bonsai_ipcc - a Python package for the calculation of national greenhouse gas inventories

Maik Budzinski [1], **Joao F. D. Rodrigues** [2], and **Mathieu Delpierre** [2]

1 Department of Sustainability and Planning, Aalborg University, Denmark. **2** 2.-0 LCA consultants, Aalborg, Denmark.

Summary

The aim of the `bonsai_ipcc` Python package is to enable users to calculate national greenhouse gas (GHG) inventories based on the guidelines provided by the International Panel on Climate Change (IPCC) ([Intergovernmental Panel on Climate Change, 2023](#)). In implementing the equations and parameter data from these guidelines, the package adheres to the organizational structure outlined in the guidelines' PDF documents, which include volumes and chapters. The package allows users to add their own data. In addition to computing default GHG inventories, the software includes tools for error propagation calculation, such as analytical error propagation and Monte Carlo simulation, both of which are endorsed by the IPCC reports.

Statement of need

Gathering greenhouse gas (GHG) data is an important step when developing models and scenarios in many environmental sciences. The official guidelines for estimating national GHG inventories have been widely used in the modelling community, e.g. to create environmentally extended input-output models ([Merciai & Schmidt, 2018](#); [Stadler et al., 2018](#)) or datasets for life cycle assessment ([Nemecek & Kaegi, 2007](#); [Schmidt et al., 2021](#)). The IPCC guidelines contain equations and default data that can be used to calculate country-based greenhouse gas inventories, taking into account different production and treatment activities. However, calculating GHG inventories directly from the report is cumbersome and time consuming, requiring manual data extraction and visual inspection to identify the sequence of formulas that must be implemented. To facilitate the compilation of GHG inventories, we developed an open-source Python package which stores the default data and implements the formulas of the IPCC report. To the best of our knowledge, `bonsai_ipcc` is the first work that aims to implement all information of the IPCC report. Presently, open-source Python packages exclusively incorporate particular equations from the IPPC reports. Furthermore, the implemented information serve merely as a component rather than the central focus of the Python package. An example is the Python package `hestia_earth.models` that includes specific equations of the volume agriculture to calculate environmental impacts for farms and food products ([Hestia, 2024](#)).

Structure of the package

The structure of `bonsai_ipcc` Python package is illustrated in figure 1. The equations (in the following elementary equations) of a chapter are used to define the sequence (tier approach) to calculate the corresponding GHG inventory. Data for default parameter values of the guidelines is provided within the package. We use the Python package `frictionless` to describe and validate the provided data tables ([frictionlessdata, 2024b](#)).

As a user, you choose the sequence and specify the dimensions (e.g., year, region) of the

involved parameters. The result is a sequence of steps that store the involved parameter values and values that are calculated by elementary equations (represented by circles and rectangles, respectively in figure 1), as well as the involved uncertainty.
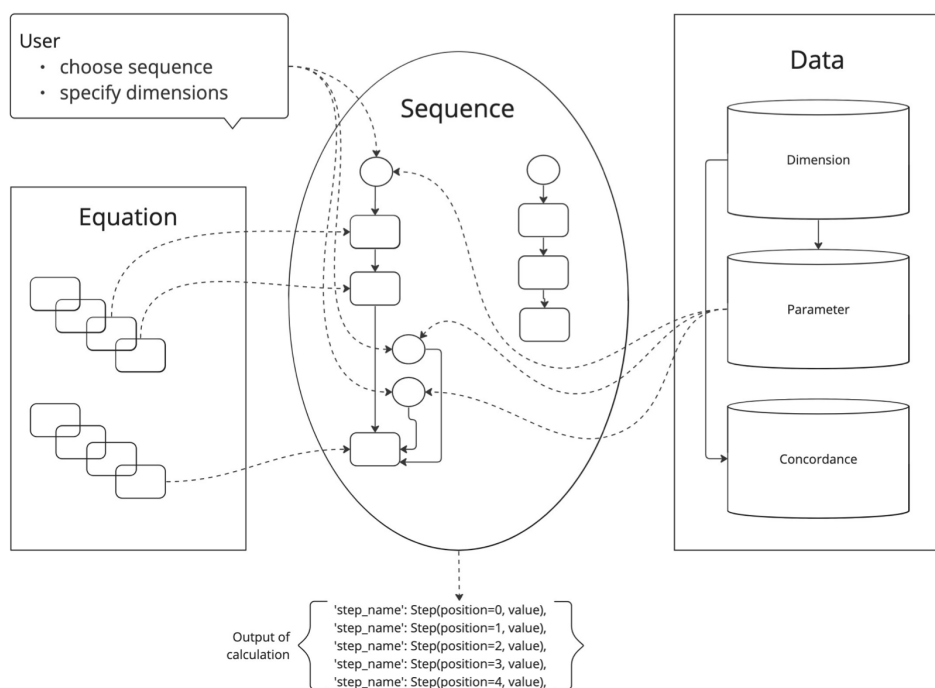
Figure 1: Structure of the bonsai_ipcc Python package

The package structure also follows the structure of the guidelines for estimating national GHG inventories. Each of the four core `<volume>`s (i.e., energy, agriculture, energy and waste) contains `<chapter>`s with elementary equations, which can be used to define the tier 1, 2 and 3 sequences calculating the inventories for GHG emissions (e.g., CO2, CH4 and N2O).

```
bonsai_ipcc.<volume>.<chapter>.sequence
bonsai_ipcc.<volume>.<chapter>.elementary
```

To distinguish between the different tiers 1, 2 and 3 when calculating the inventories for GHG emissions, the naming convention of the corresponding methods is as follows.

```
bonsai_ipcc.<volume>.<chapter>.sequence.tier<number>_<GHG>()
```

Functionality

The core feature of the `bonsai_ipcc` package is to determine GHG emissions for different tiers based on the provided data.

```
tier<number>_<ghg>(year,region,<producttype>,<activitytype>,uncertainty)
```

Since the IPCC guidelines specify tier methods for each GHG separately, we decided to make this distinction in the name of the function instead of using an argument. The outcome is recorded as a series of sequential steps, encompassing all input data and interim findings. Each step provides the parameter's name, along with its corresponding value and unit. This meticulous recording of data input and intermediate results enhances transparency and facilitates comprehensive analysis within the calculation of GHG inventories.

