
DELIVERING CLIMATE DATA TO A BROAD USER COMMUNITY – A GEOSHARE PROJECT

Lan Zhao

Rosen Center for Advanced Computing

Purdue University

PURDUE
UNIVERSITY

THE TEAM

From Purdue University

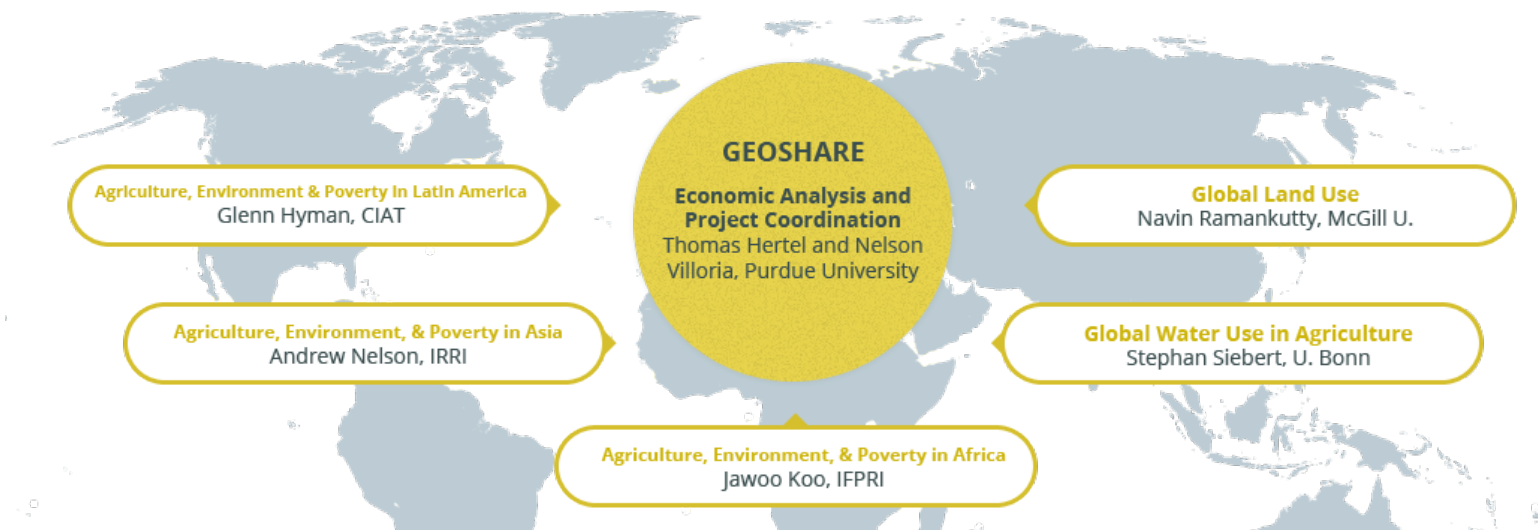
- Jaewoo Shin
- Carol X. Song
- Steve M. Clark
- Lan Zhao

From Kansas State University

- Nelson Villoria

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**
- GeoShare hub: <https://geoshareproject.org>



THE CMIP5 DATA INITIATIVE

- CMIP - Coupled Model Intercomparison Project:
 - Established by the Working Group on Coupled Modelling (WGCM) under World Climate Research Programme (WCRP)
 - CMIP provides a community-based infrastructure in support of climate model diagnosis, validation, intercomparison, documentation and data access.
 - enables a diverse community of scientists to analyze GCMs in a systematic fashion
 - Virtually the entire international climate modeling community has participated in this project since its inception in 1995.
 - Current Intercomparison - CMIP5, provided much of the new material underlying IPCC Fifth Assessment Report (AR5).
- Large amount of data. <http://cmip-pcmdi.llnl.gov/cmip5/availability.html>

AN OPEN LETTER TO YOU CLIMATE PEOPLE

“Dear Climate People (yes, I mean you IPCC WG1 types):

.. An economist to be precise. I am the type of person who is greatly interested in projecting impacts of climate change on human and natural systems. We use fancy statistics, spend tons of time and effort collecting good data on observed weather/climate and outcomes of interest....

But you know what's a huge pain in the neck for all of us? Trying to get climate model output in a format that is useable by someone without at least a computer science undergraduate degree. While you make a big deal out of having all of your climate model output in a public repository, we do not have the skills to read your terabytes and terabytes of netCDF files into our Macbooks and put them in a format we can use.”

