gcamreport: An R tool to process and standardize GCAM outputs

Clàudia Rodés-Bachs[1], **Jon Sampedro**[1], **Dirk-Jan Van de Ven**[1], **Ryna Yiyun Cui**[2], **Alicia Zhao**[2], **Matthew Zwerling**[2], and **Zarrar Khan**[3]

1 Basque Centre for Climate Change (BC3), Leioa, Spain **2** Center for Global Sustainability, School of Public Policy, University of Maryland, College Park, MD, USA **3** Pacific Northwest National Laboratory's Joint Global Change Research Institute, College Park, MD, USA

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Martin Fleischmann

Reviewers:

- @bpbond
- @ibarraespinosa

Submitted: 24 July 2023
Published: unpublished

Summary

There is an urgent need to perform multi-model studies to deal with models' heterogeneity and uncertainty with the aim of building a more reliable and transparent framework to inform policy makers in the design and implementation of climate policies (Guivarch et al., 2022). In response to this challenge, multiple institutes and organizations have adopted the standardized data template developed by the Integrated Assessment Modeling Consortium (IAMC). This template is maintained by the International Institute for Applied Systems Analysis (IIASA) and aims to standardize and facilitate model intercomparison exercises. For the latest Assessment Report (AR6), the Intergovernmental Panel on Climate Change (IPCC) required all contributors to homogenize their data to enable comparisons and ensure full transparency (Krey et al., 2014). This practice has set the foundation for a new open management of the outputs in the area of global scenario analysis.

In the case of the Global Change Analysis Model (GCAM) (Calvin et al., 2019), a well-reputed model that has been extensively used by different international and national scenario analysis, the harmonization code has never been documented nor standardized, making it difficult to reproduce the outputs and hindering the transparency of the results. To overcome these limitations, we have developed `gcamreport`, an R package that systematizes the transformations of GCAM outputs, generates figures to facilitate the analysis of the results, and allows user interaction with the produced outputs. Furthermore, the tool can be used embedded in a Docker image, which allows users to use the package in a virtual environment without having to install any specific software or library. Finally, each `gcamreport` release is linked to either a version of GCAM or a study in which GCAM was used, ensuring reproducibility, interoperability, accessibility, and findability, which is in line with the well-known open science principles FAIR and TRUST (Lin et al., 2020; Wilkinson et al., 2016).

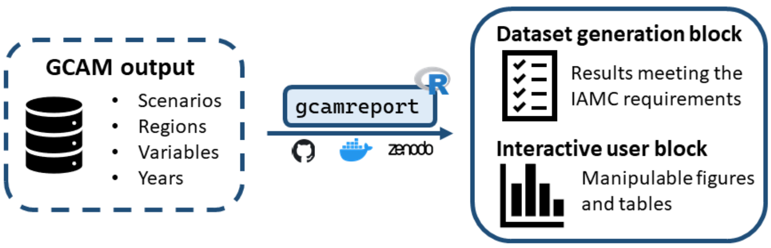

Figure 1: Figure 1: Structure of the `gcamreport` package.

The `gcamreport` package is accessible online at the public domain https://github.com/bc3LC/

gcamreport. To run it to generate and save the dataset meeting the IAMC requirements and to launch the user interface, you can follow this simplified code:

```
# install the package
remotes::install_github("bc3LC/gcamreport")
library(gcamreport)

# run the main function with all the funcionalities integrated
run(project_path = awesomeProject.dat, final_year = 2100,
save_output = TRUE, launch_ui = TRUE)
```

The additional instructions to run the package through a Docker image and with further functionalities can be found in the online documentation

