Environmental Modelling & Software 75 (2016) 193–201

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

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

Article history:
Received 4 November 2014
Received in revised form
9 September 2015
Accepted 18 October 2015
Available online 11 November 2015

Keywords:
Crop models
AgMIP
CMIP5
Global agriculture
Climate change

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.

http://dx.doi.org/10.1016/j.envsoft.2015.10.016
1364-8152/

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 on-line 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.