Software data news

Rapid aggregation of global gridded crop model outputs to facilitate cross-disciplinary analysis of climate change impacts in agriculture

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 Research Computing, Purdue University, West Lafayette, IN 47907, USA
[e] Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

ABSTRACT

We discuss an on-line tool that facilitates access to the large collection of climate impacts on crop yields produced by the Agricultural Model Intercomparison and Improvement Project. This collection comprises the output of seven crop models which were run on a global grid using climate data from five different general circulation models under the current set of representative pathways. The output of this modeling endeavor consists of more than 36,000 publicly available global grids at a spatial resolution of one half degree. We offer flexible ways to aggregate these data while reducing the technical barriers implied by learning new download platforms and specialized formats. The tool is accessed trough any standard web browser without any special bandwidth requirement.

Software availability

A tool for aggregating outputs from the AgMIP's Global Gridded Crop Model Intercomparison Project (GGCMI) is freely available at the GEOSHARE website (https://mygeohub.org/resources/agmip) using any standard Internet browser. All the programs — a java graphical user interface (GUI) and a set of R functions — can be freely downloaded and reused. The tool is free under a GNU General Public License (www.gnu.org) agreement. Documentation and support for users include a User's Manual,[1] as well as a set of default regional maps and weighting schemes.

1. Introduction

We discuss an on-line tool that facilitates access to a large

collection of climate impacts on crop yields produced by the Agricultural Model Intercomparison and Improvement Project (AgMIP; Rosenzweig et al., 2013) as part of the Global Gridded Crop Model Intercomparison Initiative (GGCMI; Elliott et al., 2014b) and the Inter-Sectoral Impacts Model Intercomparison Project (ISI-MIP; Warszawski et al., 2014). As displayed in Table 1, this collection comprises the output of seven crop models which were run on a global grid using climate data from five different general circulation models (GCM) under the current set of representative pathways (RCPs; Moss et al., 2010). The output of this modeling endeavor consists of more than 36,000 publicly available global grids at a spatial resolution of one half degree.

This information has been used to gain an understanding of the interactions among water supply, irrigation, and climate change in global caloric production (Elliott et al., 2014b); multisectoral impacts of climate change (Piontek et al., 2014); and endogenous economic responses to increases in temperature (Nelson et al., 2014). The wide range of applications of these data can be expected to greatly expand the quantitative assessment of global climate change impacts at different levels of global warming as well

* Corresponding author.
  E-mail address: nvilloria@ksu.edu (N.B. Villoria).
[1] Included as an Appendix for the reviewers convenience.

Table 1
Models and crops.

Model	Crops
EPIC	All
GEPIC	Wheat, maize, soy, rice
pDSSAT	Wheat, maize, soy
LPJmL	All
IMAGE-AEZ	All
PEGASUS	Wheat, maize, soy
LPJ-GUESS	Wheat, maize, soy

Notes. All crops: maize, soybeans, wheat, rice, managed grass, rapeseed, barley, millet, sorghum, sugarcane, sugar beets and others. In addition, each crop model was run under five different scenarios (historical and four representative concentration pathways) generated by the global circulation models HadGEM2-Es, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M, NorESM1-M.

as geographic scales.[2] Importantly, the richness in crops and GCM models can inform our understanding of model uncertainty (e.g., Piontek et al., 2014) and therefore help to impose boundaries on the potential effects of climate warming as well as to illuminate research priorities. However, and despite being open access, the technical skills required to access these data are likely to represent an important barrier for many researchers, reducing the potential impact of this information.

These barriers are particularly important for cross-disciplinary research. For instance, Hertel et al. (2010) point out that technical access barriers to geo-referenced data have slowed down our understanding of the effects of global environmental change on the long-run sustainability of the food system. From a more general perspective, Craglia et al. (2011) discuss the opportunities for online geoprocessing services to foster multidisciplinary collaboration. The advantages of on-line geoprocessing tools are many. In particular, shared access to common geospatial data results in considerable savings (Kiehle, 2006), allows users to leverage shared cyberinfrastructure for intensive computing via services such as HUBzero (McLennan and Kennell, 2010), and share workflow elements across different study areas (Yue et al., 2010; Hertel and Villoria, 2012).

Against this background, in order to facilitate the use of these data, we have built a publicly-available, open-source tool that aggregates the data from the grid-cell level to larger geographic aggregates using harvested area and production as alternative weighting schemes. The tool is implemented in GEOSHARE's HUBzero cyberinfrastructure (McLennan and Kennell, 2010) using the statistical language R (R Core Team, 2014). Both HUBzero and R are open-source systems, thus saving users costly licenses. Moreover, because the data are entirely handled and processed by GEOSHARE's HUBzero computing resources, users do not require special processing capabilities nor Internet connectivity beyond what is required for ordinary web browsing.

The main audience of the data are researchers modeling the effects of climate change on agriculture at global scales (Rosenzweig et al., 2014). Typically, these users are interested in the "shocks" to the crop productivity of a given region that are attributable to climate change (e.g., Nelson et al., 2014). The level of aggregation used in these models ranges from grid-cells (Piontek et al., 2014) to few global regions (Baldos and Hertel, 2015). Access to this type of data requires considerable skills in the use of GIS

tools or the cyberinfrastructure needed to transfer large datasets in a systematic way. Therefore, this tool bridges the gap between the crop modelers generating the data repository and the final users of this information. The tool also facilitates obtaining summary statistics which can be useful for rapid understanding of the data, regional summaries, or visualization.

The data in the AgMIP repository joins other large data collections, particularly those produced by the Coupled Model Intercomparison Project Phase 5 (CMIP5[3]). To date, the standard way of delivering these data are simple data portals where users can download the raw data, usually compressed, directly into their computers (e.g., http://www.cru.uea.ac.uk/cru/data/hrg/). A typical user of the GGCMI archive would need to set up a Globus Online client (Foster, 2011), search the desired dataset in a multi-layered folder hierarchy, download the data, and use specialized tools to extract the information from the NetCDF files in which the data are stored. Given the potentially large volume of information, data download and storing may consume significant bandwidth and hardware resources. Aggregation from the grid-cells to the desired geographic units requires significant dexterity using specialized geoprocessing tools that involve constructing aggregation weights as well as concordances between coordinates and the desired geographic regions.

