

Aggregation of Gridded Emulated Projections at the National or Regional Level: Rainfed and Irrigated Crop Yields and Irrigation Water Requirements

BY ELODIE BLANC^a

This article and companion code is an update to Blanc (2017b) which provided a tool to use statistical emulators of global gridded crop models described in Blanc and Sultan (2015) and Blanc (2017) and aggregated the projections at the regional level. This new version includes, in addition to rainfed yields of maize, rice, soybean and wheat, irrigated crops yields as well as associated irrigation water requirements as estimated in Blanc (2020).

JEL codes: Q19, Q54.

Keywords: Crop yields; Crop model; Statistical model; Climate change.

1. Introduction

To benefit from the capabilities of processed-based models while preserving the simplicity and accessibility of statistical models, Blanc and Sultan (2015), Blanc (2017) and Blanc (2020) developed an ensemble of statistical emulators of global gridded crop models (GGCM). These emulators provide estimates of rainfed and irrigated crop yields as well as irrigation water requirements for maize, rice, soybean, and wheat at the grid cell level (0.5°x0.5° resolution).

These crop model emulators are estimated using simulations from the ISI-MIP Fast Track experiment (Rosenzweig et al., 2013; Warszawski et al., 2014) for five different GGCMs: the Geographic Information System (GIS)-based Environmental Policy Integrated Climate (GEPIC) model (Liu et al., 2007; Williams & Singh, 1995), the Lund Potsdam-Jena managed Land (LPJmL) dynamic global vegetation and water balance model (Bondeau et al., 2007; Waha et al., 2012), the Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS) with managed land model (Bondeau et al., 2007; Lindeskog et al., 2013; Smith et al., 2001), the parallel Decision Support System for Agro-technology Transfer (pDSSAT) model (Elliott

^a Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139 (eblanc@mit.edu).

et al., 2013; Jones et al., 2003), and the Predicting Ecosystem Goods And Services Using Scenarios (PEGASUS) model (Deryng et al., 2011). Simulations from these GGCMs are driven by climate simulations from the Coupled Model Intercomparison Project, phase 5 (CMIP5) archive (Hempel et al., 2013; Taylor et al., 2012).

The statistical emulators capture the determinants of crop yields in GGCMs using a simple set of environmental variables. The preferred specification in Blanc (2017) and Blanc (2020), S1fpintsoil, represents the effect of summer mean temperature and accumulated precipitation, and annual carbon dioxide (CO₂) concentration on crop yields and irrigation water requirements using a flexible fractional polynomial specification. It also accounts for parameter heterogeneity across soil types.

The statistical emulators output annual estimates of crop yields and irrigation water requirements at the grid cell level. To obtain regional estimates, this paper and companion code¹ provide an aggregation tool facilitating the use of the emulators under user-defined climate change scenarios. Regional estimates obtained with this tool can then be used as input into diverse models (e.g. numerical equilibrium models, water models, etc). This paper is an extension of Blanc (2017b), which provided a similar tool but limited to rainfed yields only. In this new version, the emulators are available for irrigated crop yields as well as irrigation water requirements. It follows a similar methodology to Blanc (2017b), as described in Section 2. The companion code has been improved and is described in Section 3. Corresponding instructions to run the tool are provides in Section 4. Section 5 concludes.

2. Regional aggregation of gridded crop yields

Similarly to Blanc (2017b), simulated gridded crop yields are aggregated at the regional or national level weighting yields and irrigation water requirements by the area cultivated for each grid cell using the MIRCA2000 (Portmann et al., 2010) data of harvested for each crop. Annual harvested areas of rainfed and irrigated crops are represented in Figure 1 and 2 respectively.

Gridded crop yields are aggregated at the regional aggregation following the equation:

$$\overline{YRF}_{c,r,y,ggcm} = \frac{\sum_{g=1}^G YRF_{c,g,y,ggcm} * ARF_{c,g}}{\sum_{g=1}^G ARF_{c,g}} \quad (1)$$

¹ The companion code is included in the supplementary materials published with this article.

where for each crop, c , region, r , year, y , and GGCM, $ggcm$, average rainfed yield, $\overline{YRF}_{c,r,y,ggcm}$ (in t/ha), is calculated by first multiplying yield estimates, YRF , at the grid cell level, g , by the corresponding rainfed harvested area, ARF , at the grid cell level, g , and summing over all grid cells within the region. The sum of rainfed production is then divided by the total sum of rainfed harvested area within the region.

