The Journal of Open Source Software

gaia: An R package to estimate crop yield responses to temperature and precipitation

Mengqi Zhao [1], **Stephanie T. Morris** [2], **Claudia Tebaldi** [2], **and Abigail Snyder** [2]

1 Pacific Northwest National Laboratory, Richland, WA, USA **2** Pacific Northwest National Laboratory, Joint Global Change Research Institute, College Park, MD, USA

DOI: 10.21105/joss.07538

Software

- Review
- Repository
- Archive

Editor: Taher Chegini

Reviewers:

- @klau506
- @jrising

Submitted: 14 October 2024
Published: 16 July 2025

Summary

gaia is an open-source R package designed to estimate crop yield shocks in response to annual weather variations and CO_2 concentrations at the country scale for 17 major crops. This innovative tool streamlines the workflow from raw climate data processing to projections of annual shocks to crop yields at the country level, using the response surfaces from an empirical econometric model developed and documented in Waldhoff et al. (2020), which leverages historical weather, CO_2, and crop yield data for robust empirical fitting for 17 crops. gaia uses these response surfaces with monthly temperature and precipitation projections (e.g., from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (O'Neill et al., 2016) climate data bias-adjusted and statistically downscaled by the ISIMIP3BASD approach (Lange, 2019) in the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) (Warszawski et al., 2014)) to project yield shocks that can be applied to agricultural productivity changes at the country level for use in multisectoral economic models. The historical and future projections use gridded, country-and-crop specific monthly growing season precipitation and temperature data, aggregated to the national level, and weighted by cropland area derived from the global Monthly Irrigated and Rainfed Crop Areas around the year 2000 (MIRCA2000) dataset (Portmann et al., 2010). These annual, country, and crop-specific yield shocks can be aggregated to different definitions of regions, crop commodities, and time periods, as needed by specific multisectoral economic models. gaia serves as a lightweight, powerful tool that can aid exploration of crop yield responses under a broad range of future climate projections, enhancing human-Earth system analysis capabilities.

Statement of need

Agricultural production is highly responsive to annual weather patterns, which are becoming increasingly variable in response to shifts in climate conditions (e.g. Iizumi & Ramankutty (2016)). Agricultural markets are global, and crop yield shocks in one region will therefore impact others due to the variability in international trade. In addition, agricultural production is closely connected to energy and water sectors. Improving the representation of agricultural production and its yield responses to changing weather patterns in global multi-sector dynamics models, such as the Global Change Analysis Model (GCAM)(Calvin et al., 2019; Joint Global Change Research Institute, 2023), is key to better understanding the economic impacts of climate variations that account for human-Earth system interactions.

Most multisectoral economic modeling of climate impacts on agriculture addresses only the long-term trends in temperature and precipitation effects on crop yields for a few key crops (Lampe et al., 2014). There is a need to include projections of inter-annual weather impacts on crop yields for multiple crops under a wide range of climate projections (Ray et al., 2015).

However, the current crop modeling landscape lacks globally comprehensive, computationally efficient, and validated models capable of meeting the aforementioned needs (Waldhoff et al., 2020). To address this gap, we have developed `gaia`, which uses the empirical model described in Waldhoff et al. (2020). `gaia` produces projections of both future annual variations and long-term trends in crop yield responses at the country level for a wide range of crops (barley, cassava, cotton, ground nuts, maize, millet, rape seed, rice, rye, potatoes, pulses, sorghum, soybean, sugar beet, sugarcane, sunflower, and wheat) using gridded monthly temperature and precipitation inputs for any future climate scenario.

State of the field

The exploration of the effects of future weather conditions on agricultural production is crucial for providing insights into global food, energy, water, and economic development. Various crop models, including process-based and empirical models, have been developed to simulate crop yields under different climate scenarios (Chapagain et al., 2022; Rauff & Bello, 2015). Müller et al. (2017) conducted a Global Gridded Crop Model (GGCM) intercomparison experiment involving 14 process-based crop models, which allow users to simulate diverse crop management options, soil types, and weather conditions. However, GGCMs require large amounts of site-specific data for calibration, making extensions to many crops and regions more challenging (Müller et al., 2017). In addition, these models can take significant time to run, and the yield projections are often available for only a limited number of climate scenarios.