The AgMIP tool goes a step forward and offers the possibility of aggregation to increase the impact of the data by facilitating its use by scientific communities working on global modeling of the effects of climate change in agriculture.[4] At the most basic level the tool is simply a download platform of the raw data. The user using this service is probably literate with the GIS formats used in the climate community and have considerable experience using programming languages in large data environments. For this user, the tool saves the need to set up a Globus Online client. The tool also serves users who prefer to leave all the aggregation tasks up to the tool, but that provide customized aggregation schemes.

The rapid uptake of the tool[5] underscores the need for offering data tools in the climate realm that can lift some of the burden required by data preprocessing. This facilitates a more efficient use of resources because researchers in different communities can focus on their areas of expertise rather than on data handling and preprocessing. Therefore, by building this tool we hope to encourage other groups producing large and complex datasets to develop their own accompanying data delivery/processing tools. In the next section we demonstrate that this can be accomplished using common-use scientific software and shared cyberinfrastructure.

2. Materials and methods

The AgMIP tool is hosted in GEOSHAREs HUBzero cyberinfrastructure (https://geoshareproject.org/). HUBzero (McLennan and Kennell, 2010), is an open source software platform specializing in disseminating simulation and data tools via the world wide web. The HUBzero environment is highly flexible and the main requirement for researchers that wish to make their code public is that the code runs on Linux (McLennan and Kennell, 2010). HUBzero provides the Rappture toolkit that facilitates the development

---

[2] For instance, Blanc and Sultan (2015) combined these data with climate data and estimated regressions that permit predicting yields from changes in climate without the need of running the underlying crop models.

[3] http://cmip-pcmdi.llnl.gov/cmip5/.

[4] The size of these user communities is potentially large. For example, the Global Trade Project Analysis (GTAP) network has approximately 12,000 members, a fourth of which are active users of the GTAP data and model, and for which the study of the economic effects of climate change on agriculture is a main focus.

[5] As of September 1, 2015, since its release in March 2014, there has been 5563 runs (or aggregations) by 71 unique users.

of interactive tools with a graphical user interface. Tools can also be developed using other popular programming languages such as Java, Python, C/C++, etc. HUBzero has a well-defined tool development process to help and guide tool developers to develop, deploy, and publish their tools. Users access the AgMIP tool at GEOSHARE using an ordinary Web browser without having to download or compile any code specific to the tools. The AgMIP tool GUI is written in Java. Each HUBzero tool (whether written in Java or not) is a desktop application that runs in an isolated light-weight virtual machine container. Users can access the tool via a web browser using a graphical desktop sharing technology called Virtual Network Computing (VNC). HUBzero currently provides two types of VNC clients embedded in the web browser: HTML5, or Java Applet based client. GEOSHARE is open and users can freely access all the data and simulation tools.[6]

Fig. 1 displays the conceptual workflow of the tool separated into three different steps. Each step in Fig. 1 is implemented as a tab in a graphical user interface shown in Figs. 2—4. The GGCMI output is stored in NetCDF files. Each file is identified by a 12 part file name that specifies crop model, climate model, representative concentration pathway, socio-economic pathway, CO_2 fertilization (yes/no), irrigation (yes/no), a crop, and a time period (See Table 1 in the User's Manual for further details). For instance:

```
pdssat_hadgem2-es_rcp2p6_ssp2_co2_noirr_yield_mai_annual_2005_2010.nc4
pdssat_hadgem2-es_rcp2p6_ssp2_co2_noirr_yield_mai_annual_2011_2020.nc4
                                    .
                                    .
                                    .
pdssat_hadgem2-es_rcp2p6_ssp2_co2_noirr_yield_mai_annual_2091_2099.nc4
```

are global grids of maize yields (one grid for each year in the period 2005—2099), projected by pDSSAT (Jones et al., 2003; Elliott et al., 2014a), using the climate projections of the Global Circulation Model HADGEM2-E, under representative concentration pathway RCP2.6, assuming no irrigation and allowing for CO_2 fertilization.

In order to retrieve the data, the first step in Fig. 1 requires the user to select a unique combination of crop model, climate model/scenario, and other choices, which are all presented in the tool's user front-end (Fig. 2). The user's selections create a character string that matches the file names stored in 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 (Foster, 2011), 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 (step 2 in Fig. 1 and GUI implementation in Fig. 3).

2.1. Aggregation

Aggregation is performed by two R functions. The first function reads the data using the R NetCDF package by Pierce (2013). The second 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 for country and country-AEZ regions. 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, or the "Help" button of the tool's GUI.

Rosenzweig et al. (2014) recommends using the gridded harvested area from the M3 crop dataset (Monfreda et al., 2008) to obtain regional production. It is important to keep in mind that this aggregation assumes that harvested areas are constant over time.

This recommendation is implemented as the first option in the Aggregation tab (Fig. 3). The underlying R code performs the following operation:

$$Production_R = \sum_{g \in R} Yield_g \times Area_g. \tag{1}$$

where R is a region comprised by a number of grid-cells indexed by g.

In many applications, users need to weight the contribution of each grid-cell to obtain a regional average. A common practice is to weight each grid cell's yield by the harvested area of the chosen crop (e.g. Nelson et al., 2014). This option is also implemented in the tool, by choosing the option "Regional Yields Weighted by each Grid-cell's Harvested Area," under Aggregation Options in Fig. 3. The operation performed by the underlying R scripts is given by:

$$Production_R = \sum_{g \in R} Yield_g \times Area_g \bigg/ \sum_{g \in R} Area_g. \tag{2}$$

As before, the default is to use the grid-cell level harvested areas from the M3 crop dataset.

