

Aggregation of Gridded Emulated Rainfed Crop Yield Projections at the National or Regional Level

BY ELODIE BLANC^a

To estimate the impact of climate change on yields, researchers traditionally use process-based models or statistical models. To benefit from the capabilities of processed-based models while preserving the application simplicity of statistical models, Blanc and Sultan (2015) and Blanc (2017) provide an ensemble of statistical tools emulating crops yields from global gridded crop models at the grid cell level using a simple set of environmental variables. This paper and companion code provide a tool for researcher to use those statistical emulators and estimate crop yields of rainfed maize, rice, soybean and wheat at the regional level. Crop yields estimates for various regional delineations can then simply be used as input into a variety of numerical equilibrium models and other analyses.

JEL codes: Q19, Q54.

Keywords: Crop Yields; Crop Model; Statistical Model; Climate Change.

1. Introduction

The vulnerability of crops to environmental conditions is well known and numerous studies have attempted to estimate the impact of climate change on yields (Challinor et al. 2014). These studies generally rely on either process-based crop models (e.g., Rosenzweig and Parry 1994; Parry et al. 1999; Deryng et al. 2014), which simulate rainfed yield (no irrigation) or irrigated yield (optimal yield under perfect irrigation), or statistical techniques (e.g. Blanc 2012; Blanc and Strobl 2013; Lobell and Field 2007; Sue Wing et al. 2015). While process-based models are able to capture the effect of weather and other environmental conditions, they are computationally demanding and sometimes proprietary, which limits their accessibility. On the other hand, statistical models are more easily applicable but depend on the availability of observations to estimate the impact of average weather conditions on crop yields while controlling for other factors. To benefit from the capabilities of processed-based models while preserving the application

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

simplicity of statistical models, Blanc and Sultan (2015) and Blanc (2017) provide an ensemble of statistical tools emulating crops yields from global gridded crop models (GGCM) at the grid cell level (at a resolution of $0.5^{\circ} \times 0.5^{\circ}$ or roughly $50 \times 50 \text{ km}$) using a simple set of environmental variables.

These crop model emulators are based on GGCM simulations from the ISI-MIP Fast Track experiment (Warszawski et al. 2014; Rosenzweig et al. 2013) driven by climate simulations from the Coupled Model Intercomparison Project, phase 5 (CMIP5) archive (Hempel et al. 2013; Taylor, Stouffer, and Meehl 2012). To statistically estimate the determinants of crop yields, Blanc and Sultan (2015) and consider a parsimonious specification that only includes monthly precipitation, temperature and annual CO_2 concentrations. Among various representations of environmental effects on crop growth, this set of variables was found to provide the best compromise in term of predictive ability and simplicity. Additionally, as the weather effect on crops is expected to differ across soil types, the preferred estimation strategy estimates separate weather response functions for each soil order. Validation exercises by Blanc and Sultan (2015) and Blanc (2017) showed that, in general, the emulator reproduces relatively well the temporal and spatial patterns of climate change impacts on crop yields projected by GGCMs. Areas of disagreement regarding the sign of climate change impact on yields are limited and generally observed in areas where the projected yield impact is close to zero.

As these crop yield emulators provide annual crop yield estimates at the grid cell level only, they require further processing to obtain regional estimates. This paper and companion code¹ provide a tool for researchers to obtain crop yields estimates of rainfed maize, rice, soybean and wheat at the regional level from the statistical emulators. Such aggregation tools have already been made available to the research community on GeoHub (mygeohub.org) for other variables, such as climate data from general circulation models using the Climate Scenario Aggregator tool (Villoria et al. 2015) and even crop yield outputs from the GGCMs considered by the emulator using the AgMIP tool (Villoria et al. 2014). When interested in obtaining crop yield projections for one of the CMIP5 climate change scenario (i.e. HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M, and NorESM1-M climate models), the AgMIP tool would be the most appropriate. However, the crop yield emulator aggregation tool would be necessary when considering alternative plausible user-defined climate change scenarios. Crop yields estimates for various regional delineations can then simply be used as input into a variety of numerical equilibrium models and other analyses.

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

2. Statistical emulator characteristics

These emulators provide response functions for rainfed maize, rice, soybean, and wheat yields while accounting for the structural uncertainty of five different GGCMS: the Geographic Information System (GIS)-based Environmental Policy Integrated Climate (GEPIC) model (Liu et al. 2007; Williams and VP 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, Prentice, and Sykes 2001), the parallel Decision Support System for Agro-technology Transfer (pDSSAT) model (Elliott et al. 2013; Jones et al. 2003), and the Predicting Ecosystem Goods And Services Using Scenarios (PEGASUS) model (Deryng et al. 2011). For each of these GGCMS, model simulations consider the effect of CO₂ concentrations to account for the CO₂ fertilization effect, and assume no irrigation to capture the effect of precipitation on crop yields.

