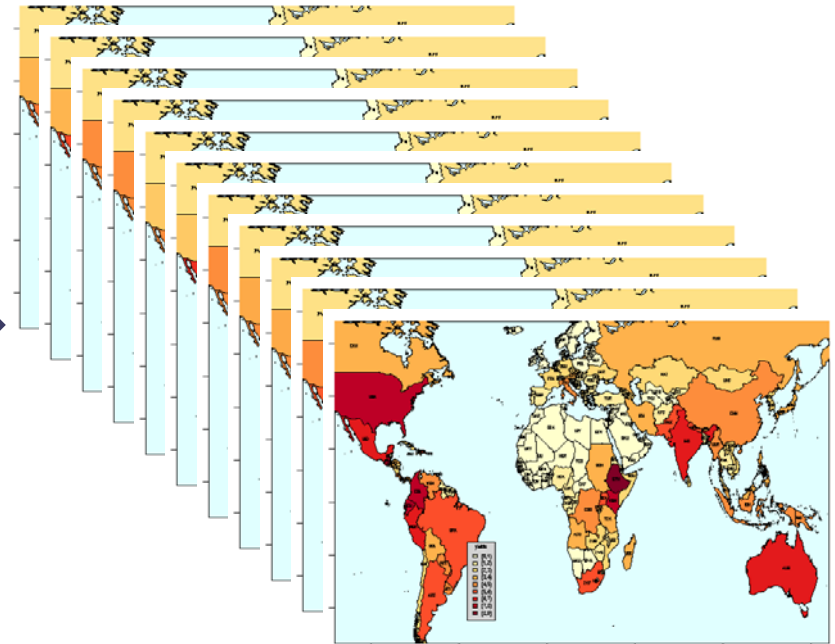
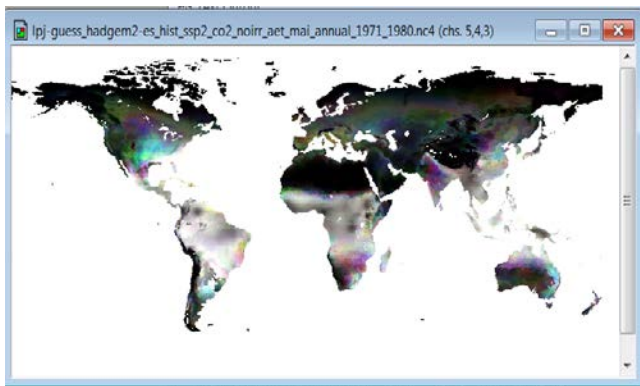
The screenshot displays a file explorer interface with the following components:

- Endpoint:** isir
- Path:** /~/k
- Folder List (Left Sidebar):**
  - NOT\_APPL
  - barley
  - maize** (highlighted in red)
  - managed\_g
  - millet
  - oat
  - others
  - rapeseed
  - rice
  - rye
  - sorghum
  - soy
  - sugar\_beet
  - sugarcane
  - wheat
- Main Content Area:** A table listing files with columns for file name and size.

File Name	Size
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_aet_mai_annual_2005_2035.nc4	3.58 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_aet_mai_annual_2035_2065.nc4	3.56 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_aet_mai_annual_2069_2099.nc4	3.56 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_biom_mai_annual_2005_2035.nc4	3.6 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_biom_mai_annual_2035_2065.nc4	3.58 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_biom_mai_annual_2069_2099.nc4	3.59 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_initr_mai_annual_2005_2035.nc4	1.95 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_initr_mai_annual_2035_2065.nc4	1.89 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_initr_mai_annual_2069_2099.nc4	1.69 MB
epic_hadgem2-es_rcp8p5_ssp2_co2_firr_pirrww_mai_annual_2005_2035.nc4	1.7 MB
epic_hadgem2-	1.7 MB
- Right-Hand Pane:** A vertical list of folders, with the top few labeled "Folder".