Some users may be interested in other weighting schemes. For example, Baldos and Hertel (2015) use grid-cell level production to calculate production-weighted yield averages while Nelson et al. (2014) use modeled area weights from You and Wood (2006). Alternative weighting schemes can be uploaded and used instead of the provided defaults.

---

[6] HUBzero was originated in the nanotechnology community, which have developed many simulation tools (https://nanohub.org/tools/). Other HUBzero projects including driNET (http://drinet.hubzero.org), WaterHub (http://water-hub.org) and GABBS (https://mygeohub.org/groups/gabbs) provide geo-spatial data analysis tools such as 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). A complete list of HUBs can be found at https://hubzero.org/sites.
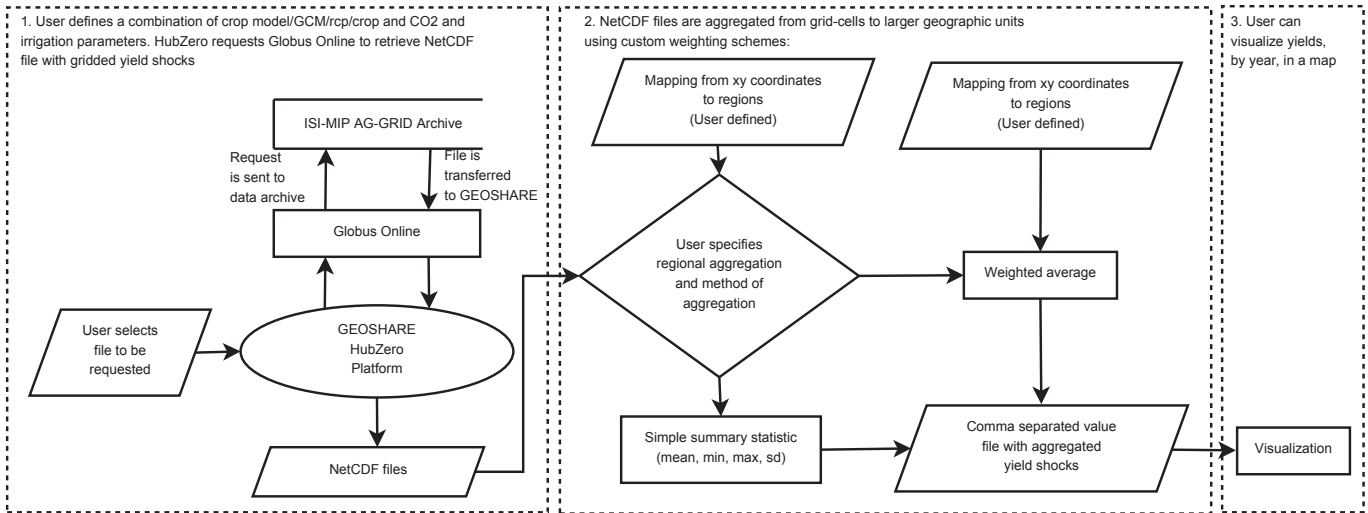
Fig. 1. AgMIP tool workflow.

2.2. Output and metadata

The output of the R function is a comma-separated-value file with two columns, one identifying the aggregated region(s) and the other the aggregated value (e.g., production weighted average maize yield). We have chosen a comma separated value (CSV) format because of its versatility. The tool also produces a detailed description (Documentation in GUI displays) of the operations performed and suggests the citations that should be included when using the tool (see example in the next section).
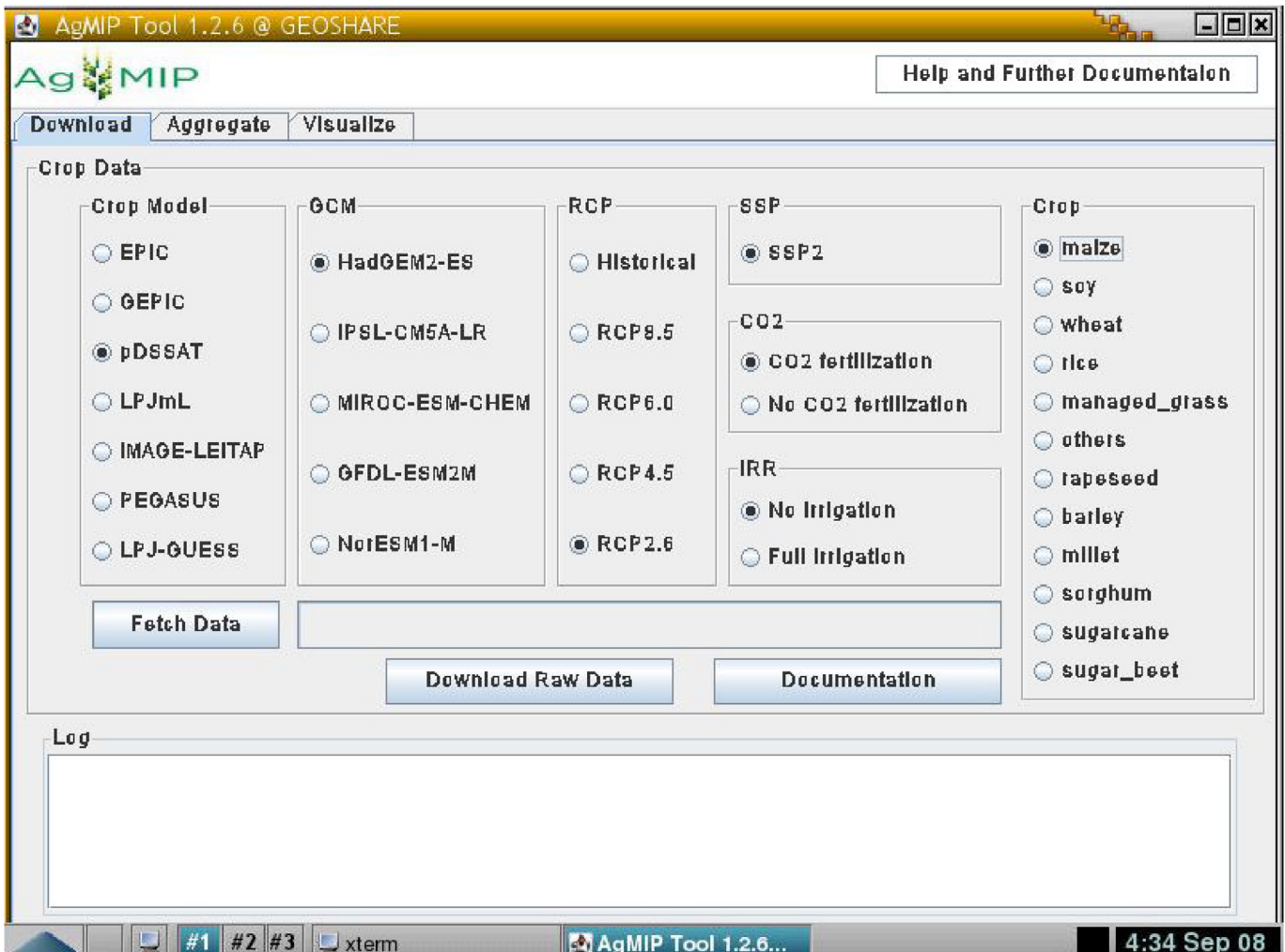