The response functions considered in the aggregation tool correspond to the preferred specification in Blanc (2017), S1fpintsoil, which provides a flexible fractional polynomial specification of the effect of weather on crop yields and account for parameter heterogeneity across soil types. The regression results showed that precipitation and temperature during all the months of the growing seasons and annual CO₂ have a significant non-linear effect on crop yields from all GGCMS. In general, temperature and precipitation curves are concave and skewed toward low values, especially for low precipitation. Examples of response functions of temperature, precipitation and CO₂ effects on maize yields for the LPJmL model estimated using the S1fpint specification for the Mollisol soil type subsample are provided in Figure 1. More details regarding the estimation of the response functions can be found in Blanc (2017).

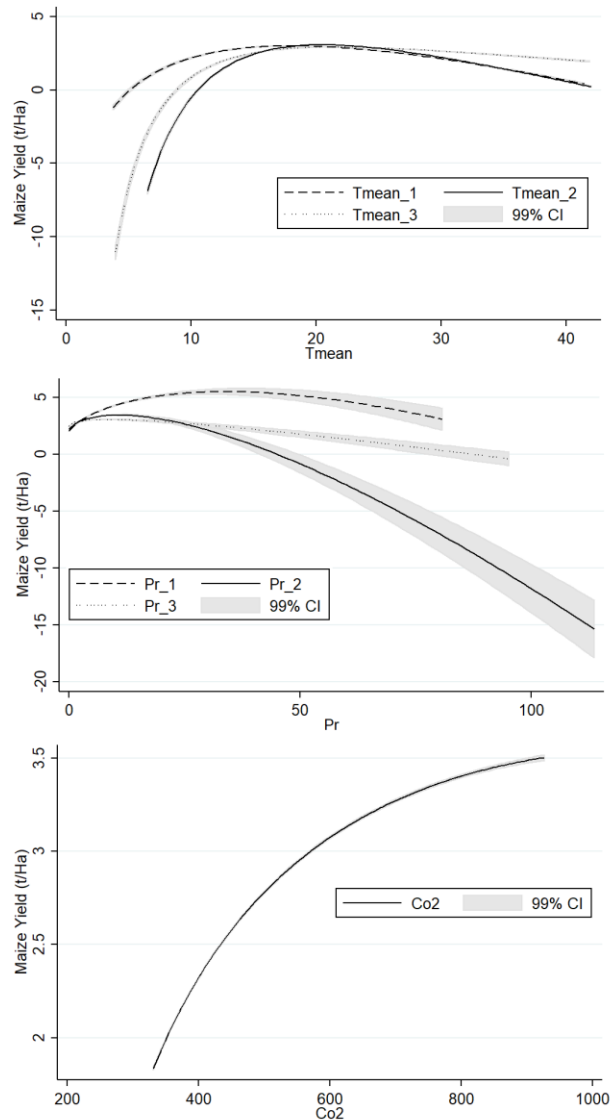


Figure 1. Examples of response functions of temperature, precipitation and CO₂ effects on maize yields for the LPJmL model estimated using the S1fpint specification for the Mollisol soil type subsample

Notes: The variables Tmean, Pr, and Co2 represent monthly mean temperature, precipitation and presents annual CO₂ concentration respectively; The terms _1, _2 and _3 refer to the first, second and third month of summer respectively.

All GGCMS provide estimates of actual annual crop yields, except for the LPJ-GUESS model, which simulates potential yields (yield non-limited by nutrient or management constraints). Blanc and Sultan (2015) and Blanc (2017) showed that the statistical emulators are overall able to replicate reasonably well the spatial

patterns of yields crop projected by crop models in levels but also in term of changes overtime. However, as noted in Blanc (2017), “due to GGCM specificities, simulations are more suited to assess long-term trends in yields rather than inter-annual yield variability”. Users of the aggregation tool should therefore ensure that it is used to simulate changes in crop yields between multi-year periods (e.g. one decade to another). When considering the ensemble of GGCMs, the user should also consider crop yield changes in percentage terms rather than in levels to account for the discrepancy between actual and potential yields.

