The Journal of Open Source Software

Cobwood: Enhancing Forest Economics Model Reusability Through labelled Panel Data Structures

Paul Rougieux [1¶]

1 European Commission, Joint Research Centre, Ispra, Italy ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Ethan White

Reviewers:

- @tomkeH
- @AtmaMani

Submitted: 25 June 2025
Published: unpublished

Summary

Forest Sector Models provide quantitative insights into how policy changes, resource scarcity or climate impacts propagate through international forest products markets. These models balance demand, trade and wood supply in each country and project future market dynamics over decades. The cobwood package provides a labelled data structure that enhances code clarity and facilitates model inspection compared to traditional approaches. To demonstrate cobwood's application, we present a reimplementation of the Global Forest Products Model (GFPMx). The reusable data structure positions cobwood as an ideal component for integration into a greater modelling tool chain.

Statement of need

Trees grow over decades or centuries and wood markets can be very localized. Yet markets for processed wood and paper products are interconnected through global trade networks. To simulate the future condition of wood markets, forest economists rely on macroeconomic forest sector models. These models operate in static and dynamic phases. The static phase balances supply and demand within a single year. The dynamic phase projects future demand and supply driven by exogenous factors like GDP growth and changes in the forest stock.

Several global forest sector models currently exist, including the Global Forest Products Model (GFPM)(Buongiorno et al., 2003), the European Forest Institute Global Trade Model (EFI-GTM)(Kallio et al., 2004), the Global Forest and Agriculture Model (G4M)(Gusti, 2020), the Global Forest Trade Model (GFTM) (Jonsson et al., 2015) and an adaptation called Timba (Fsm et al., 2025). There are also multiple regional and national forest sector models.

Current modelling softwares often use partial labelling approaches with unclear variable names that make source code difficult to interpret. Many models are not open source, and limited data labelling makes model outputs difficult to reuse. The cobwood library addresses transparency limitations by organizing forest sector datasets as labelled multi-dimensional arrays with explicit country, product, and year dimensions. Metadata and dimension names are stored directly within datasets (both in memory and on disk), making model inputs and outputs easier to interpret and share.

Existing modelling softwares typically relies on ad-hoc input file management to distinguish scenarios, with scenario assumptions scattered across multiple data files, in the input file name itself , or in separate Excel sheets. Cobwood uses explicit scenario configuration files, improving reproducibility and enabling transparent version control of modeling assumptions."

By standardizing data representation and exposing model internals, cobwood facilitates collaboration beyond forest economics. Adjacent fields such as forest management, vegetation dynamics, and life cycle analysis rely on projections of roundwood harvest and wood products consumption. An interoperable python package enables these communities to reuse and extend existing models to simulate new drivers and analyze new policy questions.

Input, output

A yaml file in the `cobwood_data/scenario` directory defines the particular input data used for a given scenario. Cobwood can load input data from any tabular source that pandas support. For instance, the GFPMx data is stored inside a single Excel spreadsheet containing many sheets for consumption, production, import, export, and prices of major forest products. A script first converts sheets to CSV files, which the `GFPMXData.convert_sheets_to_dataset` then transforms into an Xarray data structure. Other methods make it possible to load forest products market data from the FAOSTAT API and to transform them into xarray datasets.

The `write_datasets_to_netcdf` combines many products 2D datasets into one larger 3D dataset, by adding a third coordinate called "`product`" before saving the model output datasets to NetCDF files. These files include metadata labels for units. While not commonly used in economics, NetCDF is a widely accepted format in earth systems modelling, making it ideal for integrated modelling systems.