Fig. 2. AgMIP tool. Interface for data selection and retrieval including crop model, climate model, RCP, CO_2 and irrigation scenarios, and crops.
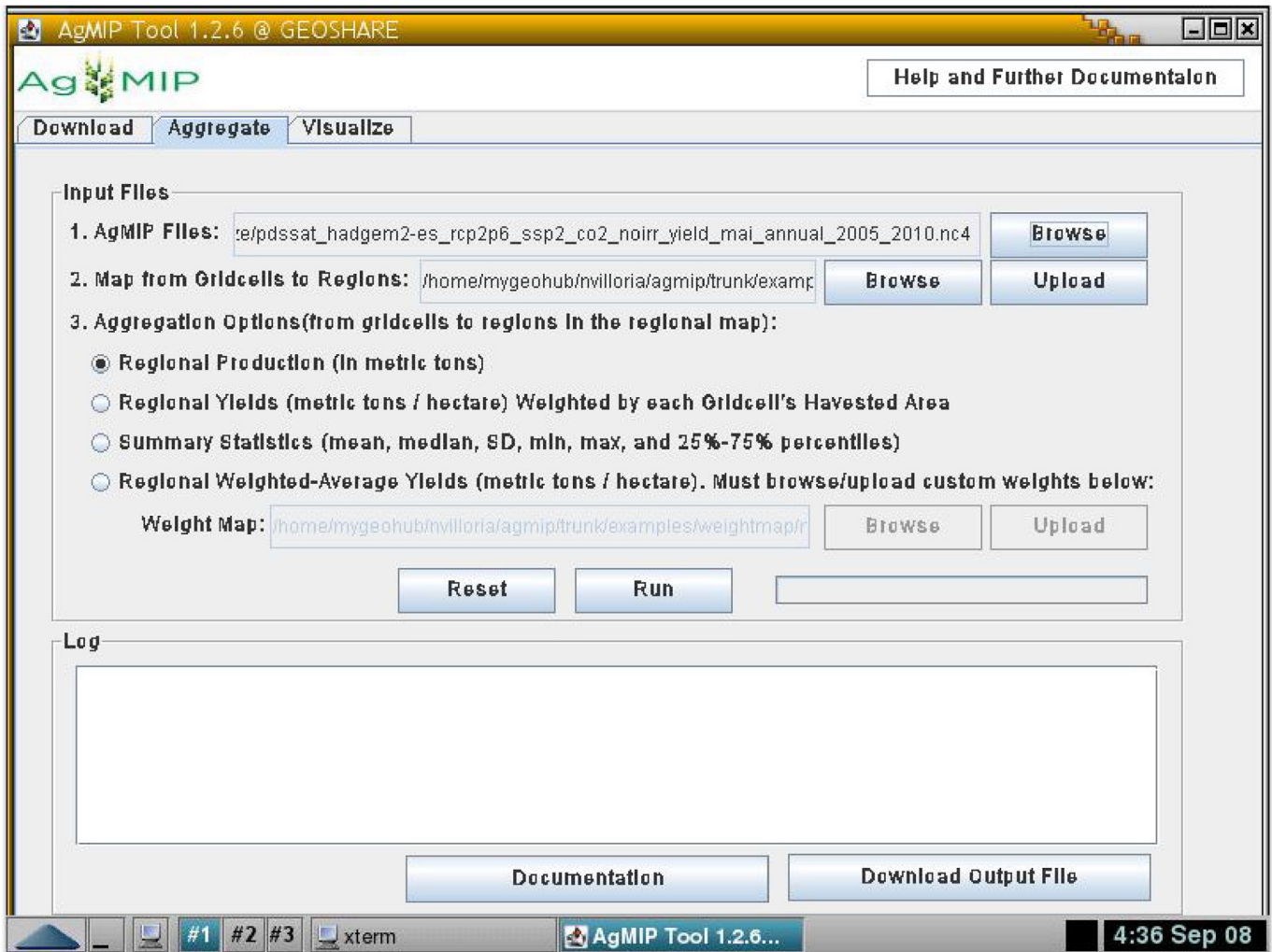
Fig. 3. AgMIP tool. Aggregation choices include production, harvested area-weighted yields, as well as weighted-averaged yields using weights defined by the user (see user manual for details). The tool also offers summary statistic for the selected regions. Users can upload their own geographic aggregation schemes.

2.3. Visualization

The last step in the workflow is to visualize the underlying information. As shown in Fig. 4, the tool creates a map of the chosen regions for each year in the dataset. These maps can be downloaded (in PNG format) for use in other applications.