3. Regional aggregation of gridded crop yields

To aggregate simulated gridded crop yields at the regional or national level, information on crop-specific harvested areas is required. Considering four different land use datasets, Porwollik et al. (2017) find that the choice of land use dataset has an effect on the mean and temporal dynamics of aggregated gridded crop yields. However, as noted by the authors, none of the four dataset is superior to the others, we use the MIRCA2000 (Portmann, Siebert, and Döll 2010) data of harvested area of each rainfed crop to calculate the area cultivated for each grid cell. Grid cells are assigned to each region of interest, and where grid cells overlap different regions, the grid cell is assigned to the region having the largest share of area within that grid cell.

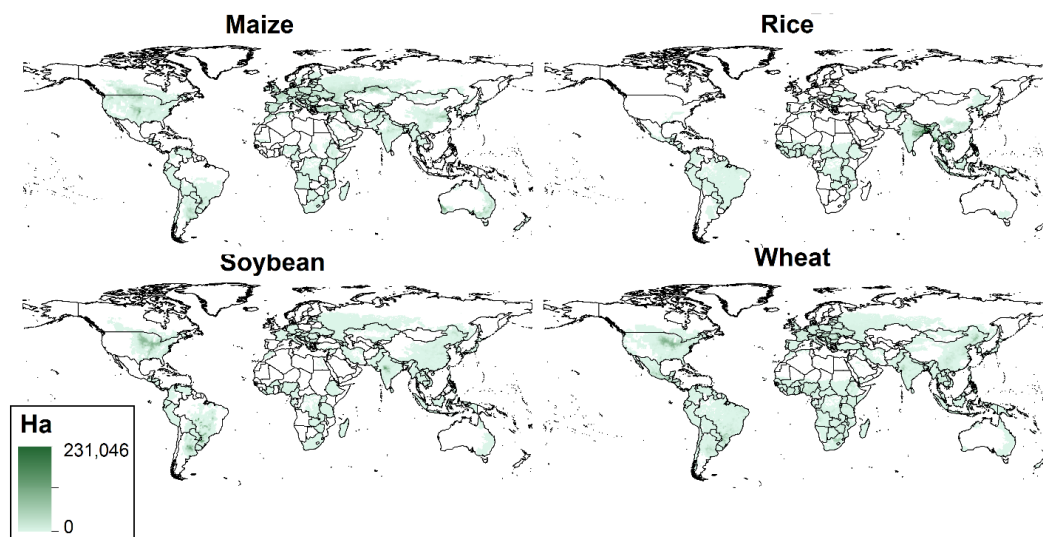


Figure 2. MIRCA2000 harvested rainfed crop areas

The regional aggregation of gridded crop yields follows the equation:

$$\overline{Yield}_{c,r,y,ggcm} = \frac{\sum_{g=1}^G Yield_{c,g,y,ggcm} * Area_{c,g}}{\sum_{g=1}^G Area_{c,g}} \quad (1)$$

where for each crop, c , region, r , year, y , and GGCM, $ggcm$, the average yield, $\overline{Yield}_{c,r,y,ggcm}$ (in t/ha), is given by first multiplying yield estimates, $Yield$, at the grid cell level, g , by the corresponding rainfed harvested area, $Area$, 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 harvested area within the region.

4. Processing tool folder structure

This paper provides a code (included in the supplementary materials published with this article) to enable one to estimate crop yields at the regional level using the response functions of the emulator under given climate data inputs. The program is written in Stata 14 and is part of the folder \Stata code. In this folder are also four subfolders: (i) \Data, containing weather and land use inputs; (ii) \Parameters, containing response function parameters; (iii) \Shapefile, containing shapefiles for output regions delineation; and (iv) \Results, containing the outputs of the program. Each components is described below.

4.1. Data

The \Stata code\Data folder is composed of the subfolder \Climate, containing weather input data, and \Land use, containing the land use mask (crop growing area).

4.1.1. Climate

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 northern hemisphere. The user must also provide annual CO₂ concentration data.

As an example, the code comes with three climate input data located in \Data\Climate. The climate data correspond to those used to estimate the response functions in Blanc (2017).

4.1.2. Land use

The \Land use subfolder contains the MIRCA2000 data of harvested area of each rainfed crop at the grid cell level necessary to average crop yields over regions following the methodology described in Section 2.

4.2. Parameters

The \Stata code\Parameters folder is composed of three subfolders. The subfolder \Estimates contains Stata coefficients estimates of explanatory variables for the S1fpinsoil specification. Estimates are stored in .ster files for each crop, model and soil category. The subfolder \Fixed Effects contains the fixed effects coefficients for each crop, model and soil category. The subfolder \FP contains the file FP_formula.dta which includes all fractional polynomial transformation associated with each variable, for each crop, model and soil category.

