Web-based access, aggregation, and visualization of future climate projections with emphasis on agricultural assessments.

Nelson B. Villoria[a,*], Joshua Elliott[b], Christoph Müller[c], Jaewoo Shin[d], Lan Zhao[e], Carol Song[e]

[a]*Department of Agricultural Economics, Kansas State University, 331i Waters Hall, Manhattan, KS,66506, USA.*
[b]*University of Chicago and ANL Computation Institute Center for Robust Decision-making in Climate and Energy Policy (RDCEP) 5735 S. Ellis Ave. Chicago, IL 60637, USA.*
[c]*Potsdam Institute for Climate Impact Research, Telegraphenberg A 31 14473, Potsdam, Germany.*
[d]*Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA.*
[e]*Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN 47907, USA.*

Abstract

Access to climate and spatial datasets by non-specialists is restricted by technical barriers involving hardware, software and data formats. We discuss an open-source online tool that facilitates downloading the climate data from the global circulation models used by the ISI-MIP project. The tool also offers temporal and spatial aggregation capabilities for incorporating future climate scenarios in applications where spatial aggregation is important. We hope that streamlined access to these data facilitates analysis of climate related issues while considering the uncertainties derived from future climate projections and temporal aggregation choices.

Keywords: Climate data, CMIP5, Climate models, HUBzero, Growing season temperature, Growing season precipitation, Climate change and agriculture

*Corresponding author. Department of Agricultural Economics, Kansas State University, 331i Waters Hall, Manhattan, KS,66506, USA. Email: nvilloria@ksu.edu.

Email address: `nvillori@ksu.edu` (Nelson B. Villoria)

1. Motivation and significance

Studies of the effects of climate change on agriculture typically involve using observational data to determine the parameters connecting climate variables to agricultural productivity and then using future climate projections from global circulation models (GCM) to evaluate potential future impacts or the effects of alternative policies [e.g., 1]. Given the uncertainty surrounding future climate projections it is considered best practice to use the output of several GCM in order to obtain a range of potential outcomes [2]. Despite increases in the availability of climate data stemming from the coordination between climate modeling groups through the Coupled Model Intercomparison Project Phase 5 (CMIP5) and their collaboration with the Intergovernmental Panel on Climate Change, access by non-specialists is hindered by technical barriers including software, hardware and the need of specialized skills to handle non-standard formats [3]. In addition to access to data, spatial processing is not trivial requiring expertise in geographic information systems (GIS) methods to process both climate data as well as auxiliary datasets [4].

The tool discussed in this article seeks to reduce the technical barriers to access climate model outputs through a web-based facility that facilitates downloading and aggregating global grids (0.5 degree) of bias-corrected, monthly mean historical and future temperature and precipitation from the five General Circulation Models (GCMs) used by the Inter-Sectoral Impacts Model Intercomparison Project [ISI-MIP; 5, 6]. (See table 2 for included models). The scientific problem the tool contributes to solve is to facilitate the analysis of future climate scenarios in applications where spatial aggregation is important. This includes a wide range of economic analysis focused on either impact assessment [7, 8, 9] or policy analysis [10].

The tool targets mainly, but not exclusively, researchers interested on the effects of climate change on agriculture. At the most general level, the Climate Scenario Aggregator (CSA) tool can be used as a downloading platform of the raw GCM data in the ISI-MIP archive. The target user of this functionality is skilled with NetCDF formats, has a relatively powerful computer, reasonable bandwidth, and is comfortable with the scripting and/or programming languages needed for manipulating and processing spatially-explicit data. A second target user may need some assistance with basic preprocessing of the data, such as temporal and spatial aggregation. This user will benefit from the aggregation programs as well as preprocessed datasets for temporal aggregation (crop calendars) and spatial aggregation (e.g., from gridcells to countries.) Finally, a third target user may be interested in the download and aggregation capabilities of the tool, but wishes to employ alternative spatial aggregation schemes (e.g., gridded population.)

The CSA tool is related to other tools that seek to facilitate access to and spatial geoprocessing of climate data while leveraging shared resources and expertise. Examples of these tools are given by [3], who developed user-friendly software applications for downscaling climate data for ecological modeling applications. Meanwhile, [11] have built an aggregation tool that facilitates access to the gridded forecast of yield changes produced by the The Agricultural Model Intercomparison and Improvement Project [AgMIP; 12].