3. Demonstration & discussion

The GGCMI archive has the potential to be used in many settings. For instance, Baldos and Hertel (2015) explore the extent to which the future changes in crop productivity linked to future climate trajectories may affect global food security; for this, they aggregate the GGCMI shocks to a dozen of regions encompassing the world, and run an economic model that estimates future price changes. In a different setting Blanc and Sultan (2015) combine the GGCMI shocks with climate data to create crop models "emulators" that can serve as economic alternatives to running the global crop models. As stated above, we expect the present tool to encourage further applications of these data.

We demonstrate the tool using projected maize yields from pDSSAT obtained under HadGEM2 (Collins et al., 2008) climate predictions for representative concentration pathways RCP2.6 and RCP8.5 (Moss et al., 2010). We include output with and without CO_2 fertilization. pDSSAT (Jones et al., 2003; Elliott et al., 2014b) is a

modification of the DSSAT crop model that runs in global grids, given the basic information on biophysical attributes. The user of these yield shocks is advised to consult Rosenzweig et al. (2014) for a discussion of the different models and attributes.

Fig. 2 shows the front-end of AgMIP tool with the relevant options checked. Fig. 3 displays the "Aggregation" tab. Beyond the included default mapping and weighting schemes, the tool is intended to give the user flexibility regarding choices of aggregation. To demonstrate such flexibility we created two additional regional mapping schemes. One maps each coordinate pair to a unique aggregate unit, which we label "World." The other maps each coordinate pair into one of two categories: low latitude (regions within the tropics) and mid-latitude (between 23.5° and 66°).

Fig. 5 displays the global production-weighted average maize yield from 2005 to 2099. To gain some perspective on the size of this job, consider that each RCP-CO2 fertilization scenario of pDSSAT is stored in 9 NetCDF array files, each containing 10 years worth of data. And for each year there are $720 \times 360 = 259,200$ grid-cells, including non-land pixels. Thus, for the four sets of outputs illustrated in Fig. 3 (2 RCPs, with and without CO_2), the tool downloads 36 NetCDF files to the HUBzero server and processes $324\,(36 \times 9)$ grids to aggregate to a single summary annual statistic. In the process, the final user is shielded from all the technical barriers implied by learning new download platforms and specialized formats, as well as pre-processing of the weighting
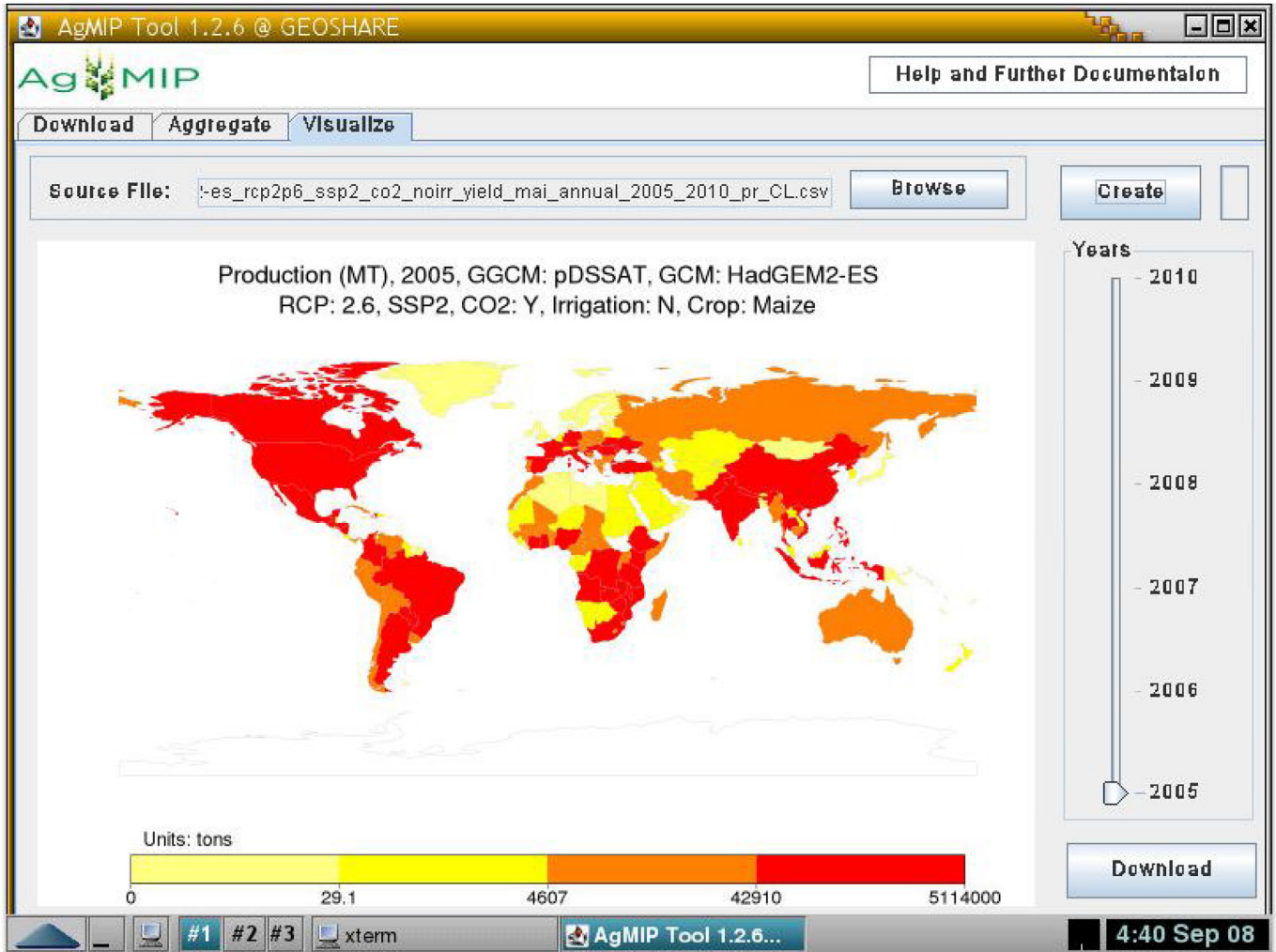
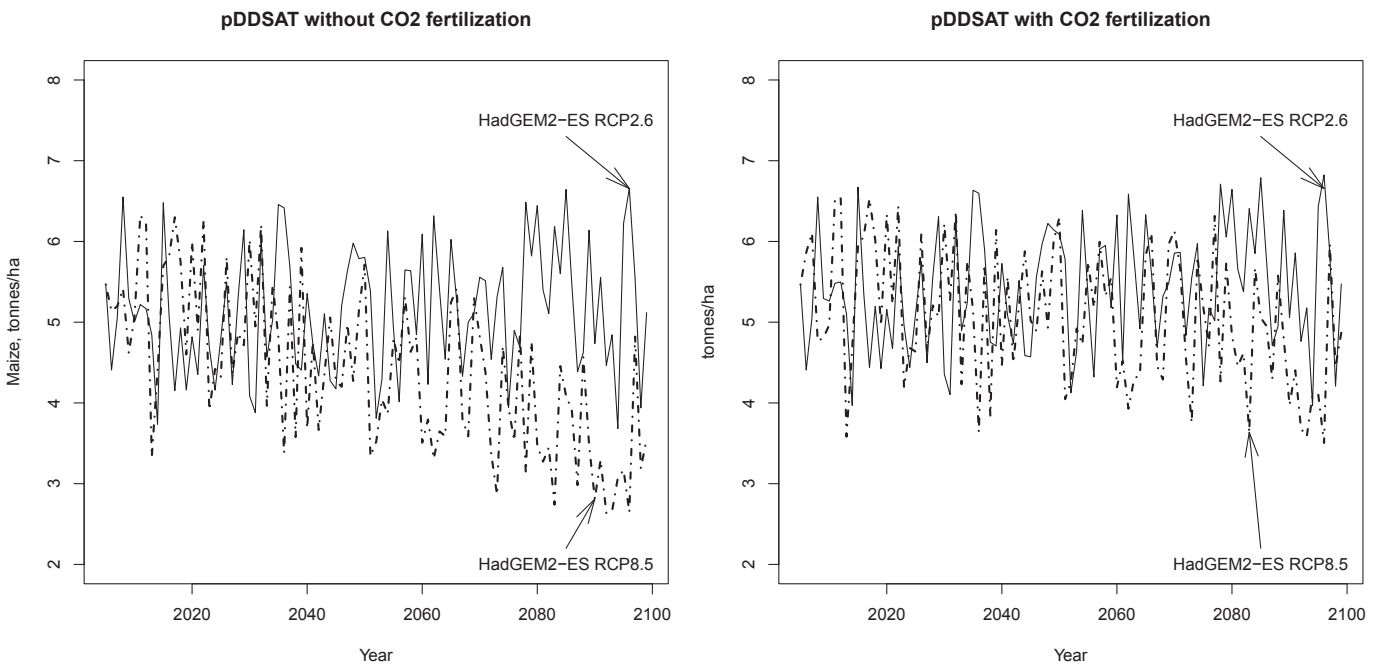**Fig. 4.** AgMIP tool. Visualization.