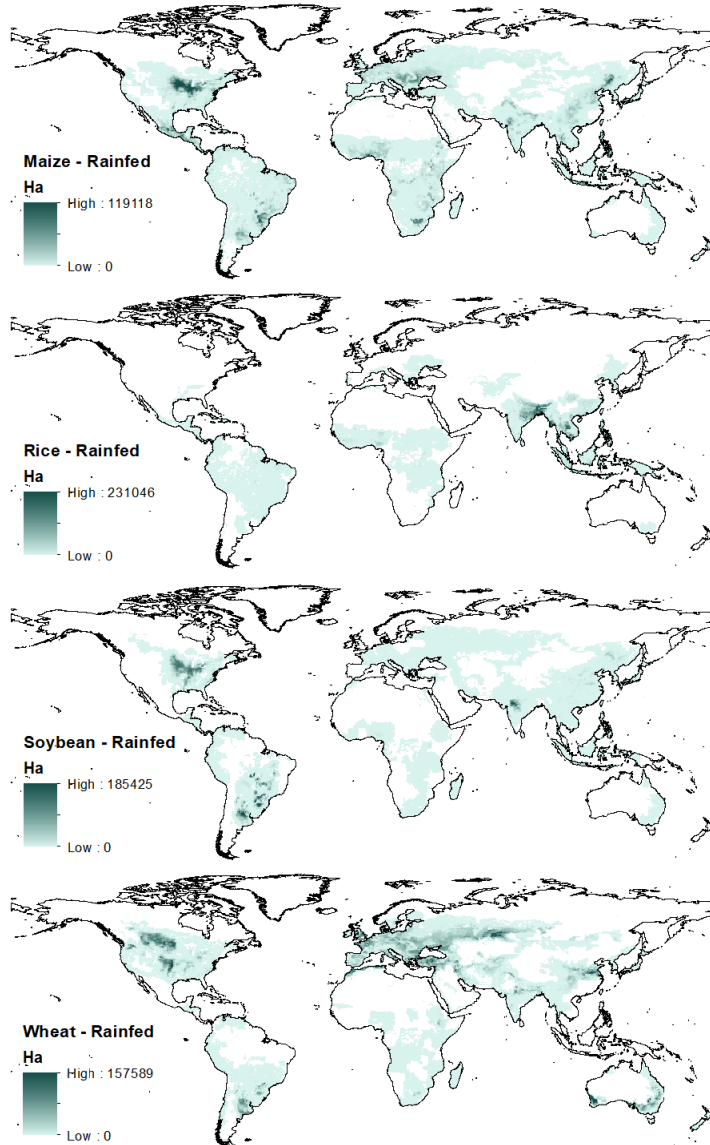


Figure 1. MIRCA2000 annual harvested areas of rainfed crop

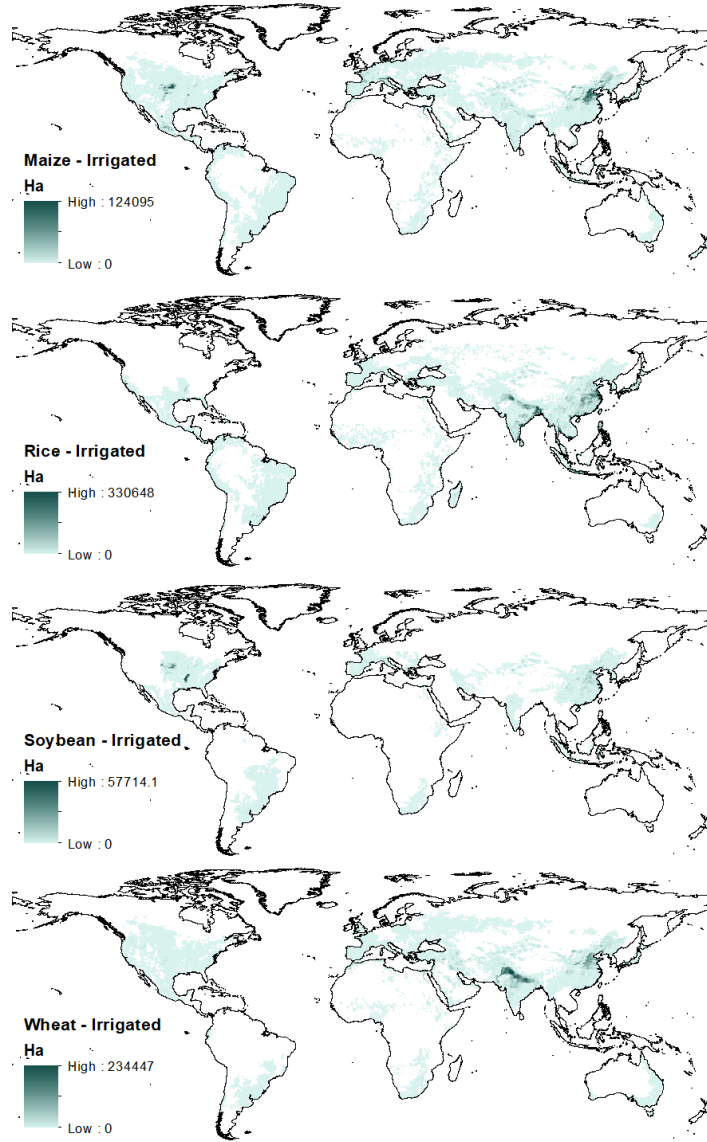


Figure 2. MIRCA2000 annual harvested areas of irrigated crop

Similarly, irrigated yields are aggregated at the regional level as follow:

$$\overline{YIR}_{c,r,y,ggcm} = \frac{\sum_{g=1}^G YIR_{c,g,y,ggcm} * AIR_{c,g}}{\sum_{g=1}^G AIR_{c,g}} \quad (2)$$

with YIR (in t/ha) corresponding to irrigated yields and AIR referring to irrigated areas.

For irrigation water requirement, regional estimates are given as:

$$\overline{PIRRWW}_{c,r,y,ggcm} = \frac{\sum_{g=1}^G PIRWW_{c,g,y,ggcm} * AIR_{c,g}}{\sum_{g=1}^G AIR_{c,g}} \quad (3)$$

with $PIRRWW$ (in mm/year) corresponding to the irrigation water demand, or potential irrigation water withdrawal, for each crop, c , region, r , year, y , and GGCM, $ggcm$, and AIR referring to irrigated areas. As emulated $PIRRWW$ assume an irrigation use efficiency of 100%, the user will need information on irrigation efficiency during conveyance and application to obtain gross irrigation water requirements (i.e. the overall amount of water necessary for irrigation).

3. Processing tool folder structure

This paper provides a Stata code (available for download from in the supplementary material section published with this article) to facilitate the use of the statistical emulators. Using the response functions of the emulators, this tool provide estimates of crop yields and irrigation requirements under given climate data inputs and aggregate them at the specified regional level. The program is written in Stata 14 and is located in the folder `.\Stata code`. This folder contains the master process called `run_emulator.do`. It also contains five sub-folders: (i) `.\Code`, containing the code of sub-processes; (ii) `.\Data`, containing weather and land use inputs; (iii) `.\Parameters`, containing response function parameters; (iv) `.\Shapefile`, containing shapefiles for output regions delineation; and (v) `.\Results`, containing the outputs of the program. Each components is described below.