# The Challenges

lpj-guess\_hadgem2-es\_hist\_ssp2\_co2\_noirr\_aet\_mai\_annual\_1971\_1980.nc4



Total Number of Pixels = 259,200  
10 Channels  
360 lines, 720 columns

10 ascii files aggregated to country level

# The Challenges

- **Data discovery**
  - **Difficult to navigate and locate data**
- **Data download**
  - **Globus Connect**
- **Data processing**
  - **Netcdf4 format, 5-10 year slices**
- **Data storage**

# AgMIP Data Aggregator



- Integrate Globus Online into HUBzero

- Globus connect end point on the web server
- Service wrapper to globus client commands

```
$ submit isimiptransfer --help
```

```
Usage: isimiptransfer.py [huboptions]
```

Options:

```
-h, --help          show this help message and exit
```

```
-l DIRREQUESTS, --ls=DIRREQUESTS
```

```
--lsroot
```

```
-g GETREQUESTS, --get=GETREQUESTS
```

- Can be extended to connect to other end points

# AgMIP Data Aggregator

The screenshot shows the AgMIP Tool 1.2 web interface. The browser address bar displays <http://www.pnas.org/content/111/9.cover-expansion>. The interface has three tabs: "AgMIP Download", "Aggregation" (selected), and "Visualization".

The "Aggregation" tab is titled "Crop Data" and contains several selection panels:

- Model:** EPIC, GEPIG, pDSSAT, LPJmL, IMAGE\_LEITAP, PEGASUS, LPJ-GUESS (selected).
- GCM:** HadGEM2-ES (selected), IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M, NorESM1-M.
- RCP:** hist (selected), rcp8p5, rcp6p0, rcp4p5, rcp2p6.
- SSP:** ssp2 (selected).
- CO2:** co2 (selected), noco2.
- IRR:** noirr (selected), firr.
- Crop:** maize (selected), soy, wheat, rice, managed\_grass, others, rapeseed, barley, millet, sorghum, sugarcane, sugar\_beet.

A "Download" button is present, followed by a progress bar showing 100% completion.

The "Log" window at the bottom shows the following text:

```
running command: submit isimiptransfer --ls upload_area/LPJ-GUESS/HadGEM2-ES/hist/ssp2/co2/noirr/maize
{
  "upload_area/LPJ-GUESS/HadGEM2-ES/hist/ssp2/co2/noirr/maize": {
    "directories": [],
    "files": {
```



# AgMIP Data Aggregator

**AgMIP Tool 1.2** Help

AgMIP <http://www.pnas.org/content/111/9.cover-expansion>

AgMIP Download **Aggregation** Visualization

**Input Files**

1. AgMIP Files:

2. Region Map:

3. Weight Map or Aggregation Function

Weight Map:

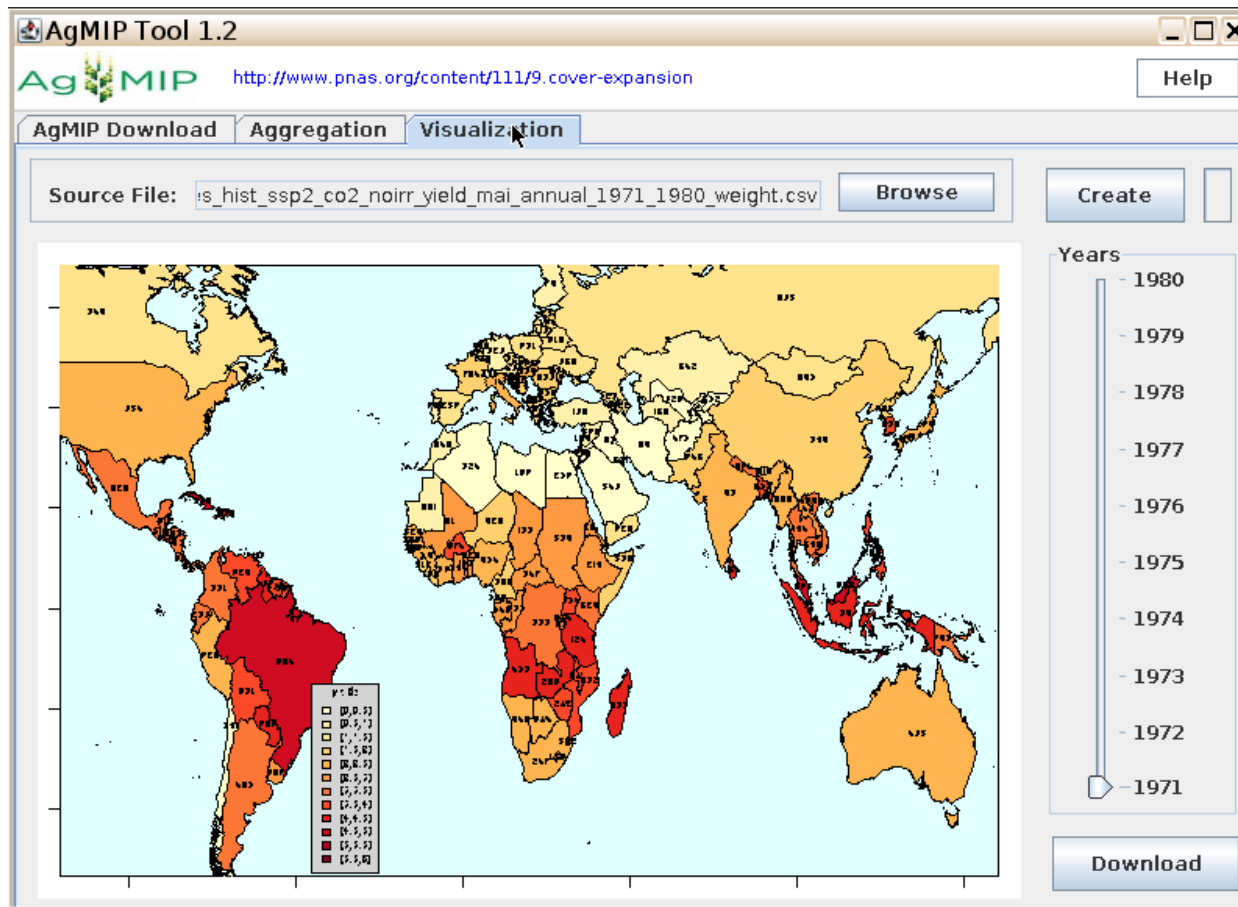
**Aggregation Functions**

Enable Functions     MAX     MIN     MEAN     SD

**Log**

```
Starting Processing...
Processing Input... [1/1]
=====
AgMip File: /data/transfers/gotransfer/upload_area/LPJ-GUESS/HadGEM2-ES/hist/ssp2/co2/noirr/maize/lpj-gue
Region File: /apps/agmip/r277/examples/regionmap/WorldId.csv
Function: null
Weight File: /apps/agmip/r277/examples/weightmap/maize_yield30r.csv
```