Fig. 5. Future maize yields (world average, weighted by grid-cell level production) projected by pDSSAT using HadGEM2-ES future climates under representative concentration pathways scenarios RCP2.6 and RCP8.5.
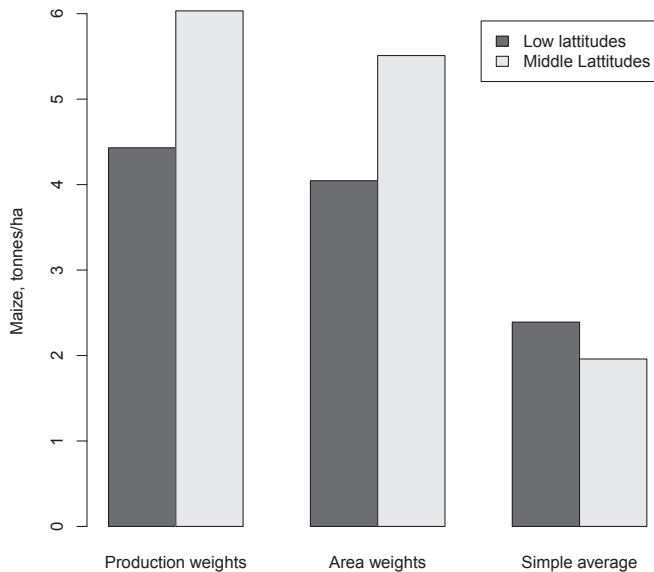
Fig. 6. Average maize yields in low and middle latitudes in the year 2050 under alternative weighting schemes. Output from pDSSAT under HadGEM2-ES RCP2.6.

schemes. Moreover, this can be all performed trough any standard web browser without any special bandwidth requirement.

In Fig. 6 we demonstrate an alternative mapping scheme in which the world is divided into low and middle latitudes as explained above. In addition, we demonstrate the effects of using production weights, area weights, and simple averages.

4. Limitations of the current GGCMI data and future research

Access to 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 short communication discusses a GEOSHARE tool that expands access to the outputs from the AgMIP GGCMI Project to the broader scientific community who can benefit from these data, but who may lack the resources to gain access to them.

The main advantage of the GGCMI ensemble is that the results are spatially consistent across the globe which would not be the case if regionally trained model results were combined for different regions. Nevertheless there are some limitations that should be considered when using these data (see Rosenzweig et al., 2014; for an in-depth discussion). First, some of the crop models have been calibrated to national or grid cell yield observations which may cause implicit assumptions that suit cropping systems in these regions better than in others (Rosenzweig et al., 2014). Second, most of the crop models lack formal sensitivity analysis to weather extremes and year-to-year weather variability. An additional caveat

is that the geographic weights (e.g., harvested area) used by the tool are assumed to be invariant to time (circa year 2000, see Monfreda et al., 2008).