2. Software description

The tool is available at the GEOSHARE HUBzero website (`https://mygeohub.org/tools/climatetool`) using any standard Internet browser. The CSA tool allows users to calculate for each half-degree land pixel a crop-specific growing season average value of temperature and precipitation using the global crop calendars from [13] (See table 2 for crop coverage.) The tool also permits aggregating the pixels to larger geographic units using crop harvested area and production from [14]. All the programs—a java graphical user interface (GUI) and a set of R functions— can be freely downloaded and reused. Documentation and support for users include a User's Manual[1] as well as a set of default regional maps and weighting schemes.

2.1. Software Architecture

HUBzero [15], is an open source software platform specializing in disseminating simulation and data tools via the world wide web. Originated in the nanotechnology community (`https://nanohub.org/tools/`), HUBzero has evolved to constitute a flexible environment and recent efforts have focused on developing capabilities for processing and delivering spatially explicit data using shared remote resources which have given rise to a series of user communities and shared tools[2]. Users access the CSA tool at GEOSHARE using

[1]Included as an Appendix for the reviewers convenience.

[2]Including the tools from the GeoSpatial Analysis and Building Blocks Project (GABBS, `https://mygeohub.org/groups/gabbs`): SWATShare (`https://mygeohub.org/groups/water-hub/swatshare`), MultiSpec (`https://mygeohub.org/tools/multispec`), Water Deficit Viewer (`https://mygeohub.org/tools/deficitviewer`), and the Active Learning Tool (`https://mygeohub.org/tools/act`); as well as of those from the GEOSHARE project: the AgMIP tool, which aggregates outputs from the AgMIP's Global Gridded Crop Modeling Initiative at https://mygeohub.org/tools/agmip, and FLAT, a tool for downscaling national and sub-national level statistics on harvested area available at https://mygeohub.org/tools/flat.

an ordinary Web browser without having to download or compile any code specific to the tools. The tool runs in an isolated light-weight virtual machine container and is displayed in the user's web browser using a graphical desktop sharing technology called Virtual Network Computing (VNC).

2.2. Software Functionalities

The CSA tool has four main functionalities: data download, data aggregation, output and metadata, and visualization. Download, aggregation, and visualization are implemented as tabs in the graphical user interface shown in figures 1 and 2. The climate data is stored in NetCDF files. Each file is identified by a file name with seven components that specifies a variable: temperature (minimum, maximum, average) or precipitation; a climate model: HadGEM2-ES [16], IPSL-CM5A-LR [17], MIROC-ESM-CHEM [18], GFDL-ESM2M [19], and NorESM1-M [20]; a representative concentration pathway [21]: historical, RCP2.6, RCP4.5, RCP6.0 and RCP8.5; and a time period that ranges from 1960 to 20099. For instance:

```
tas_bced_1960_1999_noresm1-m_rcp2p6_2006-2010.mm.nc
tas_bced_1960_1999_noresm1-m_rcp2p6_2011-2020.mm.nc
                          .
                          .
                          .
tas_bced_1960_1999_noresm1-m_rcp2p6_2091-2099.mm.nc
```

are global grids of monthly air surface temperature means (one grid for each year in the period 2006-2099), projected by NorESM1-M [20], under representative concentration pathway RCP2.6 [21].

In order to retrieve the data, the user selects a unique combination of variable, climate model and scenario which are all presented in the tool's user front-end (figure 1). The user's selections create a character string that matches the file names stored in the the ISI-MIP archive. This character string is used to retrieve all the available years— in most cases, each file stores information on 10 years worth of data— for the selected scenario. GEOSHARE's Hub and the ISI-MIP archive are connected through Globus Online [22], a service that facilitates transfer of large datasets.

Once in GEOSHARE's Hub, the files are stored in a common server workspace. Before each data request, the tool checks whether the data has already been downloaded, and if so, indicates this to the user. This feature avoids downloading the same data more than once. At this point, the user can either download the raw NetCDF files for custom processing on her desktop, or proceed to aggregate the data through the GUI implementation in figure 2).

Aggregation is performed by three R functions. The first function reads the data using the R NetCDF package by [23]. The second function estimates pixel and crop specific growing-season averages of the chosen climate variable. Planting and harvesting months for each pixel are from [13]. In many cases, the harvesting month is in a different year than the planting month. For example, planting of corn in most of Argentina occurs in October and the crop is harvested in April of the following year. Meanwhile, corn planting in the U.S. starts in May with crops harvested in September. In order to avoid ambiguities we assign the average value of the variables (e.g. temperature) over the growing season for the month in which the harvest occurs. So, the value of the average growing season temperature for year 2000 corresponds the the Argentinean harvest of April 2000 and the US harvest of September 2000 (see figure 3).

