The Journal of Open Source Software

AgriFoodPy: a package for modelling food systems

Juan P. Cordero[1¶], **Kevin Donkers**[2], **Ian Harrison**[3], **Sarah L. Bridle**[1], **Angelina Frankowska**[4], **Michelle Cain**[4], **Neil Ward**[5], **Jez Frendenburgh**[5], **Edward Pope**[2], **Alana Kluczkovski**[1], **Ximena Schmidt**[6], **Jacqueline Silva**[7], **Christian Reynolds**[8], **Katherine Denby**[9], **Bob Doherty**[10], and **Aled Jones**[11]

1 Department of Environment and Geography Wentworth Way, University of York, Heslington, York, YO10 5NG, UK **2** Met Office Hadley Centre, Met Office, Fitz Roy Road, Exeter, Devon, UK **3** School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, UK **4** Centre for Environmental and Agricultural Informatics, School of Water, Energy and Environment, Cranfield University, Cranfield MK43 0AL, UK **5** School of Environmental Sciences, University of East Anglia, Norwich, UK **6** Equitable Development and Resilience Research Group, College of Engineering, Design and Physical Science, Brunel University London, London, UB8 3PH, UK **7** Global Academy of Agriculture and Food Systems, Alexander Robertson Building, Easter Bush Campus, Roslin, Midlothian EH25 9RG **8** Centre for Food Policy, City, University of London, Northampton Square, London, EC1V 0HB, UK **9** Centre for Novel Agricultural Products (CNAP), Department of Biology, University of York, York, YO10 5DD, UK **10** School for Business and Society, University of York **11** Global Sustainability Institute, Anglia Ruskin University, Cambridge CB1 1PT, UK ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

`AgriFoodPy` is an open-source Python package for dataset manipulation, interoperability, simulation, and modeling of agrifood systems.

By employing xarray (Hoyer & Hamman, 2017) as the primary data structure, AgriFoodPy provides access methods to manipulate tabular data by extending xarray functionality via accessors.

A separate repository, `agrifoodpy_data`, is also maintained in parallel to provide access to local and global agrifood datasets, including geospatial land use and classification data (Morton, 2022), food supply (FAO, 2022), life cycle assessment (Poore & Nemecek, 2018), and population data (Nations", 2022). The `AgriFoodPy` framework is region-agnostic and provides facilities to model and simulate processes and interventions regardless of their geographic origin. As an example, current functionality can be used to construct predictions on the effect of dietary change on greenhouse emissions, the carbon sequestration of transforming agricultural land use, or consequences of modifying farming practices on operational costs.

Open-source code and community development will allow a transparent view into analysis choices and data sources, which can help provide trustworthy evidence-based support for data-driven policymaking. AgriFoodPy is developed and maintained by a diverse community of domain experts with a focus on software sustainability and interoperability.

Version 0.1 provides basic table manipulation methods to extend the coordinate dimensions of xarray Datasets and DataArrays, extract summary statistics and indicators, plotting methods to analyze and display data. It also includes a library of intervention models such as diet change, afforestation and agroecology and land carbon sequestration. We have also built an interface to external an atmospheric model (Leach et al., 2021). We plan to extend this library of interfaces to incorporate other open source tools, including socioeconomic, land use, environmental, health, and policy modelling packages.

Future releases will provide access to more models and community-contributed datasets formatted using xarray. Additionally, AgriFoodPy will implement a pipeline manager to perform end-to-end simulations of agrifood systems, which can be used to speed up the comparison of multiple scenarios.

Statement of need

Providing food for an ever-growing population while reducing the impact of human activity on the environment has become one of the main global challenges. Local and intergovernmental independent committees (https://www.theccc.org.uk/, https://www.ipcc.ch/) have reported the importance of food production on climate change. The scenarios and projections in their reports also highlight the need for precise and transparent modeling of different aspects of the food system to help stakeholders understand the effects of consumption patterns and farming practices.

Coordinated efforts to achieve a sustainable food system must originate from effective policy-making based on evidence, careful choice of metrics and indicators to describe the state of the food system, and accurate estimates of how these metrics change under different scenarios and decisions/interventions.

Existing datasets and analysis software usually rely on non-standardized data structures and predominantly closed-source code. This hinders research and independent scrutiny of food system intervention projections and the impact of policy on environmental, socio-economic, and health indicators. Moreover, this forces researchers to routinely expend significant effort replicating or re-developing existing code to reduce and analyze data. Additionally, the opacity of some data sources and analysis choices makes it difficult to draw conclusions from equivalent comparisons between different interventions and policy decisions.

Few open initiatives exist focused on analysis and modelling of agrifood and enviromental related data, e.g., The Environmental Data Science book (https://edsbook.org/). The research community has developed open-source packages that adress some individual aspects of modelling agrifood systems, such as geospatial imaging (e.g., `Geopandas` (Jordahl et al., 2020), `Rasterio` (Gillies & others, 2013--)), atmospheric and climate modeling (`Fair` (Leach et al., 2021)) in Python, and other open softwares in other languages, such as agriculture and farming (`APSIM` (Holzworth et al., 2014)) and life cycle assestment (`OpenLCA`, www.openlca.org).

AgriFoodPy provides a consistent standard for agrifood data distribution, while also allowing external models and packages to coexist and interoperate, allowing a holistic approach to agrifood modeling.

Plans for future use in research and communication include the FixOurFood agrifood calculator (https://fixourfood.streamlit.app/), an interactive modelling tool to evaluate the effect of food system transformations in the UK. There are also plans to publish a paper on global diets their social and environmental impacts.

Acknowledgements

JPC, SLB, AF, MC, KD, BD, AK and FixOurFood's work was supported by the UKRI Transforming Food Systems Strategic Priority Fund (grant number BB/V004581/1). This article has benefited from comments and suggestions of the following people: Daniel Lewis, Joe Kennedy, Nada Saidi, Adam Amara.

References

FAO. (2022). *FAO. FAOSTAT. License: CC BY-NC-SA 3.0 IGO. Extracted from: Https://www.fao.org/faostat/en/#data/FBS. Date of access: 22-06-2023*.

Gillies, S., & others. (2013--). *Rasterio: Geospatial raster i/o for Python programmers*. Mapbox. https://github.com/rasterio/rasterio

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P. M., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., … Keating, B. A. (2014). APSIM – evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, *62*, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009

Hoyer, S., & Hamman, J. (2017). Xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, *5*(1). https://doi.org/10.5334/jors.148

Jordahl, K., Bossche, J. V. den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020). *Geopandas/geopandas: v0.8.1* (Version v0.8.1). Zenodo. https://doi.org/10.5281/zenodo.3946761

Leach, N. J., Jenkins, S., Nicholls, Z., Smith, C. J., Lynch, J., Cain, M., Walsh, T., Wu, B., Tsutsui, J., & Allen, M. R. (2021). FaIRv2.0.0: A generalized impulse response model for climate uncertainty and future scenario exploration. *Geoscientific Model Development*, *14*(5), 3007–3036. https://doi.org/10.5194/gmd-14-3007-2021

Morton, C. G. ;O'Neil., R. D.;Marston. (2022). *Land cover map 2020 (1km summary rasters, GB and n. ireland)*. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/d6f8c045-521b-476e-b0d6-b3b97715c138

Nations", "United. (2022). *United nations, department of economic and social affairs, population division (2022). World population prospects 2022: Summary of results. UN DESA/POP/2022/TR/NO. 3.*

Poore, J., & Nemecek, T. (2018). Reducing food’s environmental impacts through producers and consumers. *Science*, *360*(6392), 987–992. https://doi.org/10.1126/science.aaq0216