These considerations may limit the usefulness of crop yield shocks from process models as inputs to global economic, multisector models, which include a wider range of crops and many future climate scenarios. `gaia` achieves this balance by employing computationally efficient statistical methods to model 17 major crops globally, while maintaining the sensitivity to capture both gradual shifts in climate and interannual variability impacts on crop yields. In addition, users can estimate yield shocks for additional crops by providing the appropriate input data.

Functionality

The primary functionality of `gaia` is encapsulated in the `yield_impact` wrapper function, which streamlines the entire workflow shown in Figure 1. The modular design also facilitates comprehensive diagnostic outputs, enhancing the tool's utility for researchers and decision makers. Users can also execute individual functions to work through the main steps of the process. Detailed instructions on installing and using `gaia` with an example dataset can be accessed at https://jgcri.github.io/gaia.

1. `weighted_climate`: Processes CMIP6 daily or monthly climate NetCDF data, formatted in accordance with the ISIMIP simulation protocols and calculates cropland-weighted precipitation and temperature at the country level, differentiated by crop type and irrigation type.
2. `crop_calendars`: Generates crop planting months for each country and crop, based on crop calendar data (Sacks et al., 2010).
3. `data_aggregation`: Calculates crop growing seasons using climate variables processed by `weighted_climate` and crop calendars for both historical and projected periods. This function prepares climate and yield data for subsequent model fitting.
4. `yield_regression`: Performs regression analysis fitted with historical annual crop yields, growing season monthly temperature and precipitation, CO_2 concentrations, and GDP per capita. The default econometric model applied in `gaia` is based on Waldhoff et al. (2020). Users can specify alternative formulas that are consistent with the data processed in `data_aggregation`.
5. `yield_shock_projection`: Projects country-level yield shocks for future climate using

The Journal of Open Source Software

the fitted model obtained from `yield_regression`, temperature, precipitation, and CO_2 projections. The outputs can be integrated into non-GCAM modeling frameworks requiring country-level yield responses to climate.

6. `gcam_agprodchange`: Generates GCAM-compatible outputs for agricultural productivity changes by GCAM commodity at the region-basin level. The function remaps country-level yield shocks from `yield_shock_projection` to the spatial scales required by GCAM (i.e., by region-basin-technology intersections) using harvested areas, and aggregates crops into GCAM commodities. This function applies these projected shocks to GCAM scenario-specific agricultural productivity growth rates (the metric used to project future yields in GCAM), and creates ready-to-use XML outputs for GCAM.