Data structure and implementation

Figure 1 illustrates the data structure:

- Global consumption, production, trade flows, and prices for all countries, all years and for each forest product are stored as an Xarray dataset (e.g., `model["sawn"]` for sawnwood)
- Within each dataset for one product, specific variables are accessible as two-dimensional arrays with country and year coordinates (e.g., `model["sawn"]["cons"]` for consumption)

To explore available variables, users can access the `variables` property (e.g., `model["sawn"].variables`). Array properties are used to store metadata, the example below displays the roundwood production unit :

```
model["indround"]["prod"].unit
# '1000m3'
```

The cobwood model has been used to produce scenario analysis Mubareka et al. (2025) and Rougieux et al. (2024). The first model programmed inside cobwood is a reimplementation of a simple forest sector model called GFPMx (Buongiorno, 2021). Labelled data arrays allow developers to write Python functions that closely mirror the mathematical equations found in the academic papers describing the models, with explicit time and country dimensions. For example the demand function in `cobwood/gfpmx_equations.py` is implemented on an xarray dataset `ds` where a dependent variable such as GDP is selected for all countries at time `t` with `ds["gdp"].loc[ds.c, t]`.

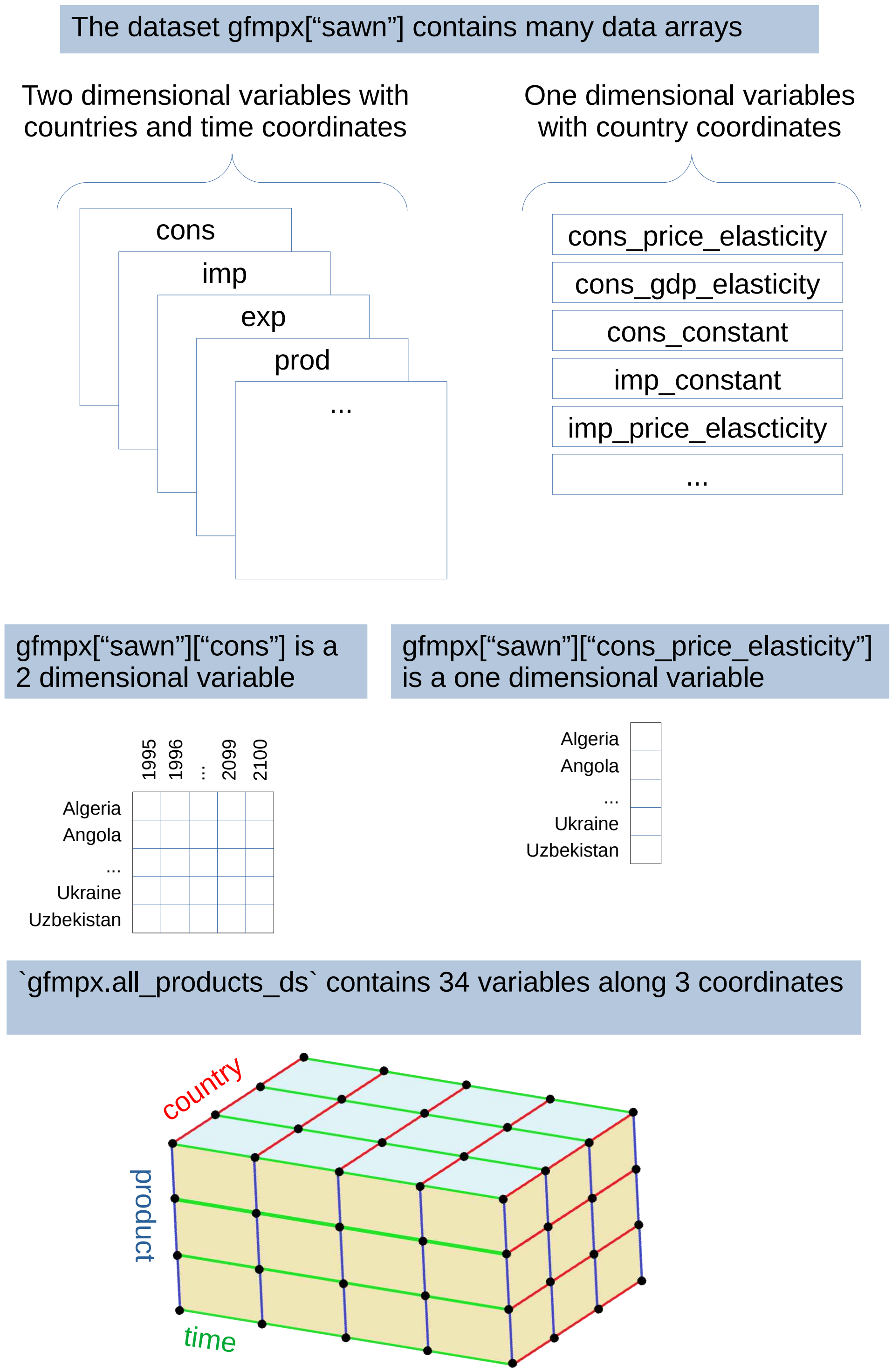

Figure 1: Data structure

Model run

The following code instantiates a GFPMX model object from a scenario yaml file. The `rerun=True` argument erases previous model runs, while `compare=True` compares output with the reference Excel implementation of GFPMx

```
from cobwood.gfpmx import GFPMX
gfpmxb2021 = GFPMX(scenario="base_2021", rerun=True)
gfpmxb2021.run(compare=True, strict=False)
```

The model output data is saved inside the model's `output_dir` directory. When re-using the model later, specify the argument `rerun="False"` (default) to load the output data without the need to run the model.

Visualisation

The following python code draws a faceted plot of industrial roundwood consumption, import, export, production and price with one line by continent.

```
gfpmxb2021.facet_plot_by_var("indround")
```