3.1. Code

The `.\Stata code\Code` folder is composed of `inc_gridcell_regions.do`, `inc_install.do`, `cr_projections.do` and `cr_maps.do`, the three sub-processes called by the master process `run_emulator.do`.

3.2. Data

The `.\Stata code\Data` folder is composed of the sub-folder `.\Climate`, which is contains weather data, and `.\Land use`, containing the land use mask (crop growing area).

To run the program, the user must provide monthly average temperature and precipitation during summer: June, July, and August for maize, rice and soybean

and May, June, and July for wheat in the northern hemisphere; and December, January, and February for maize rice and soybean, and November, December, and January for wheat in the southern hemisphere. The user must also provide annual CO₂ concentration data. As an example, the code comes with three climate input data in excel format located in `.\Data\Climate\raw`. The climate data correspond to those used to estimate the response functions in Blanc (2017). A code `cr_climate_summer.do` is available to process these example in excel format, rename and assign the weather variables to the proper growing season. It outputs the climate data in Stata format into the sub-folder `.\Climate\wheat` for the wheat growing season, and `.\Climate\oth` for other the growing season of all other crops.

The `.\Land use` sub-folder contains the MIRCA2000 data of rainfed and irrigated harvested area of each crop at the grid cell level. As described in Section 2, these files are used to average crop yields over regions.