A third R function performs the aggregation from grid-cells to larger geographic units. The user has the opportunity to select different aggregation schemes or upload her own. For example, aggregation from the grid-cell to country level requires a mapping that correlates each latitude and longitude pair with a unique country name. The mapping schemes are simple comma separated value files. By default, we have included regional mappings from grid-cells to countries, country-AEZ regions, and global. Simple guidelines for preparing these data files are in the User's Manual, which can be retrieved from either the description page of the tool.In addition the tool allows for weighted and unweighted aggregations. Files are provided from weighted aggregations using harvested areas and production based on the gridded crop harvested area and yield statistics from [14] .

The CSA tool also keeps a record of the user's choice producing a text file that indicates the chosen combination of GCM, RCP and variables which can be obtained by clicking on "Data description" in the Download tab (figure 1). For users performing an aggregation in the Aggregation tab, the documentation includes aggregation choices as well as the source of the aggregation weights (see figure 2.)

3. Illustrative Examples

Figure 4 displays four plots that illustrate the versatility of the tool in terms of spatial and temporal aggregation of the the GCM outputs. Figure 4.A compares growing-season temperatures for wheat in a single gridcell near Manhattan, Kansas in the US. Figure 4.B displays historical and average temperatures during the growing season of maize for the US using projections for RCP 2.6 for the five GCMs included in the ISI-MIP archive. In this case, the individual gridcells have been weighted by their contribution to total US maize production using production weights. An interesting feature of Figure 4.B is that allows to understand the uncertainty embedded in the model and eventually include this uncertainty in modeling exercises or impact analyzes. The two following figures, C and D, display temperature and precipitation aggregated from individual gridcells to the global level using three different aggregation modalities: weighted averages using harvested area weights, weighted averages using production weights, and unweighted averages. These two figures exemplify the usefulness of the tool for evaluating different empirical choices of aggregation at different spatial scales.

4. Impact

Our software makes three contributions. First it provides straightforward access to an important number of models in the CMIP5 archive. Second, it provides important GIS functionality for data aggregation. Finally, all the downloading and processing is in remote servers. It is likely that these contributions have varying degrees of appeal for different users, nevertheless, by expanding access and lowering entry barriers to use, we expect that this tool advances the study of the impacts of climate change in world agriculture across several geographic scales. The potential research questions that benefit for streamlined access to climate data include statistical analysis of future climate patterns; modeling the human and ecological impacts of climate change; an the evaluation of adaptation and mitigation policies. The tool also facilitates streamlined descriptions of climate patterns at different spatial scales as well as exploring the effects of different aggregation mechanisms.

5. Limitations

An important consideration to keep in mind is that these models are a subset of the around thirty-six models that contributed to the CMIP5 data archive. These five models were selected because they were the first to supply data that met the minimum data requirements of the ISI-MIP project [5, p. 221]. It is also important to keep in mind that for many regions these models are likely to underestimate the uncertainty in future climate projections [24]. In particular, these authors find that "the fraction of the of the full range of future projections captured across different regions and seasons by the ISI-MIP subset varies from 0.5 to 0.9 for temperature (median 0.75) and 0.3 to 0.8 for precipitation (median 0.55)." This is a general problem in climate scenario selection. Even if dry, wet, cool or hot climate projections can be specifically selected for particular regions, including the global aggregation, these characteristics do not necessarily hold for other regions. As such, a climate projection that is specifically dry and hot compared to other projections in one region may be cool and wet in other regions. [24] find that at least 13 climate model projections are needed to cover a substantial range of the uncertainty in all regions. This tool cannot be easily extended to all climate projections from the CMIP5 archive, as these are not available in bias-corrected form as done by [5], but we encourage users to note the limited representation of scenario selection in the interpretation of their applications.

6. Conclusions

Access to climate and spatial datasets by non specialists is hindered by technical difficulties involving software and data formats as well as the need for strong Internet bandwidth and storage capacity. This article discusses a GEOSHARE HUBzero tool that expands access to the climate data that underlies the AgMIP Global Gridded Crop Model Intercomparison (GGCMI) Project to the broader scientific community who can benefit from these data, but who may lack the resources to gain access to them. We hope that this software tool enables researchers facing technical limitations to overcome these barriers.

