The Journal of Open Source Software

interflow: A Python package to organize, calculate, and visualize sectoral interdependency flow data

Kendall L. Mongird[1], **Konstantinos Oikonomou**[1], **Juliet S. Homer**[1], **and Jennie S. Rice**[1]

1 Pacific Northwest National Laboratory, Richland, WA, USA

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Frauke Wiese

Reviewers:

- @wiljnich
- @j3r3m1

Submitted: 30 March 2022
Published: unpublished

Summary

Many economic sectors rely on an uninterrupted "upstream" supply of a resource to conduct their primary functions, leaving them vulnerable to adverse effects should that resource flow be interrupted or compromised (OECD, 2017; U.S. EPA, 2010). Well-known examples of these relationships include water demand by the energy sector (e.g., thermoelectric cooling for nuclear generation) (Grubert & Sanders, 2018; Webber, 2017) and energy demand by the water sector (e.g., electricity required to treat or move water in the public water sector) (Congressional Research Service, 2017) though many others exist. Being able to calculate and document these interdependencies and evaluate where the greatest cross-sectoral intensities and flows exist can reveal opportunities to enhance the overall network. Despite the implications and potential impacts, however, these interconnections and flows have been historically complex to analyze and understand.

The `interflow` package provides a flexible tool to organize, calculate, and visualize (using Sankey diagrams and other visualizations) sectoral interdependency flows for multiple subsectors and resources (Figure 1). This tool can help decision-makers, researchers, and other audiences more easily pull meaning from these interdependencies to reveal multi-faceted opportunities and risks. `interflow` can help investigate questions such as (1) which sectors have high cross-resource dependencies, (2) how does demand for a resource in various sectors compare across regions, and (3) where the sectoral and regional opportunities are for enhanced efficiency, security, and resiliency.

The Journal of Open Source Software

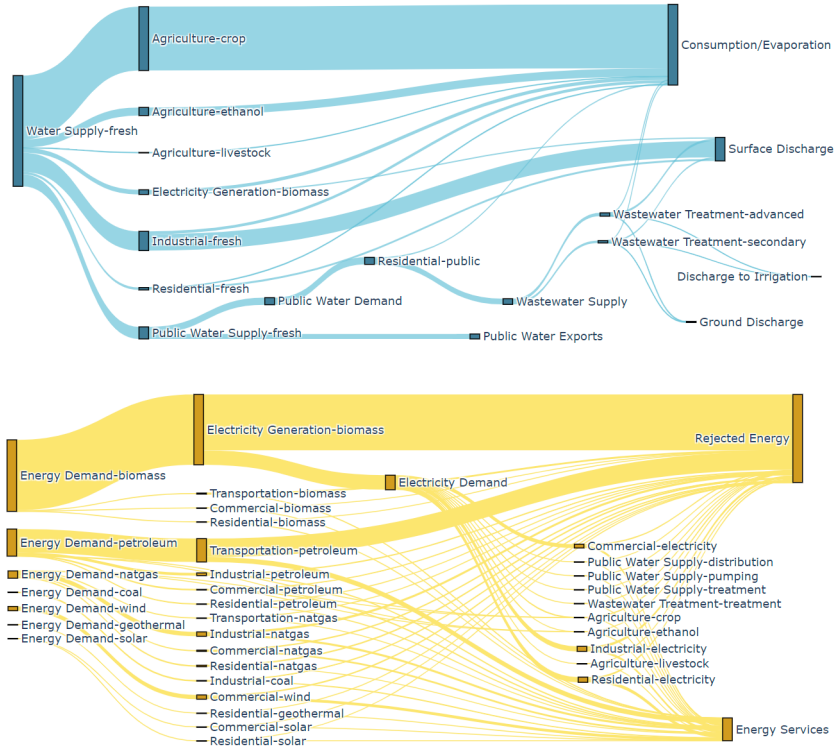

Figure 1: Example Sankey Diagram Demonstrating the Flow of a Resource Across Sectors.

Statement of Need

To the best of our knowledge, there is no open-source software option available that facilitates the calculation, organization, and visualization of multiscale, multisector interdependencies in one place. Software exists for visualizing pre-calculated sectoral flow values as Sankey diagrams such as the proprietary e!Sankey software (iPoint-systems gmbh, 2022) or the Python library matplotlib (Hunter, 2007), but these tools do not offer a way to calculate the demand of a cross-resource type and build out new sectoral connections based on the result. The `interflow` package aims to fill this gap and serve as a flexible and open-source option for conducting multi-resource sectoral interdependency data calculations, producing multisectoral datasets, and visualizing the results.