3.3. Parameters

Within the `.\Stata code\Parameters` folder, the sub-folder `.\Estimates` contains coefficient estimates for the S1fpinsoil specification in `*.ster` format for each crop, model and soil category. The sub-folder `.\Fixed Effects` contains the corresponding fixed effects coefficients in `*.dta` format. In the `.\FP` sub-folder, the file `FP_formula.dta` contains all fractional polynomial transformations to be applied to each variable, for each crop, model and soil category.

3.4. Shapefiles delineating output regions

To average grid-cell level projections of crop yields and irrigation water requirements at the regional level, spatial delineations of regions (i.e. shapefiles) are provided in the folder `.\Stata code\Shapefile`. Files for each region are located in different sub-folders: world countries are located in `.\world`, GTAP9 regions (Aguiar et al., 2016) in `.\gtap9`, EPPA6 regions (Chen et al., 2015) in `.\EPPA6`, EPPA5 regions (Chen et al., 2017) in `.\EPPA5` and WRS regions (Strzepek et al., 2013) in `.\WRS`. Within each of these sub-folders, a shapefile of the regions of interest is included along with a corresponding gridded shapefile at the 0.5°x0.5° degree resolution. The centroid coordinates of each grid cell of these gridded shapefiles and the corresponding country name are extracted in a excel file labeled `*_grid.xlsx`. Maps representing the different regions already provided with the code are represented in Figures 3 to 7. The sub-folder folder `.\grid` contains a gridded shapefile of the world representative of the emulators' resolution. This file is necessary to create new regional delineations (see instructions from Blanc (2017b)).

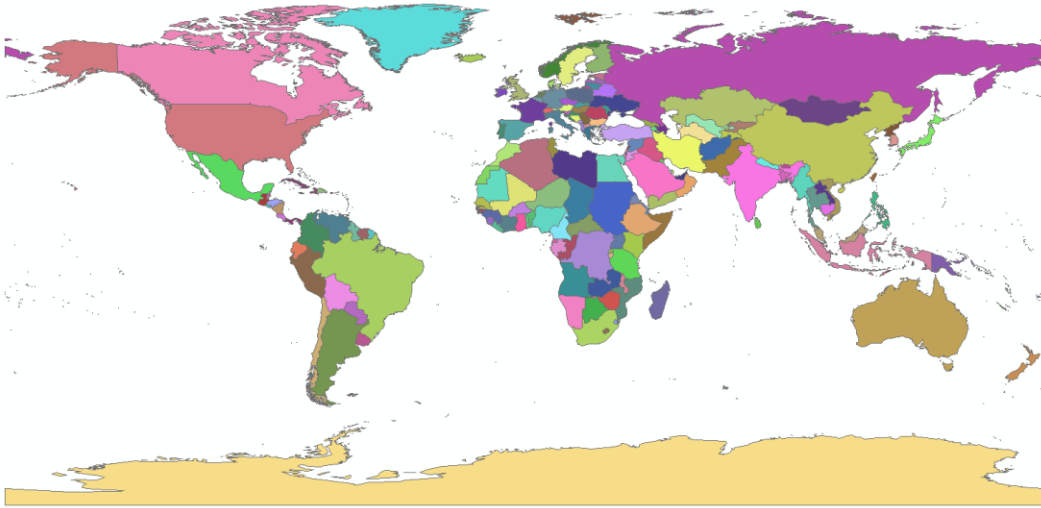


Figure 1. World regions (countries)