By **Maximilian Auffhammer** (Professor of International Sustainable Development, UC-Berkley)

<http://www.g-feed.com/2014/09/an-open-letter-to-you-climate-people.html>

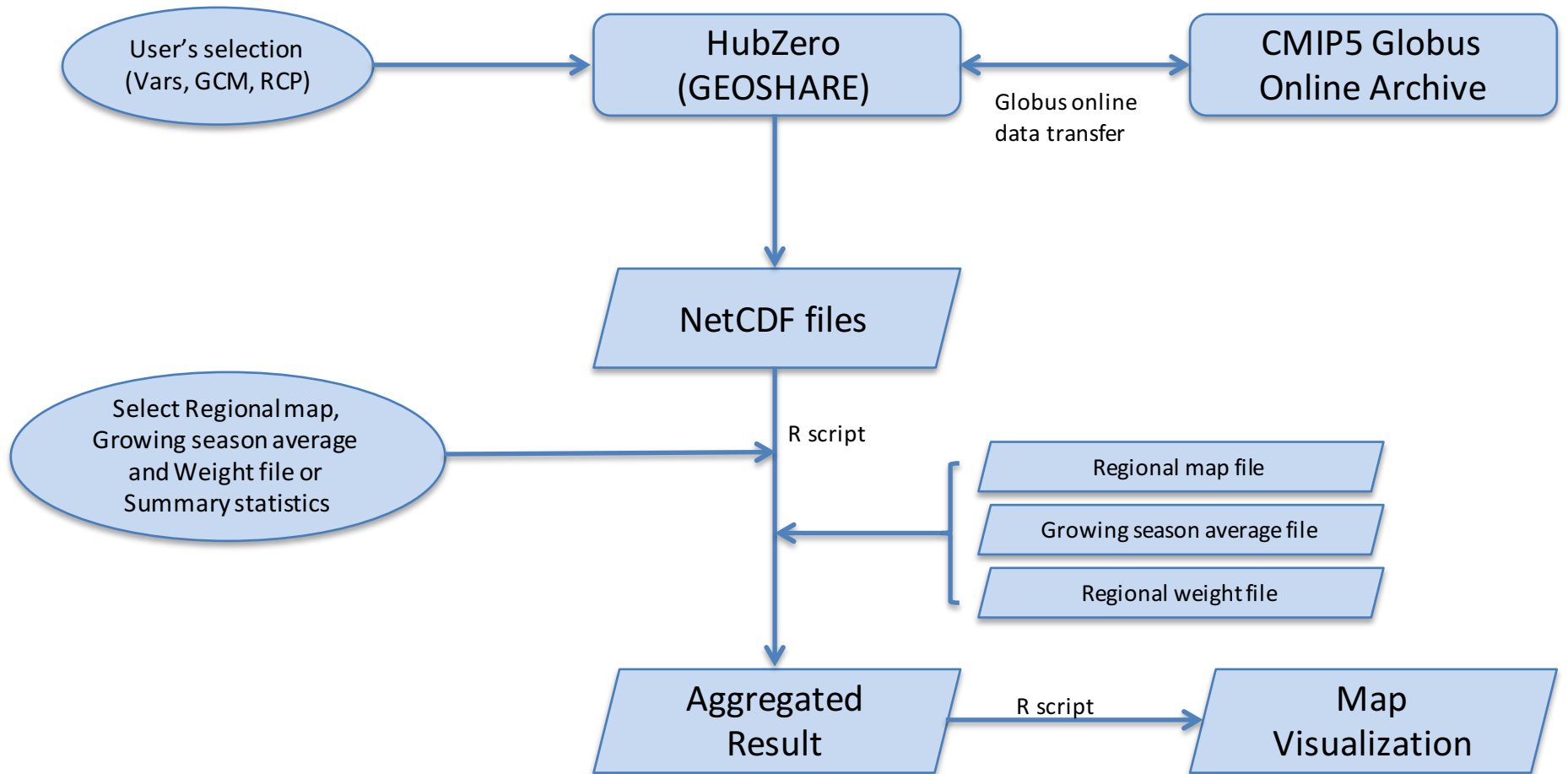
DATA ACCESS CHALLENGE

- Large data (more than 2 million GB of data).
- Requires specialized skills and computing resources
- A big hurdle for experts on climate modeling as well
- Stocker and Plattner (2014) report survey results coming from the Working Group 1 of the IPCC tasked with assessing the physical science of the world's climate system reveals
 - “more than 60% (of respondents) encountered hurdles when processing big data associated with the analysis of model simulations for climate projections and had trouble gaining timely access to model results.”