Acknowledgments

We are grateful to the ISI-MIP coordination team at the Postdam Institute for Climate Impact Research (Postdam, Germany) for making the climate data discussed in this paper available as part of the ISI-MIP Fast Track project. The output of these models is in turn available to the public thanks to the efforts and coordination of the different climate modeling groups through the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for the Coupled Model Intercomparison Project. This project was supported by the Economic Research Service of the U.S. Department of Agriculture through Agreement 58-3000-1-0058; the Agriculture and Food Research Initiative Competitive Grant no. 2015-67023-25258 from the USDA National Institute of Food and Agriculture; and by the National Science Foundation Competitive Grant no. 1261727. We also thank Bowen Chen for superb research assistance during the development and testing of the tool.

[1] M. Auffhammer, S. M. Hsiang, W. Schlenker, A. Sobel, Using weather data and climate model output in economic analyses of climate change,

Review of Environmental Economics and Policy 7 (2) (2013) 181–198. doi:10.1093/reep/ret016.

[2] M. Burke, J. Dykema, D. B. Lobell, E. Miguel, S. Satyanath, Incorporating climate uncertainty into estimates of climate change impacts, Review of Economics and Statistics 97 (2) (2015) 461–471. doi:10.1162/REST_a_00478.

[3] T. Wang, A. Hamann, D. Spittlehouse, C. Carroll, Locally downscaled and spatially customizable climate data for historical and future periods for north america, PLOS ONE 11 (6) (2016) e0156720. doi:10.1371/journal.pone.0156720.

[4] S. Wang, Y. Liu, A. Padmanabhan, Open cyberGIS software for geospatial research and education in the big data era, SoftwareX-doi:10.1016/j.softx.2015.10.003.

[5] S. Hempel, K. Frieler, L. Warszawski, J. Schewe, F. Piontek, A trend-preserving bias correction – the ISI-MIP approach, Earth Syst. Dynam. 4 (2) (2013) 219–236. doi:10.5194/esd-4-219-2013.
URL http://www.earth-syst-dynam.net/4/219/2013/

[6] L. Warszawski, K. Frieler, V. Huber, F. Piontek, O. Serdeczny, J. Schewe, The Inter-Sectoral impact model intercomparison project (ISI–MIP): project framework, Proceedings of the National Academy of Sciences 111 (9) (2014) 3228–3232, PMID: 24344316. doi:10.1073/pnas.1312330110.

[7] D. Lobell, C. Field, Global scale climate–crop yield relationships and the impacts of recent warming, Environmental Research Letters 2 (1). doi:10.1088/1748-9326/2/1/014002.

[8] D. B. Lobell, W. Schlenker, J. Costa-Roberts, Climate trends and global crop production since 1980, Science (2011) 1–5doi:10.1126/science.1204531.

[9] W. Schlenker, D. B. Lobell, Robust negative impacts of climate change on african agriculture, Environmental Research Letters 5 (1) (2010) 014010. doi:10.1088/1748-9326/5/1/014010.

[10] U. L. C. Baldos, T. W. Hertel, The role of international trade in managing food security risks from climate change, Food Security (2015) 1–16doi:10.1007/s12571-015-0435-z.

[11] N. B. Villoria, J. Elliott, C. Müller, J. Shin, L. Zhao, C. Song, Rapid aggregation of global gridded crop model outputs to facilitate cross-disciplinary analysis of climate change impacts in agriculture, Environmental Modelling & Software 75 (2016) 193–201. doi:10.1016/j.envsoft.2015.10.016.

[12] C. Rosenzweig, J. Elliott, D. Deryng, A. C. Ruane, C. Müller, A. Arneth, K. J. Boote, C. Folberth, M. Glotter, N. Khabarov, K. Neumann, F. Piontek, T. A. M. Pugh, E. Schmid, E. Stehfest, H. Yang, J. W. Jones, Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison, Proceedings of the National Academy of Sciences 111 (9) (2014) 3268–3273, PMID: 24344314. doi:10.1073/pnas.1222463110.

[13] W. J. Sacks, D. Deryng, J. A. Foley, N. Ramankutty, Crop planting dates: an analysis of global patterns, Global Ecology and Biogeography 19 (5) (2010) 607–620. doi:10.1111/j.1466-8238.2010.00551.x.

[14] C. Monfreda, N. Ramankutty, J. A. Foley, Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochemical Cycles (2008) 1:19doi:10.1029/2007GB002947.

[15] M. McLennan, R. Kennell, HUBzero: a platform for dissemination and collaboration in computational science and engineering, Computing in Science & Engineering 12 (2) (2010) 48–53. doi:10.1109/MCSE.2010.41.