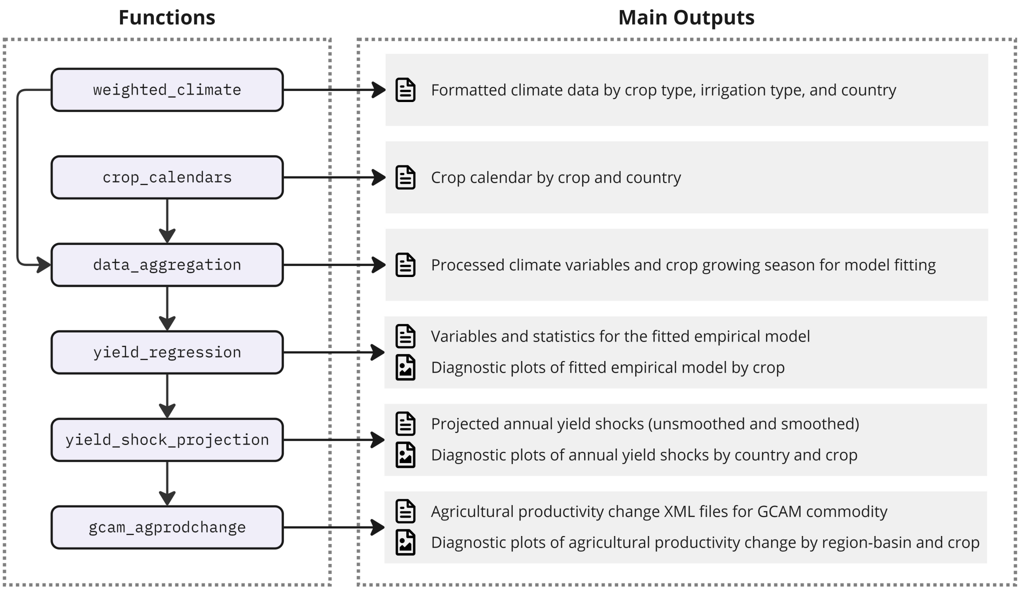

Figure 1: The gaia workflow showing the functions and the corresponding outputs of modeling crop yield shocks to weather variables using an empirical econometric model.

Acknowledgements

This research was supported by the US Department of Energy, Office of Science, as part of research in MultiSector Dynamics, Earth and Environmental System Modeling Program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. The views and opinions expressed in this paper are those of the authors alone.

References

Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y., Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., Hejazi, M., Horowitz, R., Iyer, G., Kyle, P., Kim, S., Link, R., McJeon, H., Smith, S. J., Snyder, A., … Wise, M. (2019). GCAM v5.1: Representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development*, *12*(2), 677–698. https://doi.org/10.5194/gmd-12-677-2019

Chapagain, R., Remenyi, T. A., Harris, R. M. B., Mohammed, C. L., Huth, N., Wallach, D., Rezaei, E. E., & Ojeda, J. J. (2022). Decomposing crop model uncertainty: A systematic review. *Field Crops Research*, *279*, 108448. https://doi.org/10.1016/j.fcr.2022.108448

Iizumi, T., & Ramankutty, N. (2016). Changes in yield variability of major crops for 1981–2010 explained by climate change. *Environmental Research Letters*, *11*(3), 034003. https://doi.org/10.1088/1748-9326/11/3/034003

Joint Global Change Research Institute. (2023). *GCAM Documentation* (Version 7.0). https://github.com/JGCRI/gcam-doc

Lampe, M. von, Willenbockel, D., Ahammad, H., Blanc, E., Cai, Y., Calvin, K., Fujimori, S., Hasegawa, T., Havlik, P., Heyhoe, E., Kyle, P., Lotze-Campen, H., Mason d'Croz, D., Nelson, G. C., Sands, R. D., Schmitz, C., Tabeau, A., Valin, H., Mensbrugghe, D. van der, & Meijl, H. van. (2014). Why do global long-term scenarios for agriculture differ? An overview of the AgMIP global economic model intercomparison. *Agricultural Economics*, *45*(1), 3–20. https://doi.org/10.1111/agec.12086

Lange, S. (2019). Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geoscientific Model Development*, *12*(7), 3055–3070. https://doi.org/10.5194/gmd-12-3055-2019

Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., … Yang, H. (2017). Global gridded crop model evaluation: Benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, *10*(4), 1403–1422. https://doi.org/10.5194/gmd-10-1403-2017

O'Neill, B. C., Tebaldi, C., Vuuren, D. P. van, Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

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*(1). https://doi.org/10.1029/2008GB003435

Rauff, K. O., & Bello, R. (2015). A review of crop growth simulation models as tools for agricultural meteorology. *Agricultural Sciences*, *6*(9), 1098–1105. https://doi.org/10.4236/as.2015.69105

Ray, D. K., Gerber, J. S., MacDonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications*, *6*(1), 5989. https://doi.org/10.1038/ncomms6989

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

Waldhoff, S. T., Wing, I. S., Edmonds, J., Leng, G., & Zhang, X. (2020). Future climate impacts on global agricultural yields over the 21st century. *Environmental Research Letters*, *15*(11), 114010. https://doi.org/10.1088/1748-9326/abadcb

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