4.3. Shapefiles delineating output regions

The program is set up to calculate regional average crop yields across different regions. The data required to average grid-cell level projections at the regional level are provided in the folder \Stata code\Shapefile. Files for each regions are located in corresponding subfolders: world countries are located in \world, gtap9 regions in \gtap9, EPPA6 regions in \EPPA6, and EPPA5 regions in \EPPA5. Each of these subfolders contains a shapefile of the regions of interest as well as a shapefile of these same regions at the 0.5°x0.5°-degree resolution. The centroid coordinates of each of these grid cells and the corresponding country name are extracted in a excel file labeled *_grid.xlsx. Maps representing the different regions are provided in Figures 3 to 6.

To create a new regional delineation (for regional delineations not already provided), follow the instructions provided in the Appendices. The subfolder folder \grid contains a gridded shapefile of the world necessary to create new regional delineations.

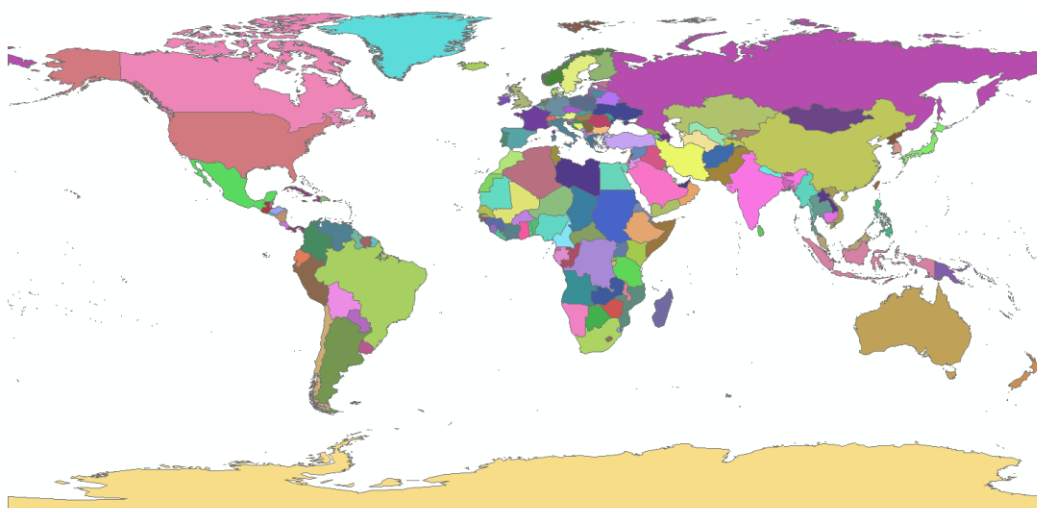


Figure 3. World regions (countries)