[16] W. J. Collins, N. Bellouin, M. Doutriaux-Boucher, N. Gedney, P. Halloran, T. Hinton, J. Hughes, C. D. Jones, M. Joshi, S. Liddicoat, G. Martin, F. O'Connor, J. Rae, C. Senior, S. Sitch, I. Totterdell, A. Wiltshire, S. Woodward, Development and evaluation of an Earth-System model – HadGEM2, Geosci. Model Dev. 4 (4) (2011) 1051–1075. doi:10.5194/gmd-4-1051-2011.

[17] J. Dufresne, M. Foujols, S. Denvil, A. Caubel, O. Marti, O. Aumont, Y. Balkanski, S. Bekki, H. Bellenger, R. Benshila, S. Bony, L. Bopp, P. Braconnot, P. Brockmann, P. Cadule, F. Cheruy, F. Codron, A. Cozic, D. Cugnet, N. d. Noblet, J. Duvel, C. Ethé, L. Fairhead, T. Fichefet, S. Flavoni, P. Friedlingstein, J. Grandpeix, L. Guez, E. Guilyardi, D. Hauglustaine, F. Hourdin, A. Idelkadi, J. Ghattas, S. Joussaume, M. Kageyama, G. Krinner, S. Labetoulle, A. Lahellec, M. Lefebvre, F. Lefevre, C. Levy, Z. X. Li, J. Lloyd, F. Lott, G. Madec, M. Mancip, M. Marchand, S. Masson, Y. Meurdesoif, J. Mignot, I. Musat,

S. Parouty, J. Polcher, C. Rio, M. Schulz, D. Swingedouw, S. Szopa, C. Talandier, P. Terray, N. Viovy, N. Vuichard, Climate change projections using the IPSL-CM5 earth system model: from CMIP3 to CMIP5, Climate Dynamics 40 (9-10) (2013) 2123–2165. doi:10.1007/s00382-012-1636-1.

[18] S. Watanabe, T. Hajima, K. Sudo, T. Nagashima, T. Takemura, H. Okajima, T. Nozawa, H. Kawase, M. Abe, T. Yokohata, T. Ise, H. Sato, E. Kato, K. Takata, S. Emori, M. Kawamiya, MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, Geosci. Model Dev. 4 (4) (2011) 845–872. doi:10.5194/gmd-4-845-2011.

[19] J. P. Dunne, J. G. John, A. J. Adcroft, S. M. Griffies, R. W. Hallberg, E. Shevliakova, R. J. Stouffer, W. Cooke, K. A. Dunne, M. J. Harrison, J. P. Krasting, S. L. Malyshev, P. C. D. Milly, P. J. Phillipps, L. T. Sentman, B. L. Samuels, M. J. Spelman, M. Winton, A. T. Wittenberg, N. Zadeh, GFDL's ESM2 global coupled Climate–Carbon earth system models. part i: Physical formulation and baseline simulation characteristics, Journal of Climate 25 (19) (2012) 6646–6665. doi:10.1175/JCLI-D-11-00560.1.

[20] M. Bentsen, I. Bethke, J. B. Debernard, T. Iversen, A. Kirkevåg, . Seland, H. Drange, C. Roelandt, I. A. Seierstad, C. Hoose, J. E. Kristjánsson, The norwegian earth system model, NorESM1-M – part 1: Description and basic evaluation of the physical climate, Geosci. Model Dev. 6 (3) (2013) 687–720. doi:10.5194/gmd-6-687-2013.

[21] R. H. Moss, J. A. Edmonds, K. A. Hibbard, M. R. Manning, S. K. Rose, D. P. van Vuuren, T. R. Carter, S. Emori, M. Kainuma, T. Kram, G. A. Meehl, J. F. B. Mitchell, N. Nakicenovic, K. Riahi, S. J. Smith, R. J. Stouffer, A. M. Thomson, J. P. Weyant, T. J. Wilbanks, The next generation of scenarios for climate change research and assessment, Nature 463 (7282) (2010) 747–756. doi:10.1038/nature08823.

[22] I. Foster, Globus online: Accelerating and democratizing science through Cloud-Based services, Ieee Internet Computing 15 (3) (2011) 70–73. doi:10.1109/MIC.2011.64.