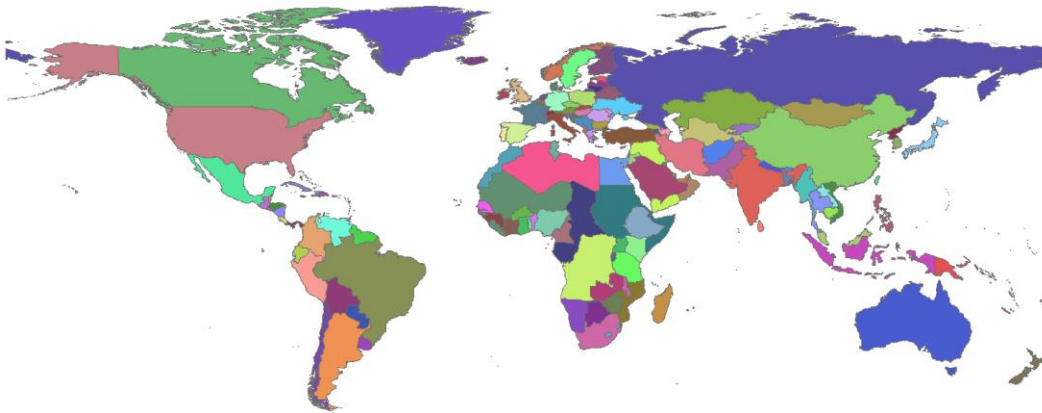


Figure 2. GTAP9 regions

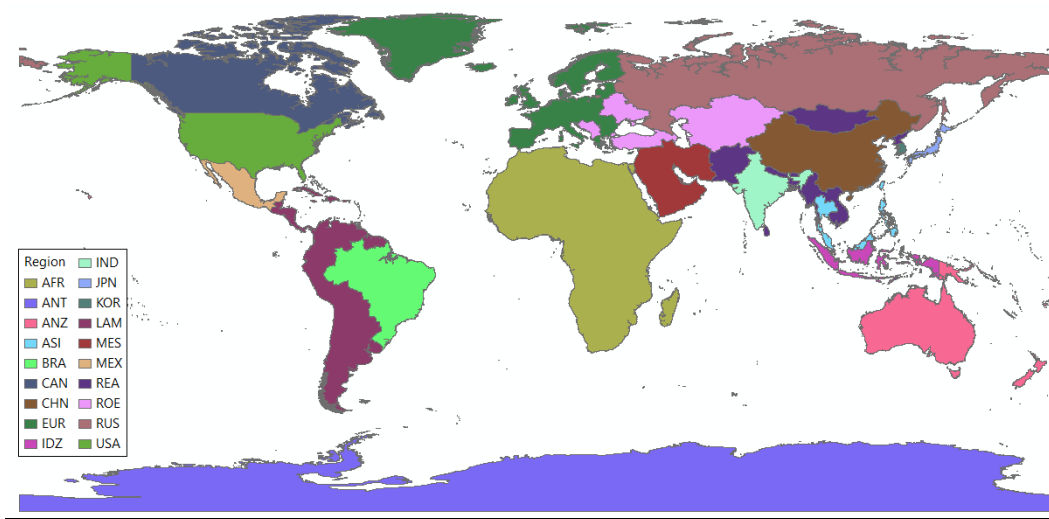


Figure 3. EPPA6 regions

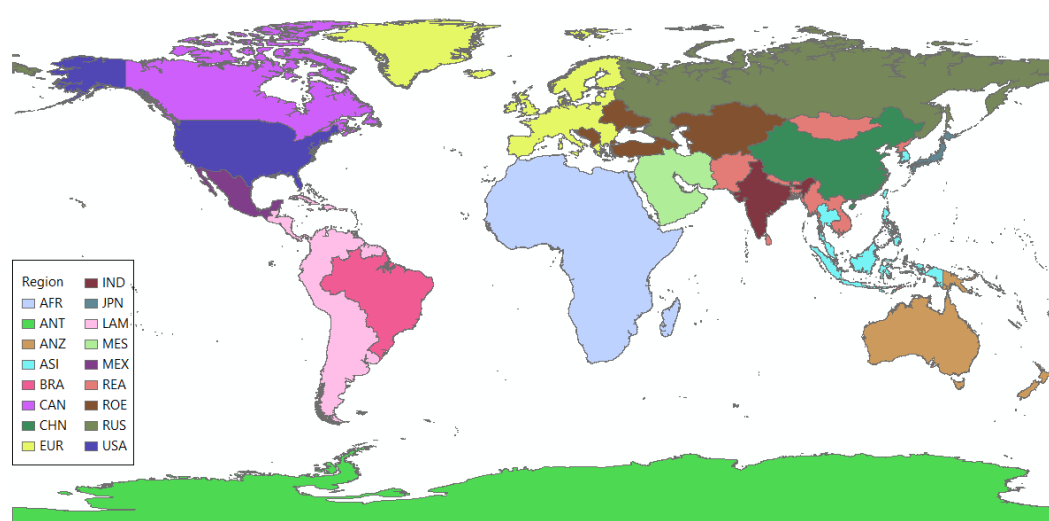


Figure 4. EPPA5 regions

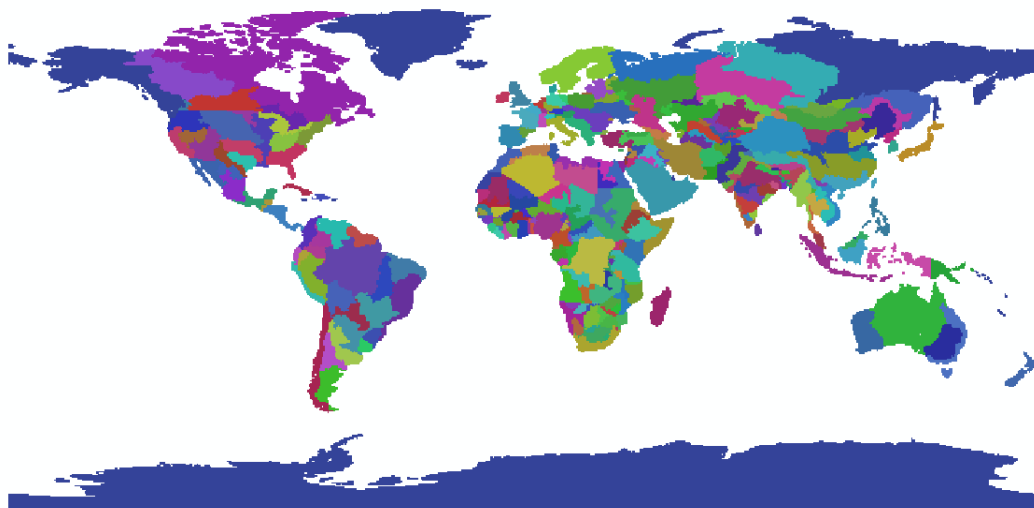


Figure 7. WRS regions

3.5. Results

The sub-process `cr_projections.do` creates regional averages of crop yields in metric tons per hectare (t/ha) for each year and place them in the sub-folder `.\Stata code\Results`. Corresponding maps created by the program `cr_maps.do` are also placed in this sub-folder. The name of the created files start by `Preds_*` and `Map_*` respectively, and the remainder of the file name details the options set up by the user at the beginning of the `run_emulator.do` code.

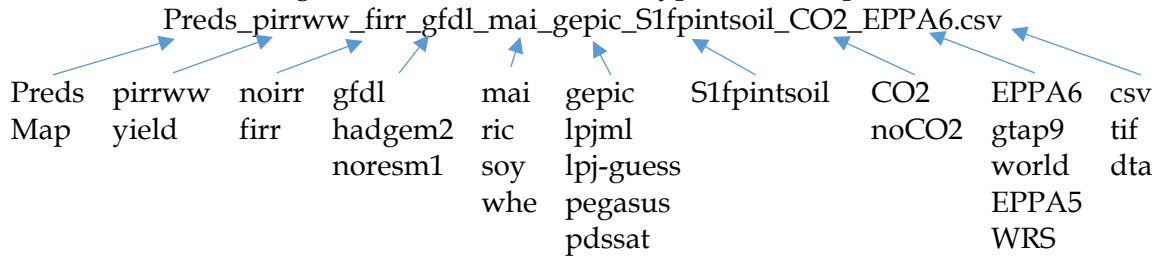
4. Running instructions

To run the emulator tool, follow the instructions below:

1. Download and unpack the `Stata code` folder to `C:\`. The folder can also be copied in another local directory, but the user will have to change the working directory (see point 6 below). The uncompressed size is 17.3 GB.
2. Prepare climate data at 0.5 degree resolution. The final climate datasets should be in Stata format and include the following variables:
 - o `gcm`: name of the climate model (or simulation).
 - o `year`: year.
 - o `pr_1`, `pr_2` and `pr_3`: monthly summer precipitation (in mm/day) for each month of summer.
 - o `Tmean_1`, `Tmean_2` and `Tmean_3`: monthly summer average temperature (in °C) for each month of summer.
 - o `co2`: annual CO₂ concentration.
 - o `lat_d` and `lon_d`: latitude and longitude of grid cells (in decimal degrees).