Data handling

The IPCC guidelines also provide default data for a large amount of parameters that are used in the elementary equations. This data is included in the Python package. When including the data into the package, we follow the frictionless standards (frictionlessdata, 2024a). These standards provide patterns to describe data, such as tables, files and datasets. The framework follows the five design principles - simplicity, extensibility, human-editable and machine-usable, reusable and applicable across different technologies. The parameter dimension and concordance tables are associated to the volume and chapter where these data is used.

```
bonsai_ipcc.<volume>.<chapter>.parameter.<table>
bonsai_ipcc.<volume>.<chapter>.dimension.<table>
bonsai_ipcc.<volume>.<chapter>.concordance.<table>
```

The data for parameters and dimensions is stored in tabular format as csv files. To query the values within the `bonsai_ipcc` package, we use pandas `DataFrame` (McKinney, 2010). To automate the process of selecting the right parameter when building the tier sequences, the package uses the concept of concordance tables. Thereby, each attribute of a dimension, e.g. country `DE` in the dimension `region`, can be associated to other more aggregated attributes (e.g., `Western Europe`). This has the advantage that parameter values can be selected from other attributes in cases where the guidelines only provide data for more aggregated ones. When reading the values from a specific parameter table, the sequence algorithm first tries to find the dimension on the left hand side and proceeds stepwise to the right until a value is found. The same principle is used for other dimensions, including year and <producttype>.

Uncertainty

Two methods for uncertainty analysis are implemented in the `ipcc` package: analytical error propagation and Monte Carlo method. When running the sequence, the type of `value` in each step depends on the selected method for uncertainty calculation (`float` for `uncertainty="def"`, the `ufloat` type of the unertainies library (Lebigot, 2024) for `uncertainty="analytical"` and a NumPy array (Harris et al., 2020) for `uncertainty="monte_carlo"`). Based on the provided uncertainty information for a parameter table, the algorithm chooses the proper type of uncertainty distribution. The following distribution types are implemented, `normal`, `lognormal`, `truncated normal`, `uniform`, `truncated exponential` and `beta` distribution. Truncated normal distributions are adjusted based on Rodrigues (2016) so that original mean and standard deviation are perpetuated.

Conclusion

The transformation of the IPCC guidelines for calculation greenhouse gas inventories into the `bonsai_ipcc` Python package is an important step towards reproducibility and automation of national GHG inventory results. Furthermore, users of the package can use the results when developing models and scenarios in different scientific fields. Due to the magnitude of the IPCC guidelines, the implementation of its volumes into the Python package is an ongoing process. To this date one volume (waste) out of the four core volumes has been fully implemented. A second one (agriculture) is in progress. The implementation of a third one (industry) has been started. And a fourth (energy) is waiting to be initialized.

Acknowledgments

References

frictionlessdata. (2024a). *Data software and standards*. https://specs.frictionlessdata.io/

frictionlessdata. (2024b). Frictionless-py. In *GitHub repository*. GitHub. https://github.com/frictionlessdata/frictionless-py

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hestia. (2024). Hestia engine models. In *GitLab repository*. GitLab. https://gitlab.com/hestia-earth/hestia-engine-models

Intergovernmental Panel on Climate Change. (2023). *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. https://www.ipcc-nggip.iges.or.jp/public/2019rf/index.html

Lebigot, E. O. (2024). Uncertainties. In *GitHub repository*. GitHub. https://github.com/lmfit/uncertainties

McKinney, Wes. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Merciai, S., & Schmidt, J. (2018). Methodology for the construction of global multi-regional hybrid supply and use tables for the EXIOBASE v3 database. *Journal of Industrial Ecology*, *22*(3), 516–531. https://doi.org/10.1111/jiec.12713

Nemecek, T., & Kaegi, T. (2007). *Ecoinvent report no. 15 - life cycle inventories of agricultural production systems*. https://db.ecoinvent.org/reports/15_Agriculture.pdf

Rodrigues, J. D. F. (2016). Maximum-entropy prior uncertainty and correlation of statistical economic data. *Journal of Business & Economic Statistics*, *34*(3), 357–367. https://doi.org/10.1080/07350015.2015.1038545

Schmidt, J., Merciai, S., Munoz, I., Rosa, M. de, & Astudillo, M. (2021). *The big climate database v1 – methodology report*. https://denstoreklimadatabase.dk/files/media/document/Methodology_report_The_Big_Climate_Database_20240116.pdf

Stadler, K., Wood, R., Bulavskaya, T., Södersten, C.-J., Simas, M., Schmidt, S., Usubiaga, A., Acosta-Fernández, J., Kuenen, J., Bruckner, M., Giljum, S., Lutter, S., Merciai, S., Schmidt, J. H., Theurl, M. C., Plutzar, C., Kastner, T., Eisenmenger, N., Erb, K.-H., … Tukker, A. (2018). EXIOBASE 3: Developing a time series of detailed environmentally extended multi-regional input-output tables. *Journal of Industrial Ecology*, *22*(3), 502–515. https://doi.org/10.1111/jiec.12715