[23] D. Pierce, ncdf4: Interface to unidata netCDF (Version 4 or earlier) format data files (Dec. 2015).
URL `https://cran.r-project.org/web/packages/ncdf4/index.html`

[24] C. F. McSweeney, R. G. Jones, How representative is the spread of climate projections from the 5 CMIP5 GCMs used in ISI-MIP?, Climate Services 1 (2016) 24–29. doi:10.1016/j.cliser.2016.02.001. URL http://www.sciencedirect.com/science/article/pii/S2405880715300170

Nr.	Code metadata description	Value
C1	Current code version	Version 1.0.0
C2	Permanent link to code/repository used for this code version	`https://mygeohub.org/tools/climatetool`
C3	Legal Code License	GNU General Public License
C4	Code versioning system used	LZ
C5	Software code languages, tools, and services used	Java, R
C6	Compilation requirements, operating environments & dependencies	None
C7	If available Link to developer documentation/manual	`https://mygeohub.org/tools/climatetool`
C8	Support email for questions	Built-in HUBzero ticket support system

Table 1: Code metadata

Climate Models	Scenarios	Crops
HadGEM2-ES [16], IPSL-CMSA-LR [17], MIROC-EXM-CHEM [18], GFDL-ESM2M [19], NorESM1-M [20]	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

Table 2: Coverage of the Climate Scenario Aggregator tool

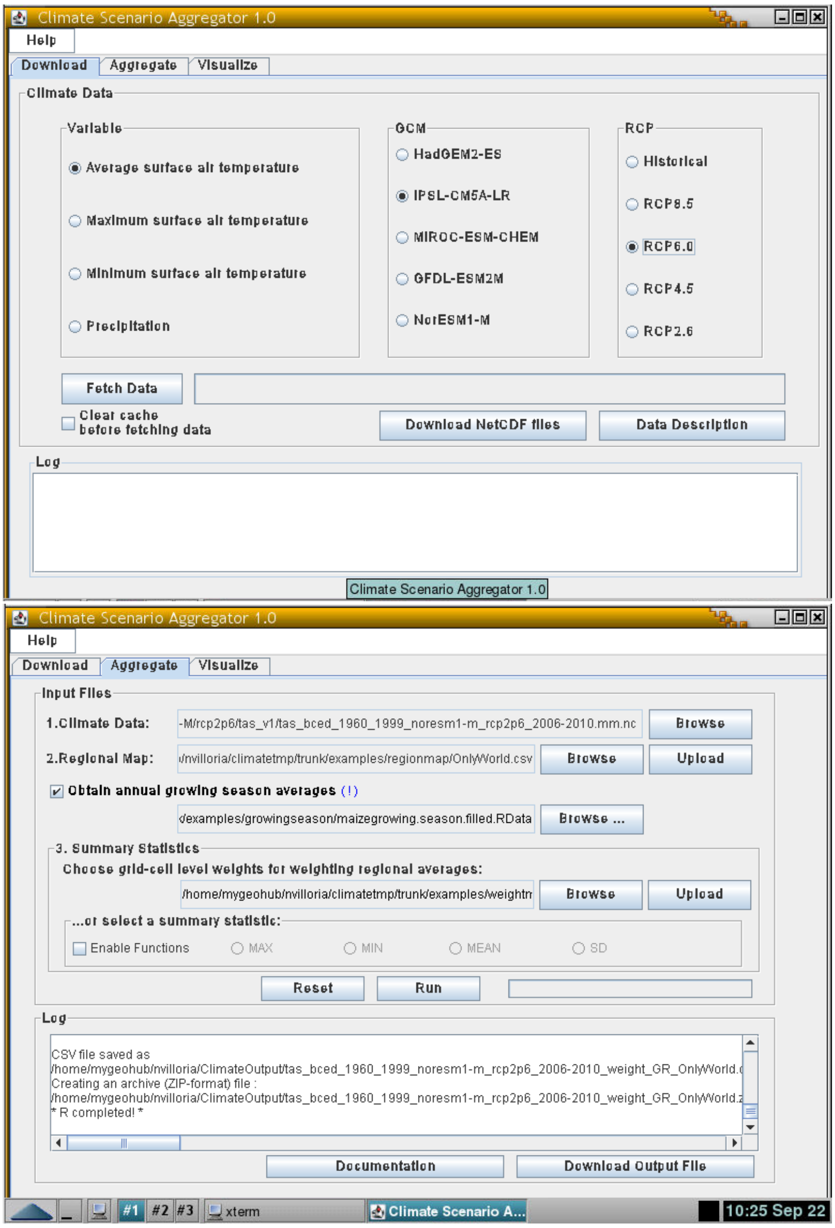

Figure 1: Climate Scenario Aggregator tool download tab. Interface for data selection and retrieval including climate model, variable, and RCP.