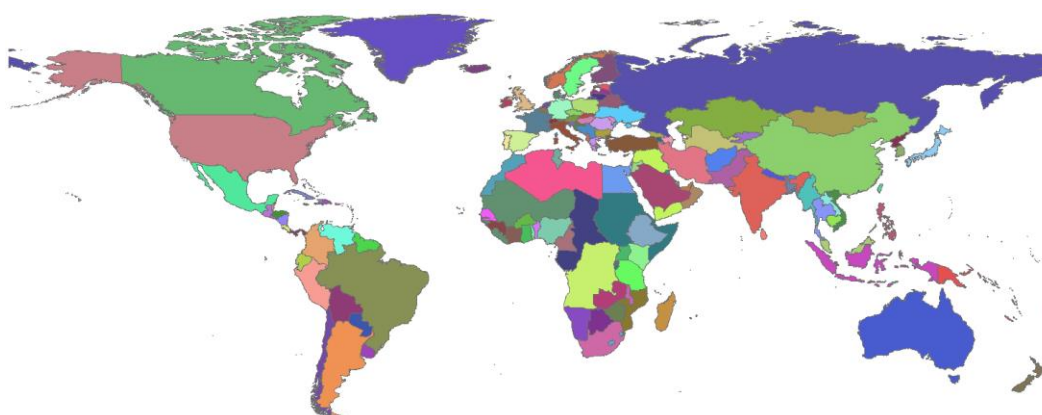


Figure 4. GTAP9 regions

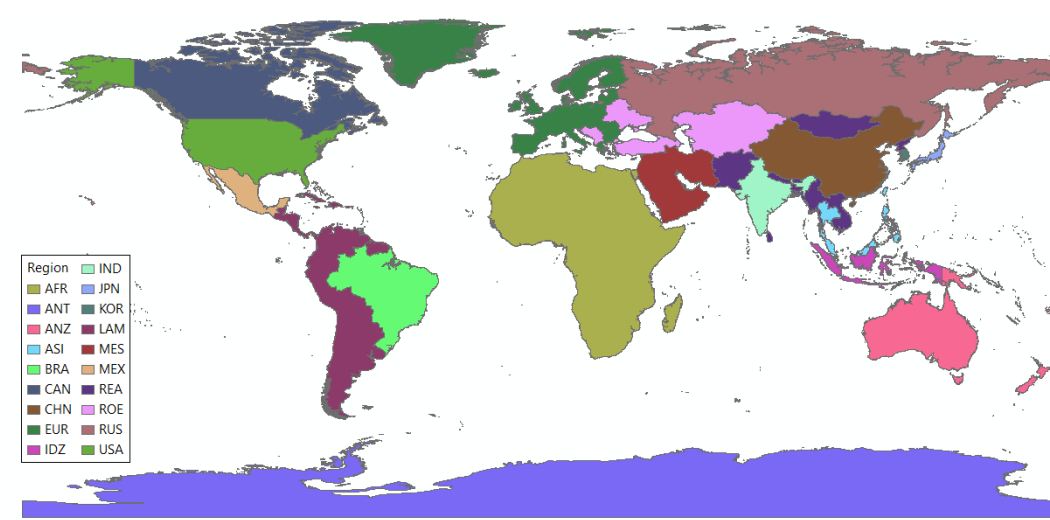


Figure 5. EPPA6 regions

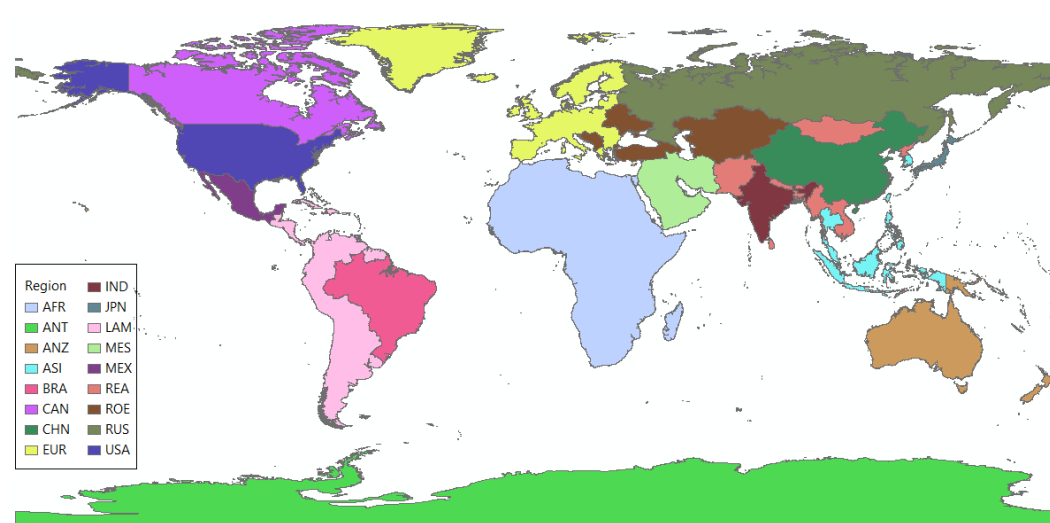


Figure 6. EPPA5 regions

4.4. Results

The program `cr_projections.do` will output regional averages of crop yields in metric tons per hectare (t/ha) for each year in the folder `\Stata code\Results`. Relatedly, the program `cr_maps.do` will output maps for the specified options. The name of the file starting by `Preds_*` and `Map_*` are composed of the options set up by the user at the beginning of the code.

As an example, Figure 7 provides two maps representing changes in maize yields (in percentage terms) between the periods 2001-2010 and 2091-2100 projected using the S1fpint of the LPJmL emulator under the GFDL rcp8p5 climate scenario. The first map represents changes in yields at the EPPA5 regional level, and the second map represents changes in yields at the country level.