The results currently delivered by the GEOSHARE AgMIP tool are part of an ongoing project (Elliott et al., 2014b). The ISI-MIP/GGCMI has started its second phase in which the different modeling groups are focusing efforts on modeling intercomparison, evaluation and improvement. The discussed tool has been designed to deliver the

new and improved datasets as they are publicly released.

The AgMIP Tool @ GEOSHARE: A GEOSHARE Tool for Aggregating Outputs from the AgMIP's Global Gridded Crop Modeling Initiative (Ag-GRID) User's Manual

How to use the AgMIP tool

This tool is freely available at geoshareproject.org.[7] Users need to register and then sign up for a free account. To access the tool, click on Resources | Tools. Then select "AgMIP Tool: A tool for aggregating …" | Launch Tool. (In order to access the GEOSHARE workspace, some users have reported the need to install a Java plug-in and/or adjust the Java security settings; the alternative is to use HTML5, which do not require Java.) After clicking on "Launch Tool", the AgMIP tool should appear in the Hub workspace.

Downloading AgMIP gridded outputs

The Download tab allows for selecting the following from the gridded archive: the crop model, climate model, representative concentration pathway, socioeconomic pathways, whether the models include CO_2 fertilization and or irrigation, and finally the crop. Clicking on download connects the tool in GEOSHARE's HUBzero with Globus Online and transfers the selected file to the workspace. The selections in the graphical user interface follow the filename conventions adopted by Elliott et al. (2014b), and summarized in Table 1.

Table 1
: Filename conventions used in the GGCMI archive.

Filename tag	Values
Crop model	epic, gepic, pdssat, lpjml, image-aez, pegasus, lpj-guess
Climate model	hadgem2-es, ipsl-cm5a-lr, miroc-esm-chem, gfdl-esm2m, noresm1-m
RCP	rcp2p2.6, rcp2p4.5, rcp2p6, rcp2p8.5, hist (historical)
SSP	ssp2
Irrigation	firr (full irrigation), noirr (noirigation)
Crop	mai (maize), soy (soybeans), whe (wheat), ric (rice), mgr (managed grass), rap (rapeseed), bar (barley), mil (millet), sor (sorghum), sug (sugarcane), sgb (sugar beets)

For instance, selecting pDSSAT | NorESM1-M | rcp2p6 | ssp2 | co2 | firr | rice, will download 10 files starting with the output for 2006—2010 until 2091—2099. The first of these would be named:

```
pdssat_noresm1-m_rcp2p6_ssp2_co2_firr_yield_ric_annual_2006_2010.nc4
pdssat_noresm1-m_rcp2p6_ssp2_co2_firr_yield_ric_annual_2011_ ...
```

and so forth. The downloaded files are stored in a string of sub-directories below "upload_area," which in turn are named after

---

these options:

```
\pdssat\noresm1-m\rcp2p6\ssp2\co2\firr\rice\
```

These files can be browsed and selected for aggregation as explained below. Notice that for each unique combination of scenarios and models there may be many files storing different years. Because of this, the tool will download all the files in the AgMIP archive whose name matches the options selected by the user.

Aggregation

The second tab offers aggregation options. As defaults, we provide a mapping from XY coordinates to ISO codes (file named WorldId.csv.) This mapping suits the needs of users wanting to aggregate to the country level. We also offer a mapping from XY coordinates to 18 national agroecological zones (Ctry18AezId.csv) a format that should be handy for users of the GTAP-AEZ framework. Beyond these aggregation schemes, users can upload their own (using the upload button.) All that is required is a.csv file with four columns labeled "","lon","lat","id" where the first column (empty label) is just a column of row numbers, the second and third columns are longitudes and latitudes, and the fourth is the unit to which "lon" and "lat" are aggregated. For instance:

```
      lon , lat ,  id
1, -179.75, 65.25, RUS14
2, -179.75, 65.75, RUS14
3, -179.75, 66.25, RUS13
4, -179.75, 66.75, RUS13
5, -179.75, 67.25, RUS13
6, -179.75, 67.75, RUS13
7, -179.75, 68.25, RUS13
8, -179.75, 68.75, RUS13
```

The "aggregation" tab also allows for either selecting or uploading a weighting map to create a weighted average yield, or if the user prefers, a summary statistic (mean, standard deviation, max and min). For convenience, we have provided grid-cell level yields and harvested area from Monfreda et al. (2008), but the user could upload alternative weighting schemes (e.g., gridded population.) As before, these files should be in plain, standard comma separated value format, with the following labels "","lon","lat","weight", where the first column (empty label) is just a column of row names. Once these options are selected, the user can run the underlying aggregating R script and either download the data, or proceed to the visualization tab (only functional for country level aggregations.)

Self documentation and attributions

The aggregation tab also includes the option to download the suggested citation along with the reference used in the process. For instance, for an aggregation that obtains the minimum values of simulated historical (1951—1960) minimum yields using the pDSSAT model, the tool self documents the different choices, producing the following text:

Maize production for the period 2005—2010 generated by the pDSSAT crop model using climate data from the HadGEM2-ES GCM under representative concentration pathway RCP 2.6 (scenario SSP2) without irrigation and with CO_2 fertilization as documented in Rosenzweig et al. (2014). Data and modeling protocols are described in Elliott et al. (2014b). Details of the aggregation procedures are in Villoria et al. (2015) which is followed by a list of the references mentioned.

Visualization

To visualize a given file, "Browse" the directory to select the source file (extension should be csv) and click on "Create". Wait until the message "Map Generated!" appears. The slider on the right allows moving across the years in the csv file.

References

Baldos, U.L.C., Hertel, T.W., Mar. 2015. The role of international trade in managing food security risks from climate change. Food Secur. 1—16. http://link.springer.com/article/10.1007/s12571-015-0435-z.

Blanc, E., Sultan, B., Mar. 2015. Emulating Maize Yields from Global Gridded Crop Models Using Statistical Estimates. Technical report. MIT Joint Program on the Science and Policy of Global Change. http://dspace.mit.edu/handle/1721.1/96707.

Collins, W., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Hinton, T., Jones, C., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., et al., 2008. Evaluation of the HadGEM2 Model. Hadley Cent. Tech. Note 74.