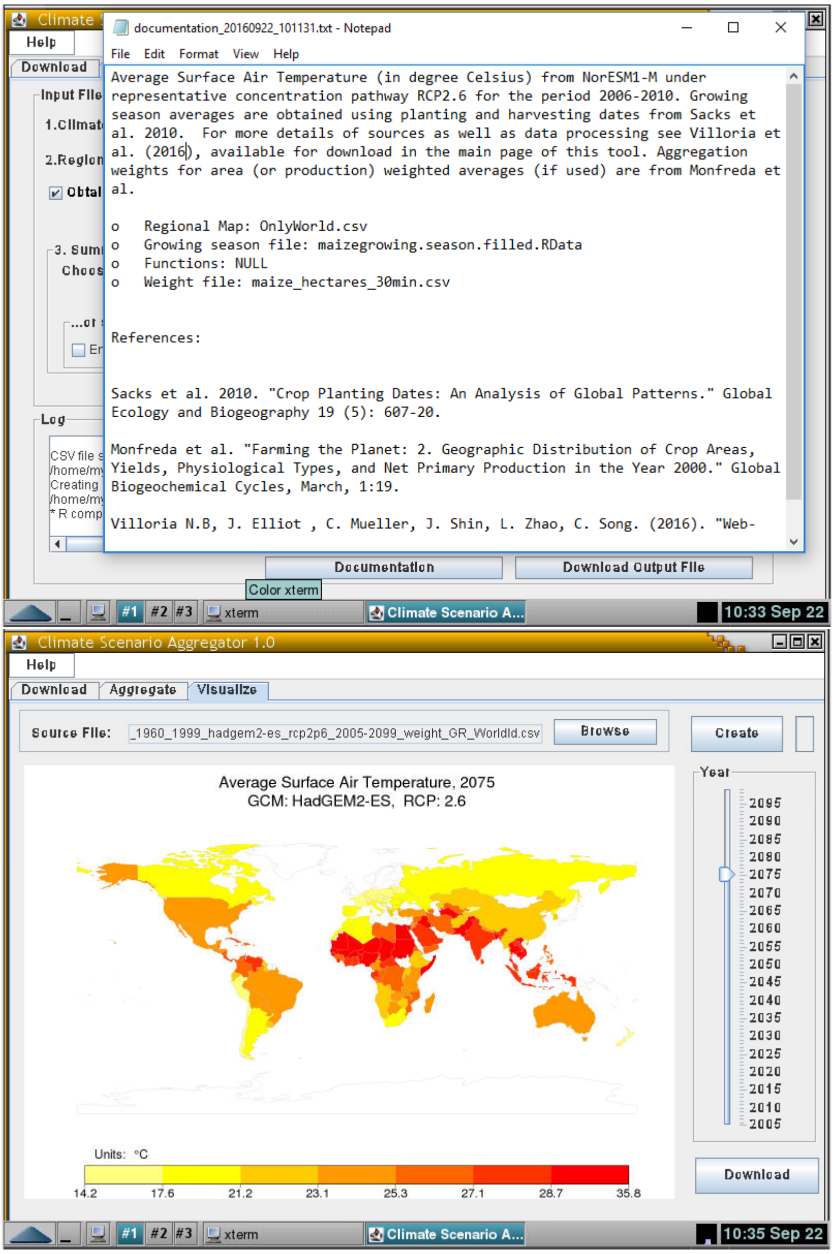

Figure 2: Climate Scenario Aggregator tool visualization, self-documentation and meta-data.

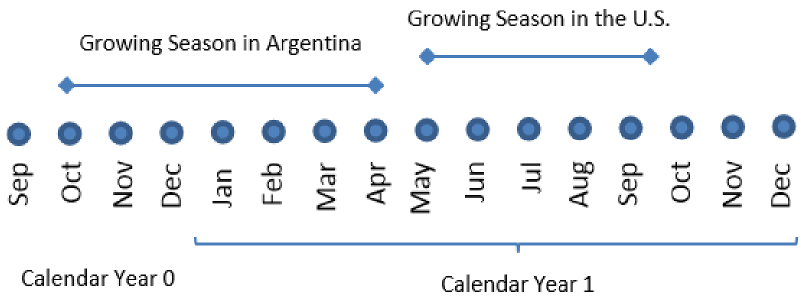

Figure 3: The average temperature/precipitation over the growing season is assigned to the calendar year in which the harvest season occurs. In the example, for Argentina, the average temperature/precipitation in Calendar Year 1 is taken over October 0-April 1 while in the U.S. (Midwest region) is taken over May 1-September 1. The planting and harvesting dates for each country are from [13].