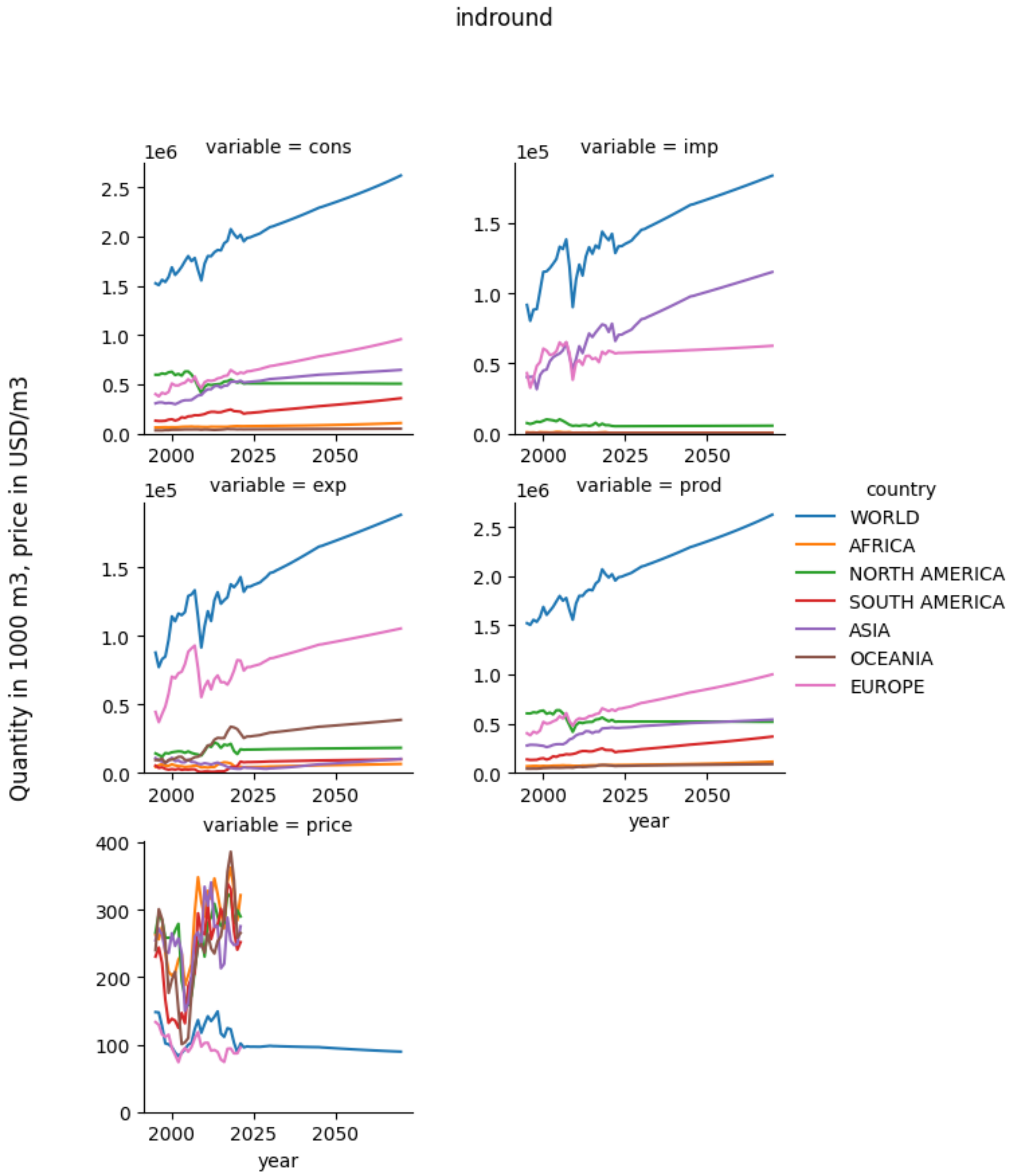

Figure 2: Industrial roundwood consumption, import, export, production and prices by continent

Specify the country argument to get one line by country

```
gfpmxb2021.facet_plot_by_var("indround", countries=["Canada", "France", "Japan"])
```