CMIP5 DATA AGGREGATOR

- A HUBzero tool on GEOSHARE
- Enable users to download growing season measures of climate variables for user-defined spatial aggregation schemes
- Targeted for economists working with econometric and simulation models
- Main features:
 - Easy to use user interface to select data to download
 - Integrate with HUBzero Globus Online service at the backend
 - R script to aggregate the CMIP5 output to any user defined levels using different functions
 - Map Visualization
 - Metadata

SYSTEM WORKFLOW



CMIP5 AGGREGATOR TOOL COVERAGE

Global Climate Models	Representative Concentration Pathways	Data on growing season available for the following crops
HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M, NorESM1-M.	Historic, RCP8.5, RCP 6.0, RCP4.5, RCP2.6.	Barley (winter, spring), cassava, cotton, groundnuts, maize, millet, oats (winter, spring), potatoes, pulses, rapeseed-winter, rice, rye-winter, sorghum, soybeans, sugarbeets, sunflower, sweet potatoes, wheat (winter, spring), and yams,

CMIP5 DATA SOURCE

- Surface air temperature (average, minimum, and maximum), precipitation
- Served from ISI-MIP globus online archive
- Processed from the original CMIP5 data into monthly means arranged in $0.5^{\circ} \times 0.5^{\circ}$ global grids
- NetCDF4 format, in 5-10 year slices
- covering historical and future climates for the period 1960-2099

GLOBUS ONLINE DATA TRANSFER

- A fast, reliable, and secure file transfer service developed by the Globus project
- Integrate Globus Online into HUBzero
 - Globus connect end point on the web server
 - Service wrapper to globus client commands

```
$ submit isimptransfer --help
```

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

Options:

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

```
-I DIRREQUESTS, --ls=DIRREQUESTS
```

```
--lsroot
```

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

- Can be extended to connect to other end points



SPATIAL AND TEMPORAL AGGREGATION

- Flexible aggregation functions – allow arbitrary spatial aggregation based on maps that relate each grid-cell centroid to an aggregated region
- Growing season based temporal aggregation
 - Reduce the summary statistics to include only the months in the growing season

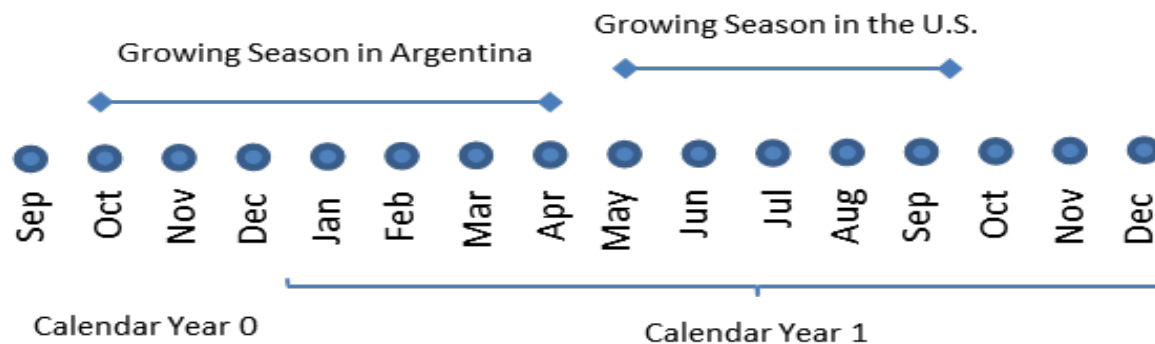


Figure 4: The average temperature/precipitation over the growing season is assigned to the calendar year in which the harvest season occurs. In the example, for Argentina, the average temperature/precipitation in Calendar Year 1 is taken over October '0-April '1 while in the U.S. (Midwest region) is taken over May '1-Sep '1. The planting and harvesting dates for each country are from Sacks et al.