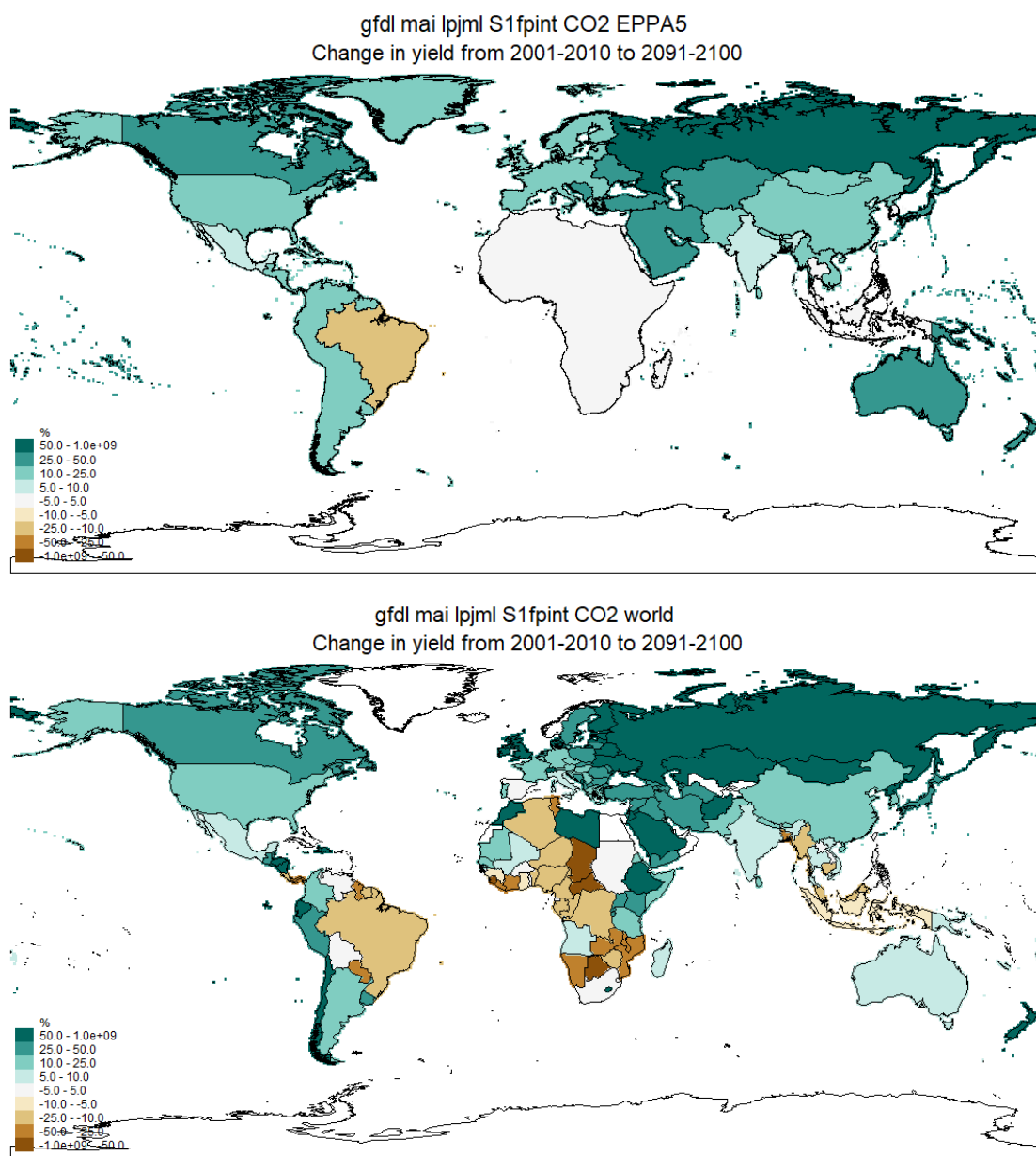


Figure 7. Example of results for two different regional delineation

5. Running instructions

To run the tool, install the 'Stata code' folder in your home directory.

Prepare desired climate data and place them in . \Stata code\Data\Climate. The name of the file should correspond to the name of the climate model (or simulation).

Open Stata and install the functions `wtmean` and `spmap` if not already installed. Type `ssc inst _gwtmean` and `ssc inst spmap`.

Open the do file `cr_projections.do` in \Stata code and amend the options located at the top of the file. For options with name ending with `list` (e.g. `model list`), the user can specify a list of options. For the other variables, only one option can be specified. The options to specify are:

`path`: directory of the \Stata code folder (e.g. `local path="C:\Users\quidam\Stata code"`)

`regions`: name of the region delineation required for aggregation. In the standard setup, four options are available: `gtap9`, `world`, `EPPA6`, and `EPPA5`.

`gcmlist`: list of climate change scenarios. In the standard setup, three predefined scenarios are available: `gfdl`, `hadgem2`, and `noresm1`.

`croplist`: list of crops to consider among maize (`mai`), rice (`ric`), soybean (`soy`) and wheat (`whe`)

`modellist`: list of GGCMS to consider among `LPJmL` (`lpjml`), `LPJ-GUESS` (`lpj-guess`), `PEGASUS` (`pegasus`) and `pDSSAT` (`pdssat`).

`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.

Model specific options:

`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.

`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.

Run the file `cr_projections.do` file. The outputs will be placed in the folder \Stata code\Results.