Craglia, M., de Bie, K., Jackson, D., Pesaresi, M., Remetey-Fülöpp, G., Wang, C., Annoni, A., Bian, L., Campbell, F., Ehlers, M., van Genderen, J., Goodchild, M., Guo, H., Lewis, A., Simpson, R., Skidmore, A., Woodgate, P., Dec. 2011. Digital earth 2020: towards the vision for the next decade. Int. J. Digital Earth 5 (1), 4—21. 10.1080/17538947.2011.638500.

Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., Foster, I., 2014a. The parallel system for integrating impact models and sectors (pSIMS). Environ. Model. Softw. 1—8. http://www.sciencedirect.com/science/article/pii/S1364815214001121.

Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K.J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R.C., Mueller, N.D., Ray, D.K., Rosenzweig, C., Ruane, A.C., Sheffield, J., Jul. 2014b. The global gridded crop model intercomparison: data and modeling protocols for phase 1 (v1.0). Geosci. Model Dev. Discuss. 7 (4), 4383—4427. http://www.geosci-model-dev-discuss.net/7/4383/2014/.

Foster, I., 2011. Globus online: accelerating and democratizing science through cloud-based services. IEEE Internet Comput. 15 (3), 70—73. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber.

Hertel, T.W., Britz, W., Diffenbaugh, N.S., Ramankutty, N., Villoria, N., Oct. 2010. A Global, Spatially Explicit, Open Source Data Base for Analysis of Agriculture, Forestry, and the Environment: Proposal and Institutional Considerations. Report to the UK science advisor. Department of Agricultural Economics, Purdue University.

Hertel, T.W., Villoria, N., Apr. 2012. GEOSHARE: Geospatial Open Source Hosting of Agriculture, Resource and Environmental Data for Discovery and Decision Making. GPRI Digital Library. http://docs.lib.purdue.edu/gpridocs/7.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., Jan. 2003. The DSSAT cropping system model. Eur. J. Agron. 18 (3—4), 235—265. http://www.sciencedirect.com/science/article/pii/S1161030102001077.

Kiehle, C., Dec. 2006. Business logic for geoprocessing of distributed geodata. Comput. Geosci. 32 (10), 1746—1757. http://www.sciencedirect.com/science/article/pii/S0098300406000756.

McLennan, M., Kennell, R., Mar. 2010. HUBzero: a platform for dissemination and collaboration in computational science and engineering. Comput. Sci. Eng. 12 (2), 48—53. http://scitation.aip.org/content/aip/journal/cise/12/2/10.1109/MCSE.2010.41.

Monfreda, C., Ramankutty, N., Foley, J.A., Mar. 2008. Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Glob. Biogeochem. Cycles 1, 19.

Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., van Vuuren, D.P., Carter, T.R., Emori, S., Kainuma, M., Kram, T., Meehl, G.A., Mitchell, J.F.B., Nakicenovic, N., Riahi, K., Smith, S.J., Stouffer, R.J., Thomson, A.M., Weyant, J.P., Wilbanks, T.J., Feb. 2010. The next generation of scenarios for climate change research and assessment. Nature 463 (7282), 747—756. http://www.nature.com/nature/journal/v463/n7282/abs/nature08823.html.

Nelson, G.C., Valin, H., Sands, R.D., Havlík, P., Ahammad, H., Deryng, D., Elliott, J., Fujimori, S., Hasegawa, T., Heyhoe, E., Kyle, P., Lampe, M.V., Lotze-Campen, H., d'Croz, D.M., Meijl, H. v., Mensbrugghe, D. v. d., Müller, C., Popp, A., Robertson, R., Robinson, S., Schmid, E., Schmitz, C., Tabeau, A., Willenbockel, D., Mar. 2014. Climate change effects on agriculture: economic responses to biophysical shocks. Proc. Natl. Acad. Sci. 111 (9), 3274—3279. PMID: 24344285. http://www.pnas.org/content/111/9/3274.

Pierce, D., 2013. ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. http://dwpierce.com/software.

Piontek, F., Müller, C., Pugh, T.A.M., Clark, D.B., Deryng, D., Elliott, J., González, F.d.J.C., Flörke, M., Folberth, C., Franssen, W., Frieler, K., Friend, A.D., Gosling, S.N., Hemming, D., Khabarov, N., Kim, H., Lomas, M.R., Masaki, Y., Mengel, M., Morse, A., Neumann, K., Nishina, K., Ostberg, S., Pavlick, R., Ruane, A.C., Schewe, J., Schmid, E., Stacke, T., Tang, Q., Tessler, Z.D.,

Tompkins, A.M., Warszawski, L., Wisser, D., Schellnhuber, H.J., Mar. 2014. Multisectoral climate impact hotspots in a warming world. Proc. Natl. Acad. Sci. 111 (9), 3233–3238. PMID: 24344270. http://www.pnas.org/content/111/9/3233.

R Core Team, 2014. R: a Language and Environment for Statistical Computing. http://www.R-project.org.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., Mar. 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. 111 (9), 3268–3273. PMID: 24344314.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., Mar. 2013. The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agric. For. Meteorol. 170, 166–182. http://www.sciencedirect.com/science/article/pii/S0168192312002857.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., Schewe, J., Mar. 2014. The inter-sectoral impact model intercomparison project (ISI–MIP): project framework. Proc. Natl. Acad. Sci. 111 (9), 3228–3232. PMID: 24344316. http://www.pnas.org/content/111/9/3228.

You, L., Wood, S., Oct. 2006. An entropy approach to spatial disaggregation of agricultural production. Agric. Syst. 90 (1–3), 329–347. http://www.sciencedirect.com/science/article/pii/S0308521X06000096.

Yue, P., Gong, J., Di, L., Yuan, J., Sun, L., Sun, Z., Wang, Q., Dec. 2010. GeoPW: laying blocks for the geospatial processing web. Trans. GIS 14 (6), 755–772. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9671.2010.01232.x/abstract.