Sectoral interdependency analysis itself is not a new area of research (Curmi et al., 2013; Greenberg et al., 2017; Liu et al., 2016), but the publications in this area typically return only the end-product of their analysis (e.g., output values or diagrams). The detailed methodology or algorithms developed to conduct their calculations are often not provided in an accessible or transparent manner. Researchers looking to build upon or modify the existing assumptions and data are left to redevelop the calculation structure from scratch as a result. The `interflow` package provides a consistent and open-source calculation framework to enable reproducibility.

Design and Methodology

The `interflow` package iterates through a user-provided tabular input data to (1) collect known resource flow values (e.g., water demand) between provided sectors, (2) calculate new sector flow values for a secondary resource type from the initial flow (e.g., energy demand based on the water flow) using provided cross-resource intensity coefficients, and (3) build upstream and downstream sector connections to carry those calculated and collected flows.

Collected and calculated flows are then compiled in a way that can be used in a variety of data visualization functions offered in the package.

The basic methodology described above is repeated for all regions provided by the user in the input data. Though the `interflow` package comes equipped with extensive sample data to evaluate U.S. county-level water and energy flows across various sectors, it can conduct analysis for any region (e.g., country, province), any sector or group of sectors (e.g., electricity generation, agriculture), and any group of resources (e.g., water, energy, food, carbon, land) and is limited only by the input data that the user provides. More information on the input data requirements can be found in the generalizability documentation. All code used to generate the sample water and energy input data for all US counties is additionally included as part of the package.

A Pandas DataFrame (McKinney & others, 2010) containing collected and calculated flow values between sectors for each region and resource type is returned as output from the `calculate()` function. The DataFrame output can be directly used with other package functions to generate a variety of visualizations which each utilize the Plotly Python package (Plotly Technologies Inc., 2015) including (1) Sankey diagrams showing the network of flows across sectors for a chosen region, (2) stacked bar charts of inflow and outflow values for sectors in a region, and (3) a choropleth map to compare flow values across regions. The visualizations can be used to compare dependencies across and between sectors at various levels of sub-sector granularity. Given that `interflow` collects and calculates values for each region provided in the input data, the output can additionally be used for region-wise comparisons of flows and intensities. For more information on the key outputs see the outputs documentation.

Acknowledgements

This research was supported by the U.S. Department of Energy, Water Power Technologies Office, as part of research in Integrated Water Power Resilience.

References

Congressional Research Service. (2017). *Energy-water nexus: The water sector's energy use*. [PDF]. https://sgp.fas.org/crs/misc/R43200.pdf

Curmi, E., Fenner, R., & Richards, K. et al. (2013). Visualising a stochastic model of californian water resources using sankey diagrams. *Water Resources Management*, *27*, 3035–3050. https://doi.org/10.1007/s11269-013-0331-2

Greenberg, H. R., Simon, A. J., Singer, S. L., & Shuster, E. P. (2017). *Development of energy-water nexus state-level hybrid sankey diagrams for 2010*. [PDF]. https://flowcharts.llnl.gov/content/assets/docs/2010_United-States_EnergyWater.pdf

Grubert, E., & Sanders, K. T. (2018). Water Use in the United States Energy System: A National Assessment and Unit Process Inventory of Water Consumption and Withdrawals. *Environmental Science & Technology*, *52*(11). https://doi.org/10.1021/acs.est.8b00139

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

iPoint-systems gmbh. (2022). *e!Sankey - show the flow*. https://www.ifu.com/e-sankey/

Liu, Y., Hejazi, M., Kyle, P., Kim, S. H., Davies, E., Miralles, D. G., Teuling, A. J., He, Y., & Niyogi, D. (2016). Global and regional evaluation of energy for water. *Environmental Science & Technology*, *50*(17), 9736–9745. https://doi.org/10.1021/acs.est.6b01065

McKinney, W., & others. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *445*, 51–56. https://doi.org/10.

25080/Majora-92bf1922-00a

OECD. (2017). *The land-water-energy-nexus: Biophysical and economic consequences*. https://read.oecd-ilibrary.org/environment/the-land-water-energy-nexus_9789264279360-en#page4

Plotly Technologies Inc. (2015). *Collaborative data science*. Plotly Technologies Inc. https://plot.ly

U.S. EPA. (2010). *Understanding water sector interdependencies*. [PDF]. https://nepis.epa.gov/Exe/ZyPDF.cgi/60001QB5.PDF?Dockey=60001QB5.PDF

Webber, M. E. (2017). Chapter 2.2.6 - Water for Electricity Generation in the United States. In *Competition for water resources: Experiences and management approaches in the US and europe*. Elsevier. ISBN: 9780128032374