

A Python Based Multi-Regional Input-Output Analysis Toolbox: Pymrio - newest updates and future developments

Topic: Advances in Open Source Software for Input-Output Compilation, Analysis and Quantitative Impact Assessment

Author: Konstantin STADLER

Co-Authors: Hazim Hussein, Mohamed Badr, Peter Maxwell, Candy Deck, Ruslan Zhuravchak

Pymrio is an open-source tool for Environmentally-Extended Multi-Regional Input-Output (EE MRIO) analysis, developed in Python. The tool provides a high-level abstraction layer for global EE MRIO databases. This enables a harmonized approach towards common EE MRIO data tasks, including

- automatically downloading (specific years and versions) from a EE MRIO database repository,
- reading the downloaded raw EE MRIOs data into a common, Pandas Dataframe based structure,
- easy access to MRIO matrix values by implementing label based indexing,
- common, unit-tested functions for calculating production/consumption based accounts, aggregation, stressor origin analysis etc.
- visualization and reporting of results

Pymrio is a (MR)IO type agnostic tool which can handle any (MR)IO database with a consistent sector classification. Pymrio currently includes parser for several openly available monetary EE MRIO databases (EXIOBASE v1-v3, WIOD, Eora26, OECD-ICIO) while others are planned to be implemented in the near future (e.g. GLORIA, FIGARO). This makes Pymrio ideally suited for data pipelines which need to process several MRIO databases, as for example for the MRIO comparisons website on www.environmentalfootprints.org.

Since Pymrio is an already established tool we will focus this presentation on the latest additions to functionality and the near future development plan. Thus, after briefly outlining the internal structure of Pymrio we will present a tutorial covering:

1. Available consistent data storage formats for EE MRIOs, including a newly implemented, efficient data storage format based on the Apache Parquet data file format.
2. Improved performance for MRIO analysis based on parallelizing calculations on either GPU or toward cloud/HPC systems.
3. How to make use of the improved performance for conducting Monte Carlo based uncertainty/sensitivity analysis with Pymrio.

Pymrio is an open source project which welcomes contributions. These can be in form of improvements on the documentation, bug-fixes or adding completely new functionality. The code is hosted on Github, and we will show how to contribute to Pymrio with a focus on how to initiate/discuss a new contribution, preferred code style and required test coverage.

We will then put Pymrio in a wider context of available Input-Output and Industrial Ecology related tools. There we touch on connected open-source projects, like a new system to add custom extensions to the EXIOBASE MRIO system, which makes use of Pymrio. We will also show how Pymrio can be used together with the open source Python package country converter CoCo to automatically build aggregation definitions for MRIO countries to regions without the need to manually build concordance matrices.

Finally, we will present how to obtain Pymrio from the standard Python packaging repositories (PyPI and conda-forge). This will also include a short introduction into the Industrial Ecology Open Source

dashboard and its Input-Output section (<https://github.com/IndEcol/Dashboard>), which can be used by any researcher to present Open Source tools for our field.

In summary, we aim to give IO researchers, with some basic coding skills, a good understanding on how to use Pymrio as well as the necessary knowledge to report issues and actively contribute to its development. In addition, we hope that by highlighting the existence of the dashboard for collecting Industrial Ecology and Input-Output tools we can contribute to more open code sharing and development in IO research and analysis.