The Journal of Open Source Software

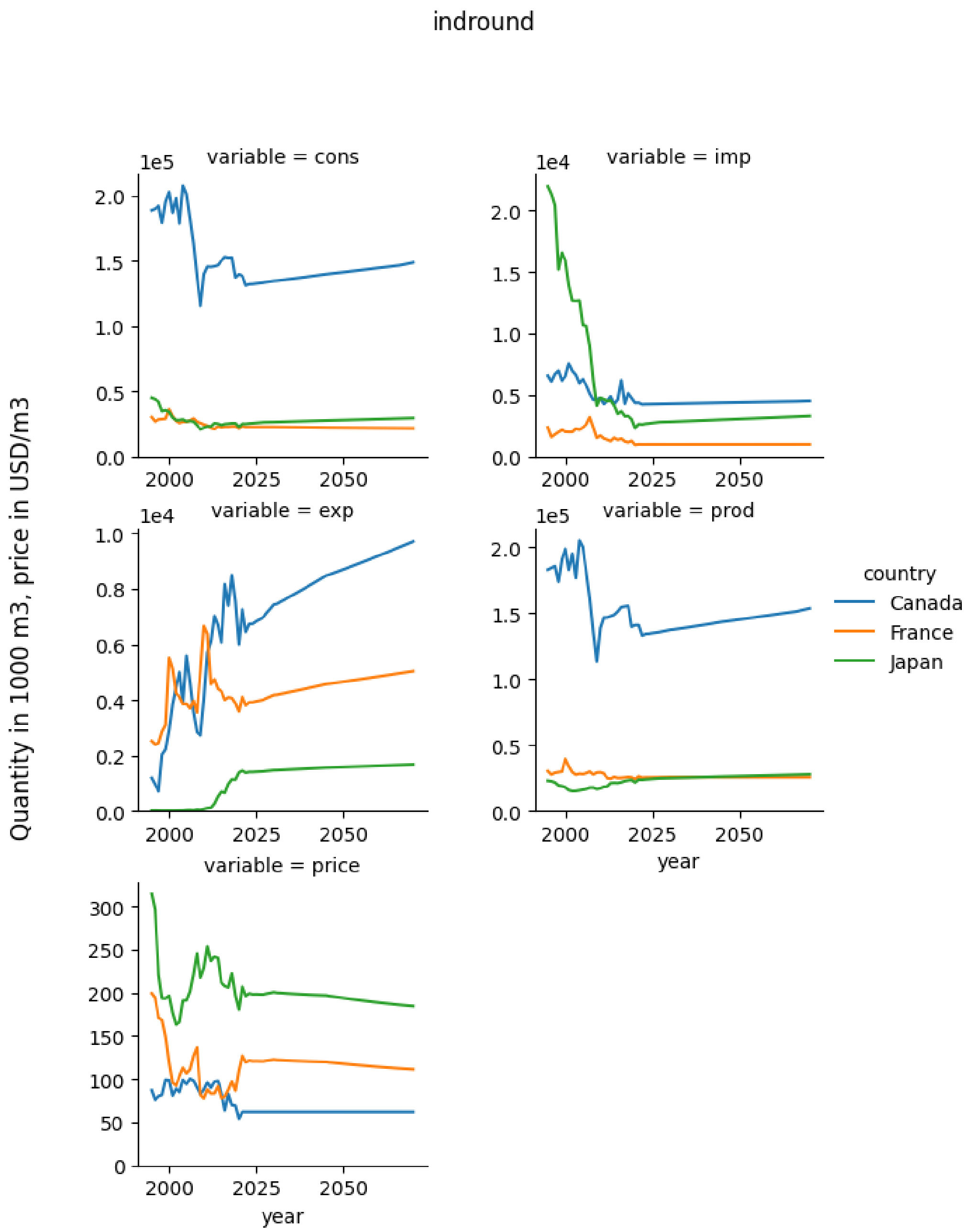

Figure 3: Industrial roundwood consumption, import, export, production and prices by country

Conclusion

The cobwood package represents macroeconomic forest products market data as N-dimensional labelled data arrays. The data structure incorporates comprehensive metadata and coordinates improving source code readability and model transparency. Additionally, the scenario configuration file enables comparison of different model implementations across variations of input parameters. Furthermore, model outputs are saved to NetCDF files, which preserve dimensions and metadata. This data structure will be reused to implement many other forest sector models. Ultimately, the goal is to facilitate the integration of forest sector models as components of interdisciplinary modelling tool chains.

References

Buongiorno, J. (2021). GFPMX: A cobweb model of the global forest sector, with an application to the impact of the COVID-19 pandemic. *Sustainability*, *13*(10), 5507. https://doi.org/10.3390/su13105507

Buongiorno, J., Zhu, S., Zhang, D., Turner, J., & Tomberlin, D. (2003). *The global forest products model: Structure, estimation, and applications*. Elsevier.