3. As growing seasons differ between wheat and the other crops, climate data for wheat should be placed in `.\Stata code\Data\Climate\wheat` and for the other crops in `.\Stata code\Data\Climate\oth`. The name of the file should correspond to the name of the climate model (or simulation).
4. Open the `run_emulator.do` file in `.\Stata code` and amend the options located at the top of the file. For options where the variable name ends with `*list` (e.g. `modellist`), the user can specify a list of options. For the other variables, only one option can be specified (make sure that there is no leading or trailing space).
5. The general options to specify are:
 - o `path`: directory of the `.\Stata code` folder (in the original version the working directory is `local path="C: \Stata code"`). If the folder is copied in a different directory, change the (e.g. `local path="C:\My folder\Stata code"`)
 - o `regions`: name of the region delineation required for aggregation. In the standard setup, five options are available: `gtap9`, `world`, `EPPA6`, `EPPA5` and `WRS`.
 - o `gcmlist`: list of climate change scenarios. In the standard setup, three predefined scenarios are available: `gfdl`, `hadgem2`, and `noresml`.
 - o `croplist`: list of crops to consider among maize (`mai`), rice (`ric`), soybean (`soy`) and wheat (`whe`).
 - o `modellist`: list of GCMs to consider among LPJmL (`lpjml`), LPJ-GUESS (`lpj-guess`), PEGASUS (`pegasus`) and pDSSAT (`pdssat`).
 - o `co2`: the option `yes` indicates to account for the CO₂ effect. The option `no` assumes that CO₂ remains constant at base year level (first year of dataset). This option can be used to tease out the specific contribution of CO₂ fertilisation effect on crop yields.
6. Model specific options:
 - o `gepic_seas`: the option `yes` will reproduce GEPIC's 10-year seasonality (GEPIC simulations are run independently for each decade to account of soil fertility erosion). The option `no` = will provide an average yields, without the GEPIC seasonality.
 - o `pdssat_seas`: the option `yes` will reproduce pDSSAT input of CO₂ every 30 years, and therefore update CO₂ every 30 years. The option `no` will take new values of CO₂ every year.
7. Specify the map options to create maps of the outputs:

- type: The option `level` provides a map of crop yields in level over a given period. The option `change` provides a map of crop yields in terms for changes between a present period and a future period.
- 8. For the option `level`:
 - `fyear`: first year of the period.
 - `lyear`: last year of the period.
- 9. For the option `change`:
 - `ch`: the option `Pctch` specify a change in percentage terms and the option `Absch` specify a change in absolute terms.
 - `fyear_present`: first year of the present period.
 - `lyear_present`: last year of the present period.
 - `fyear_future`: first year of the future period.
 - `lyear_future`: last year of the future period.
- 10. The second part of the code will run the `inc_gridcell_regions.do`, `cr_projections.do` and `cr_maps.do` files located in `.\Stata code\Code` with the specified options.
- 11. The running time for one set of option is around 8 minutes.
- 12. The outputs will be placed in the folder `.\Stata code\Results`. The naming convention for the output file is: output type, independent variable name, irrigation status, climate model, crop, GGCM, specification, CO₂ fertilization, regional delineation, and file type. For example:



5. Conclusions

This paper describes an update to the tool developed by Blanc (2017b) which allows users to calculate emulated rainfed crop yields projections for four crops and five different CGCMs at the regional level. This updated tool, allows users to also estimate irrigated crop yields and corresponding irrigation water requirements under user-given climate change scenarios. As with the previous version, users must be careful when selecting scenarios and ensure that they are within the range of the climate change scenarios used to estimate the response functions in Blanc (2017) and Blanc (2020).

Acknowledgements

We acknowledge the modeling groups and the ISI-MIP coordination team for their roles in producing, coordinating, and making available the ISI-MIP model output. We gratefully acknowledge the financial support for this work from the U.S. Department of Energy, Office of Science under DE-FG02-94ER61937, and other government, industry, and foundation sponsors of the Joint Program on the Science and Policy of Global Change. For a complete list of sponsors, please visit <http://globalchange.mit.edu/sponsors/all>.