DATA AGGREGATION

**tasmax_bced_1960_1999_ipsl
-cm5a-lr_rcp6p0_2031-
2040.mm.nc**

Monthly maximum surface air
temperature

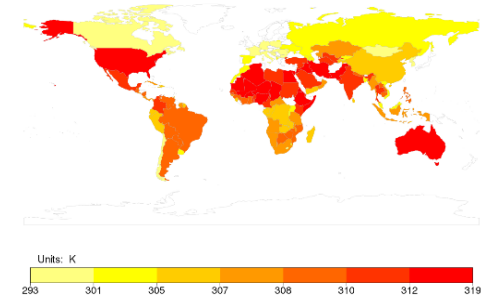
Model output climate of IPSL-
CM5A-LR r1i1p1 Interpolated to
0.5 degree and bias corrected
using observations from 1960 -
1999 for EU WAT

CH project

Total Number of Pixels = 259,200
120 Channels
360 lines, 720 columns
Netcdf4 format, 5-10 year slices



Maximum Surface Air Temperature, 2033
GCM: IPSL-CM5A-LR, RCP: 6.0
Maximum Surface Air Temperature, 2033
GCM: IPSL-CM5A-LR, RCP: 6.0
Maximum Surface Air Temperature, 2033
GCM: IPSL-CM5A-LR, RCP: 6.0
Maximum Surface Air Temperature, 2033
GCM: IPSL-CM5A-LR, RCP: 6.0
Maximum Surface Air Temperature, 2033
GCM: IPSL-CM5A-LR, RCP: 6.0
Maximum Surface Air Temperature, 2033
GCM: IPSL-CM5A-LR, RCP: 6.0
Maximum Surface Air Temperature, 2033
GCM: IPSL-CM5A-LR, RCP: 6.0
Maximum Surface Air Temperature, 2033
GCM: IPSL-CM5A-LR, RCP: 6.0



10 ascii files aggregated to
country level

CMIP5 DATA AGGREGATOR

The screenshot shows the 'Climate Tool 1.0 @ GEOSHARE' web interface. It features a navigation menu with 'Download', 'Aggregate', and 'Visualize' tabs. The 'Download' tab is active, displaying a 'Climate Data' section with three columns of radio button options: 'Variable', 'GCM', and 'RCP'. The 'Variable' column has 'Maximum surface air temperature' selected. The 'GCM' column has 'IPSL-CM5A-LR' selected. The 'RCP' column has 'RCP6.0' selected. Below these options is a 'Fetch Data' button, a progress bar at 19%, and two buttons: 'Download NetCDF files' and 'Data Description'. At the bottom, a 'Log' window shows the following text: 'status : SUCCEEDED', '}', '}', 'transferring... : 2 / 10', and 'running command: submit isimiptransfer --get inputs/monthly.mean/IPSL-CM5A-LR/rcp6p0/tasmax_v1/tasmax_bced_1960_1999_ip'.

Climate Tool 1.0 @ GEOSHARE

Help

Download Aggregate Visualize

Climate Data

Variable

- Average surface air temperature
- Maximum surface air temperature
- Minimum surface air temperature
- Precipitation

GCM

- HadGEM2-ES
- IPSL-CM5A-LR
- MIROC-ESM-CHEM
- GFDL-ESM2M
- NorESM1-M

RCP

- Historical
- RCP8.5
- RCP6.0
- RCP4.5
- RCP2.6

Fetch Data

19%

Download NetCDF files

Data Description

Log

```
status : SUCCEEDED
}
}
transferring... : 2 / 10
running command: submit isimiptransfer --get inputs/monthly.mean/IPSL-CM5A-LR/rcp6p0/tasmax_v1/tasmax_bced_1960_1999_ip
```

CMIP5 DATA AGGREGATOR

Climate Tool 1.0 @ GEOSHARE

Help

Download Aggregate Visualize

Input Files

1. Climate Data:

2. Regional Map:

Obtain annual growing season averages (1)

3. Summary Statistics

Choose grid-cell level weights for weighting regional averages:

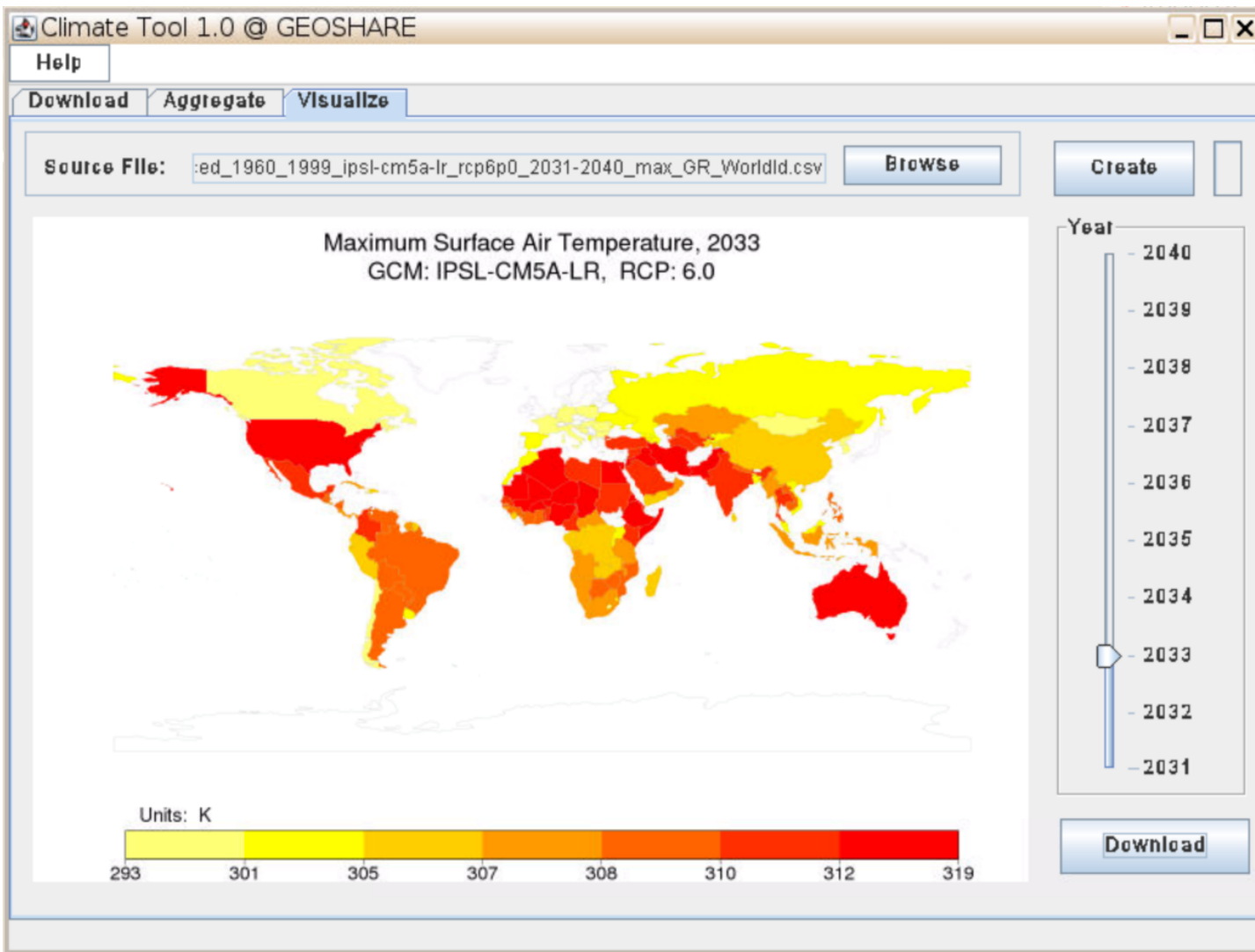
...or select a summary statistic:

Enable Functions MAX MIN MEAN SD

Log

```
Year: 2031-2040
Variable ID: tasmax
Growing File: /home/mygeohub/shin152/newTool/trunk/examples/growingseason/maizegrowing.season.unfilled.RData
=====
Running R...
Rscript /home/mygeohub/shin152/newTool/trunk/examples/do.r /home/mygeohub/shin152/newTool/trunk/examples/function
```

CMIP5 DATA AGGREGATOR



FUTURE WORK

- <https://mygeohub.org/tools/climatetool>
- 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

ACKNOWLEDGEMENT

This project was supported by the Economic Research Service of the USDA, the Agriculture and Food Research Initiative Competitive Grant no. 2015-67023-23109 from the USDA National Institute of Food and Agriculture, and by the National Science Foundation Competitive Grant no. 1261727.