Fsm, T., Morland, C., Schier, F., Tandetzki, J., & Honkomp, T. (2025). TiMBA (timber market model for policy-based analysis). *Journal of Open Source Software*, *10*(115), 8034. https://doi.org/10.21105/joss.08034

Gusti, M. (2020). *G4M model documentation*. Laxenburg, Austria: International Institute of Applied Systems Analysis.

Jonsson, R., Rinaldi, F., & San-Miguel-Ayanz, J. (2015). The global forest trade model. *Luxembourg: Publications Office of the European Union*.

Kallio, A. M. I., Moiseyev, A., Solberg, B., & others. (2004). The global forest sector model EFI-GTM—the model structure. *European Forest Institute-Internal Report*, *15*.

Mubareka, S. B., Renner, A., Acosta Naranjo, R., Armada Bras, T., Barredo Cano, J. I., Blujdea, V., Borriello, A., Borzacchiello, M. T., Briem, K., Camia, A., Canova, M., Cazzaniga, N. E., Ceccherini, G., Ceddia, M., Cerrani, I., Chiti, T., Colditz, R., De Jong, B., De Laurentiis, V., … Zulian, G. (2025). *EU biomass supply, uses, governance and regenerative actions* (S. B. Mubareka & A. Renner, Eds.). Publications Office of the European Union. https://doi.org/10.2760/6511190

Rougieux, P., Pilli, R., Blujdea, V., Mansuy, N., & Mubareka, S. B. (2024). *Simulating future wood consumption and the impacts on europe's forest sink to 2070*. Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC136526