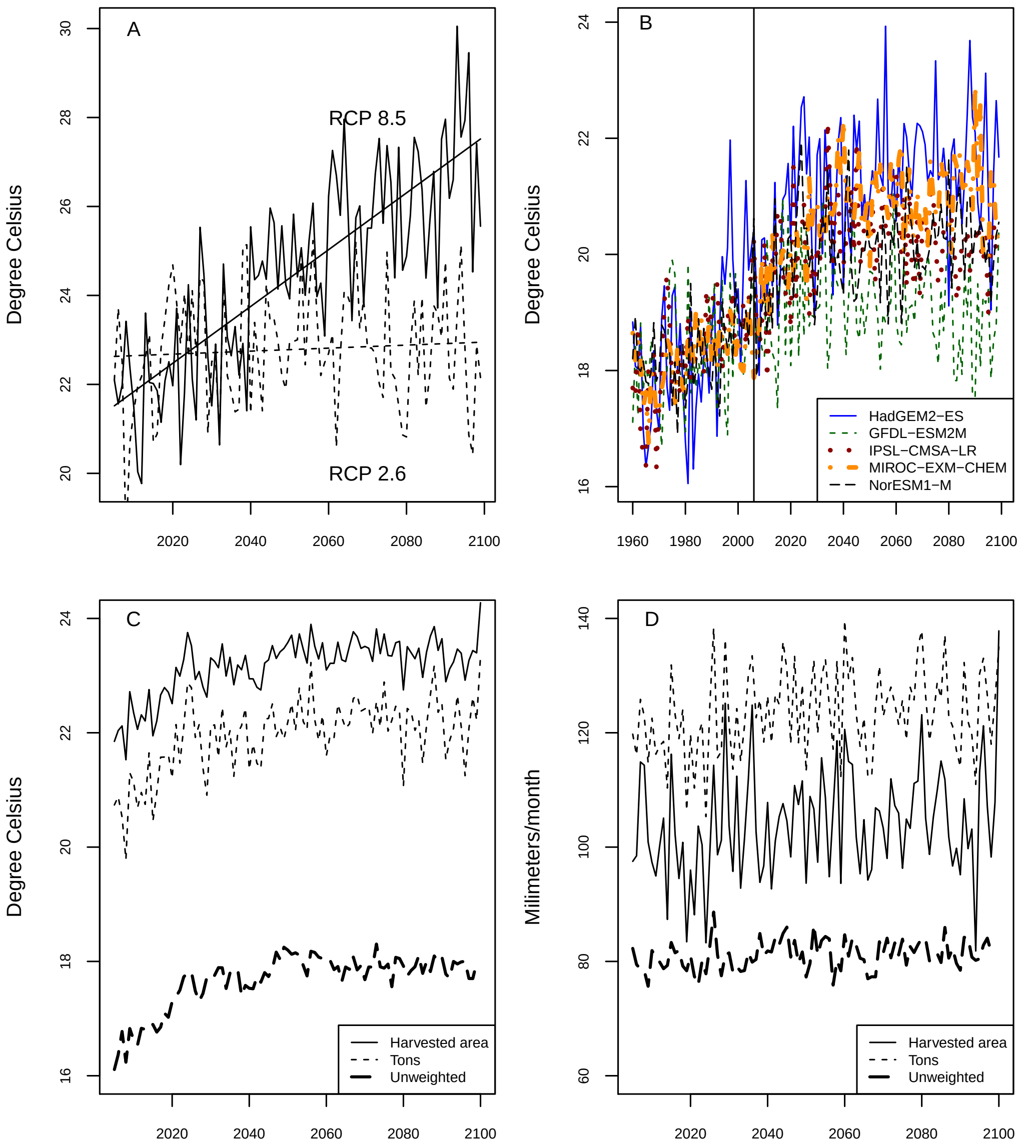

Figure 4: A: Average temperature during the wheat growing season at 96.25W-39.75N (near Manhattan, Kansas, USA) under RCP 2.6 and 8.5 from HadGEM2-ES ; B: Area-weighted average temperature during the maize growing season in the US, RCP 2.6 for the five available climate models; C: Global weighted (using harvested area and production weights) and unweighted average temperature over the maize growing season; D: Global weighted (using harvested area and production weights) and unweighted average precipitations over the maize growing season.