References

- Aguiar, A., Narayanan, B., & McDougall, R. (2016). An Overview of the GTAP 9 Data Base. *Journal of Global Economic Analysis*, 1(1), 181–208. <https://doi.org/10.21642/JGEA.010103AF>
- Blanc, É. (2017a). Statistical Emulators of Maize, Rice, Soybean and Wheat Yields from Global Gridded Crop Models. *Agricultural and Forest Meteorology*, 236, 145–161. <https://doi.org/10.1016/j.agrformet.2016.12.022>
- Blanc, É. (2017b). Aggregation of Gridded Emulated Rainfed Crop Yield Projections at the National or Regional Level. *Journal of Global Economic Analysis*, 2(2), 112–127. <https://doi.org/10.21642/JGEA.020203AF>
- Blanc, É. (2020). Statistical emulators of irrigated crop yields and irrigation water requirements. *Agricultural and Forest Meteorology*, 284, 107828. <https://doi.org/10.1016/j.agrformet.2019.107828>
- Blanc, É., & Sultan, B. (2015). Emulating maize yields from global gridded crop models using statistical estimates. *Agricultural and Forest Meteorology*, 214–215, 134–147. <https://doi.org/http://dx.doi.org/10.1016/j.agrformet.2015.08.256>
- Bondeau, A., Smith, P. C., Zaehle, S. S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., MÜLLer, C., Reichstein, M., & Smith, B. (2007). Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, 13(3), 679–706. <https://doi.org/10.1111/j.1365-2486.2006.01305.x>
- Chen, Y-H Henry, Paltsev, S., Reilly, J., Morris, J., & Babiker, Mustafa, H. (2015). The MIT EPPA6 Model: Economic Growth, Energy Use, Emissions, and Food Consumptions. In *Joint Program Report Series (Issue 278)*. <https://globalchange.mit.edu/publication/16262>
- Chen, Y.-H. Henry, Paltsev, S., Reilly, J., Morris, J., Karplus, V., Gurgel, A., Winchester, N., Kishimoto, P., Blanc, É., & Babiker, Mustafa, H. (2017). The MIT Economic Projection and Policy Analysis (EPPA) Model: Version 5. In *Joint Program Technical Note 16*. <https://globalchange.mit.edu/publication/16620>
- Deryng, D., Sacks, W. J., Barford, C. C., & Ramankutty, N. (2011). Simulating the effects of climate and agricultural management practices on global crop yield.

- Global Biogeochemical Cycles, 25(2), n/a-n/a.
<https://doi.org/10.1029/2009GB003765>
- Elliott, J., Glotter, M., Best, N., Kelly, D., Wilde, M., & Foster, I. (2013). The parallel system for integrating impact models and sectors (pSIMS). Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery (XSEDE '13), 21, 1–8.
- Hempel, S., Frieler, K., Warszawski, L., Schewe, J., & Piontek, F. (2013). A trend-preserving bias correction – the ISI-MIP approach. *Earth System Dynamics*, 4(2), 219–236.
- Jones, J., Hoogenboom, G., Porter, C., Boote, K. J., Batchelor, W., Hunt, L., & Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18(3–4), 235–265.
- Lindeskog, M., Arneeth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., & Smith, B. (2013). Implications of accounting for land use in simulations of ecosystem services and carbon cycling in Africa. *Earth System Dynamics Discussions*, 4, 235–278.
- Liu, J., Williams, J. R., Zehnder, A. J. B., & Yang, H. (2007). GEPIC – modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricultural Systems*, 94(2), 478–493.
<https://doi.org/http://dx.doi.org/10.1016/j.agsy.2006.11.019>
- Portmann, F. T., Siebert, S., & Döll, P. (2010). MIRCA2000 - Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, 24.
- 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., Baigorría, G., & Winter, J. M. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170(0), 166–182.
<https://doi.org/http://dx.doi.org/10.1016/j.agrformet.2012.09.011>
- Smith, B., Prentice, I. C., & Sykes, M. T. (2001). Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. *Global Ecology and Biogeography*, 10(6), 621–637. <https://doi.org/10.1046/j.1466-822X.2001.t01-1-00256.x>
- Strzepek, K., Schlosser, C. A., Gueneau, A., Gao, X., Blanc, É., Fant, C., Rasheed, B., & Jacoby, H. D. (2013). Modeling water resource systems within the framework of the MIT Integrated Global System Model: IGSM-WRS. *Journal of Advances in Modeling Earth Systems*, 5(1), 1–16.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4), 485.
<https://doi.org/10.1175/BAMS-D-11-00094.1>

- Waha, K., van Bussel, L. G. J., Müller, C., & Bondeau, A. (2012). Climate-driven simulation of global crop sowing dates. *Global Ecology and Biogeography*, 21(2), 247–259. <https://doi.org/10.1111/j.1466-8238.2011.00678.x>
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences*, 111(9), 3228–3232. <https://doi.org/10.1073/pnas.1312330110>
- Williams, J. R., & Singh, V. P. (1995). Chapter 25. The EPIC. In *Computer Models of Watershed Hydrology* (pp. 909–1000). Water Resources Publications.