To create maps of the outputs, open the Stata do file `cr_maps.do`. Specify the options `path`, `regions`, `gcmlist`, `croplist`, `modellist` and `co2` as instructed above. Specify the map options:

`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.

For the option level:

fyear: first year of the period

lyear: last year of the period

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

Run the `cr_maps.do` file. The outputs will be placed in the folder `\Stata code\Results`.

6. Conclusions

The current program allows users to easily obtain emulated rainfed crop yields projections for four crops and five different CGCMs at the regional level. This program is designed to be run with user-given climate change scenarios. However, users must be careful to consider scenarios that are within the range of the climate change scenarios used to estimate the response functions in Blanc (2017).

In further developments (depending on the publication of the underlying studies), the program will be updated to included irrigated crops and a larger number of crops. Possible extensions also include emulations of irrigation water requirements.

Acknowledgements

We thank two anonymous referees and Anaïs Goburdhun for helpful comments and suggestions. We acknowledge the modeling groups (listed in Appendix A, Table A1 of this paper) 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

Blanc, Élodie. 2012. "The Impact of Climate Change on Crop Yields in Sub-Saharan Africa." *American Journal of Climate Change* 1 (1):1-13.

- — —. 2017. "Statistical Emulators of Maize, Rice, Soybean and Wheat Yields from Global Gridded Crop Models." *Agricultural and Forest Meteorology* 236:145–61.
- Blanc, Élodie, and Eric Strobl. 2013. "The Impact of Climate Change on Cropland Productivity: Evidence from Satellite Based Products at the River Basin Scale in Africa." *Climatic Change* 117 (4):873–90.
- Blanc, Élodie, and Benjamin Sultan. 2015. "Emulating Maize Yields from Global Gridded Crop Models Using Statistical Estimates." *Agricultural and Forest Meteorology* 214–215:134–47. <http://dx.doi.org/10.1016/j.agrformet.2015.08.256>.
- Bondeau, Alberte, Pascale C Smith, Sonke Zaehle, Sibyll Schaphoff, Wolfgang Lucht, Wolfgang Cramer, Dieter Gerten, et al. 2007. "Modelling the Role of Agriculture for the 20th Century Global Terrestrial Carbon Balance." *Global Change Biology* 13 (3). Blackwell Publishing Ltd:679–706. <https://doi.org/10.1111/j.1365-2486.2006.01305.x>.
- Challinor, A J, J Watson, D B Lobell, S M Howden, D R Smith, and N Chhetri. 2014. "A Meta-Analysis of Crop Yield under Climate Change and Adaptation." *Nature Climate Change* 4 (4):287–91. <http://www.nature.com/nclimate/journal/v4/n4/abs/nclimate2153.html#supplementary-information>.
- Deryng, Delphine, Declan Conway, Navin Ramankutty, Jeff Price, and Rachel Warren. 2014. "Global Crop Yield Response to Extreme Heat Stress under Multiple Climate Change Futures." *Environmental Research Letters* 9 (3):34011. <http://stacks.iop.org/1748-9326/9/i=3/a=034011>.
- Deryng, Delphine, W. J. Sacks, C. C. Barford, and N. Ramankutty. 2011. "Simulating the Effects of Climate and Agricultural Management Practices on Global Crop Yield." *Global Biogeochemical Cycles* 25 (2). <https://doi.org/10.1029/2009GB003765>.
- Elliott, J, M Glotter, N Best, D Kelly, M Wilde, and I Foster. 2013. "The Parallel System for Integrating Impact Models and Sectors (pSIMS)." In *Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery (XSEDE '13)*, 21:1–8. Association for Computing Machinery.
- Hempel, S, K Frieler, L Warszawski, J Schewe, and F Piontek. 2013. "A Trend-Preserving Bias Correction & the ISI-MIP Approach." *Earth System Dynamics* 4 (2):219–36.
- Jones, J, G Hoogenboom, C Porter, K Boote, W Batchelor, L Hunt, and J Ritchie. 2003. "The DSSAT Cropping System Model." *European Journal of Agronomy* 18 (3–4):235–65.
- Lindeskog, M, A Arneth, A Bondeau, K Waha, J Seaquist, S Olin, and B Smith. 2013. "Implications of Accounting for Land Use in Simulations of Ecosystem Services and Carbon Cycling in Africa." *Earth System Dynamics Discussions* 4:235–78.