# AgMIP Data Aggregator



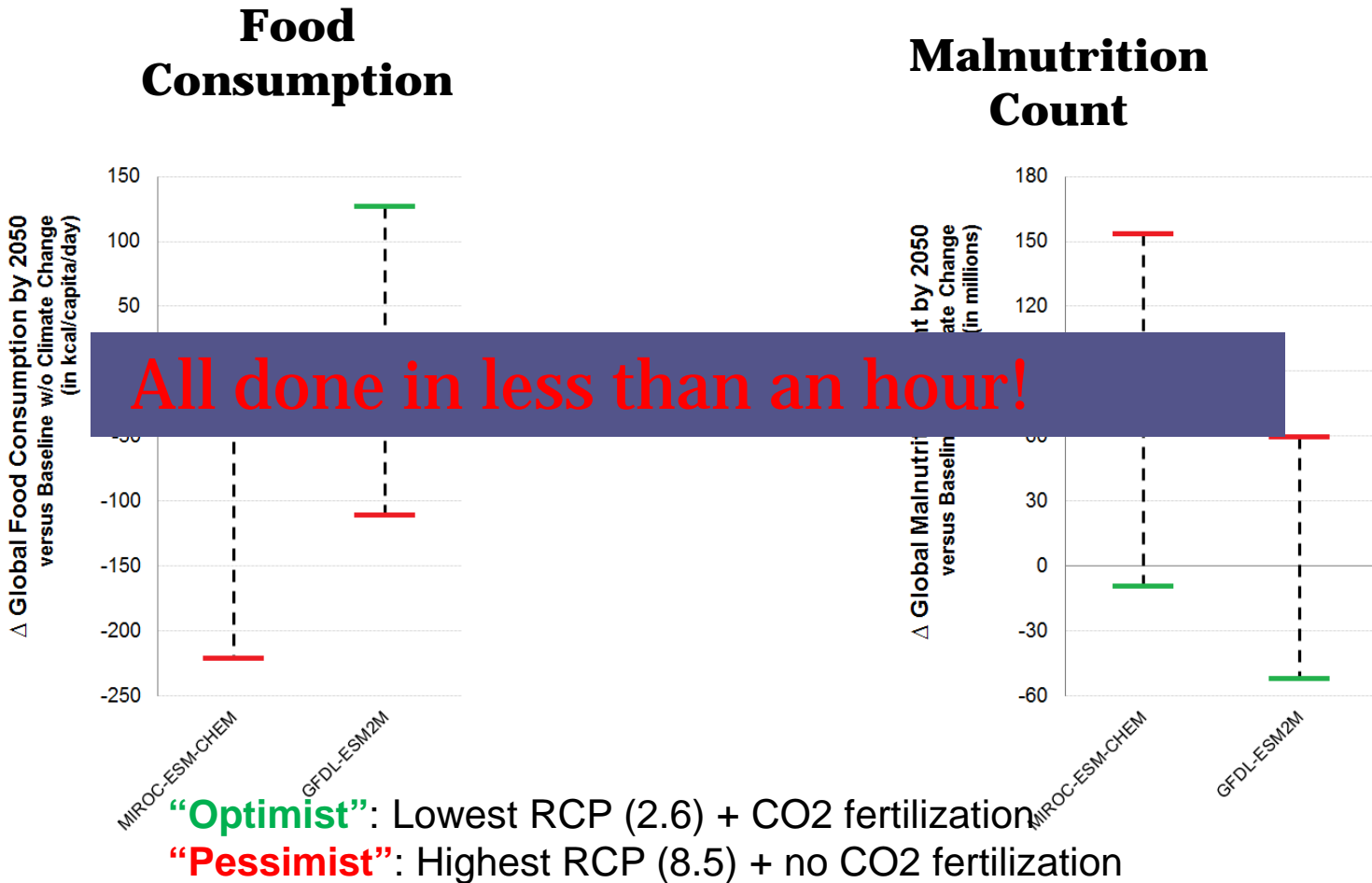
# AgMIP Data Aggregator

- Easy to use user interface to select data to download
- Integrate with HUBzero GO service at the backend
- R script to aggregate the AgMIP output to any user defined levels using different functions
- Map Visualization
- Metadata

# Use Case

- **Climate change impacts on food security in 2050 by Uris Baldos**
  - **Climate impact on yield scenarios:**
    - Crop model: PEGASUS
    - GCMs: MIROC-ESM-CHEM, GFDL-ESM2M
    - No CO<sub>2</sub> fert./RCP8p5 pessimist vs. CO<sub>2</sub> fert./RCP2p6 optimist
  - **SIMPLE: an economic model of agriculture**
  - **Change in crop prices, caloric consumption, caloric malnutrition headcount in 2050**

# Use Case



# Future Work

- **Work with the GABBS project to develop geospatial building blocks**
  - **GlobusOnline data transfer**
  - **Geospatial data aggregation**
  - **Publish/share model output**
  - **Geospatial data driven modeling workflows**

- **GeoShare hub:**  
<https://geoshareprojects.org>
- **AgMIP tool:**  
<https://mygeohub.org/resources/agmip>
- **AgMIP:** <http://www.agmip.org/>