- Liu, Junguo, Jimmy R Williams, Alexander J B Zehnder, and Hong Yang. 2007. "GEPIC - Modelling Wheat Yield and Crop Water Productivity with High Resolution on a Global Scale." *Agricultural Systems* 94 (2):478-93. <http://dx.doi.org/10.1016/j.agsy.2006.11.019>.
- Lobell, David, and Christopher Field. 2007. "Global Scale Climate-Crop Yield Relationships and the Impacts of Recent Warming." *Environmental Research Letters* 2 (1):1-7.
- Parry, M, G Fischer, M Livermore, C Rosenzweig, and A Iglesias. 1999. "Climate Change and World Food Security: A New Assessment." *Global Environmental Change* 9:551-67.
- Portmann, F T, S Siebert, and P Döll. 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.
- Porwollik, Vera, Christoph Müller, Joshua Elliott, James Chryssanthacopoulos, Toshichika Iizumi, Deepak K. Ray, Alex C. Ruane, et al. 2017. "Spatial and Temporal Uncertainty of Crop Yield Aggregations." *European Journal of Agronomy* 88 (August):10-21. <https://doi.org/10.1016/j.eja.2016.08.006>.
- Rosenzweig, C, J W Jones, J L Hatfield, A C Ruane, K J Boote, P Thorburn, J M Antle, et al. 2013. "The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and Pilot Studies." *Agricultural and Forest Meteorology* 170 (0):166-82. <http://dx.doi.org/10.1016/j.agrformet.2012.09.011>.
- Rosenzweig, C, and M L Parry. 1994. "Potential Impacts of Climate Change on World Food Supply." *Nature* 367:133-38.
- Smith, Benjamin, I Colin Prentice, and Martin T Sykes. 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). Blackwell Science Ltd:621-37. <https://doi.org/10.1046/j.1466-822X.2001.t01-1-00256.x>.
- Sue Wing, Ian, Erwan Monier, Ari Stern, and Anupriya Mundra. 2015. "US Major Crops' Uncertain Climate Change Risks and Greenhouse Gas Mitigation Benefits." *Environmental Research Letters* 10 (11). <https://doi.org/10.1088/1748-9326/10/11/115002>.
- Taylor, Karl E., Ronald J. Stouffer, and Gerald A. Meehl. 2012. "An Overview of CMIP5 and the Experiment Design." *Bulletin of the American Meteorological Society* 93 (4):485-98. <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Villoria, Nelson Benjamin, Joshua Elliott, Christoph Müller, Jaewoo Shin, and Lan Zhao. 2015. "Climate Scenario Aggregator." <https://mygeohub.org/resources/1005>.
- Villoria, Nelson Benjamin, Joshua Elliott, Christoph Müller, Jaewoo Shin, Lan Zhao, and C Song. 2014. "Rapid Aggregation of Globally Gridded Crop Model

Outputs to Facilitate Cross-Disciplinary Analysis of Climate Change Impacts in Agriculture." <https://mygeohub.org/tools/agmip/>.

Waha, K, L G J van Bussel, C Müller, and A Bondeau. 2012. "Climate-Driven Simulation of Global Crop Sowing Dates." *Global Ecology and Biogeography* 21 (2). Blackwell Publishing Ltd:247–59. <https://doi.org/10.1111/j.1466-8238.2011.00678.x>.

Warszawski, Lila, Katja Frieler, Veronika Huber, Franziska Piontek, Olivia Serdeczny, and Jacob Schewe. 2014. "The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project Framework." *Proceedings of the National Academy of Sciences* 111 (9):3228–32. <https://doi.org/10.1073/pnas.1312330110>.

Williams, J R, and Singh VP. 1995. "Chapter 25. The EPIC." In *Computer Models of Watershed Hydrology*, 909–1000. Littleton, CO: Water Resources Publications.

Appendix. Create new regional delineation

To create a new regional delineation (for regional delineations not already provided), follow the instructions below (the example are provided to create the regional delineation for the world countries, which is already included in the folder):

Create new folder and name it by the name of the region (e.g. world)

In ArcGIS:

install spatial join - largest overlap tool from <http://www.arcgis.com/home/item.html?id=e9cccd343bf84916bda1910c31e5eab2>

open regional shapefile (e.g. world.shp)

open . \Stata code\Shapefile\grid\0_5_world_grid.shp

run the 'spatial join - largest overlap' tool with the options: Target Feature = 0_5_world_grid; Join Feature = world.shp; Output Feature Class = \Stata code\Shapefile\world\world_grid.shp; and uncheck the option "keep all"

In Excel:

open world_grid.dbf

save as world_grid.xlsx

in Stata:

note the name of the variable in the shapefile corresponding to the region (e.g. in the world.shp shapefile, the name of each region (i.e. country) is provided by the variable called 'NAME').

open the do file inc_gridcell_regions.do

specify the name of the variable (e.g. include the line: if "`regions'"=="world" local regname="NAME")