Statement of need

Integrated assessment models (IAMs) are the dominant tools for global scenario analysis and for exploring system-wide implications of climate policies (Van Beek et al., 2020). They must be transparent and open in order to be an effective instrument for informing decision making. In addition, multi-model studies should be considered to strengthen the consistency and robustness of the policy recommendations (Nikas et al., 2021). However, it is not always easy to compare outputs from different models, as it involves a laborious task of bringing together modeling communities with different backgrounds. To overcome these handicaps, the IAMC developed a time-series data template (Huppmann et al., 2023) which has been used in distinguished multi-IAM studies, such as the presentation of the Shared Socio-economic Pathways (SSPs) (O'Neill et al., 2014), the design of the Representative Concentration Pathways (RCPs) (Moss et al., 2010), and the last two IPCC Assessment Reports (AR5 (Masson-Delmotte et al., 2021) and AR6 (Shukla, J. et al., 2022)), as well as the IPCC Special Report on Global Warming of 1.5ºC (Masson-Delmotte et al., 2018) and the recent climate scenarios of the Network for Greening the Financial System (NGFS) (Richters et al., 2022). For this kind of assessments, it is essential that the outputs from the different models meet strict requirements and are reported in a standardized way to facilitate the comparison and multi-analysis (Skea et al., 2021).

One of the most extensively applied integrated assessment model is the Global Change Analysis Model (GCAM) (Calvin et al., 2019). GCAM is an open-source multi-sector model developed at the Joint Global Change Research Institute (JGCRI) designed to explore the linkages between energy, water, land, climate, and economics within a single computational system. It enables users to explore potential what-if type futures dividing the world in 32 geopolitical regions and running in a 5-year time step. GCAM is in continuous development, with each new version increasing its accuracy by better representing and detailing existing elements, while adding new ones and fixing old ones (Binsted et al., 2022; Sampedro et al., 2022). Hence, the code to transform GCAM outputs into suitable outputs for model intercomparison should be adapted to each particular version/study. Otherwise, it becomes difficult to reproduce, automatize, and track changes (Stodden et al., 2018).

Therefore, `gcamreport` is a powerful R-tool, aligned with the principles of open science (Lin et al., 2020; Wilkinson et al., 2016), to guarantee the transparency of the produced outcomes by transforming GCAM outputs to the IAMC template requirements, making the results directly applicable to multi-IAM studies (including upcoming IPCC reports). Moreover, to overcome the limitations to reproduce the execution environment (e.g., packages or libraries), `gcamreport` is also available within a Docker image, which already contains all the required packages and facilitates the usage, ensuring interoperability and reproducibility (Boettiger, 2015).

Finally, to support the analysis of the model results, `gcamreport` includes a user-friendly interface that allows to visualize the standardized outputs in tabular format, allowing the user to subset the results by regions, scenarios, models, variables, and years, along with the option to download the desired results. Moreover, the `gcamreport` user interface generates plots of the selected variables within the same category, aggregated by region and/or sub-variables, which can also be downloaded for more in-depth analysis.

Functionality

The `gcamreport` package consists of a set of functions divided into two different blocks. All functions are described in a specific documentation supplemented by detailed tutorials.

- Dataset generation block: Transformation and saving of outputs from a created or pre-loaded GCAM database to meet the IAMC formatting requirements.
- Interactive user block: Dataset visualization and manipulation, and figures generation.

To read the raw GCAM outputs, `gcamreport` uses rgcam. This R package is part of the GCAM-ecosystem, which is a set of complementary tools to the GCAM model that extend its functionality and facilitate the data management. In particular, `rgcam` allows to both read and create a manageable data file containing the desired model outputs selected from the GCAM results database.

For the correct performance of `gcamreport` internal functions, the package includes user-modifiable constant values and mapping files, such as regional and sectoral aggregations, as well as external files to compute additional calculations to produce some results that are beyond the regular GCAM reporting (e.g., installed capacity). These additional calculations are necessary to meet the IAMC reporting requirements (since not all models provide the same type of outputs, but often through back-on-the-envelope calculations, harmonizing outputs in this way improves the comparability of model results). Mapping files are automatically loaded and used, but they can be customized and adjusted by the user if required by the specific study or GCAM version. In this case, it is recommended to save and tag the final version used to allow reusability and reproducibility.

The outputs generated by the functions within the *dataset generation block* consist of a Comma-Separated Values (CSV) file and a Microsoft Excel Open XML Spreadsheet (XLSX), both of which are automatically saved unless the user specifies otherwise with the `save_output` parameter. This ensures portability and format compatibility of the outputs (Krey, 2023). Besides, the outputs generated by the *interactive user block* are both CSV files and Portable Network Graphic (PNG) files, corresponding to the processed tabular data and the generated figures, which are only saved if the user clicks on the corresponding *download* button.

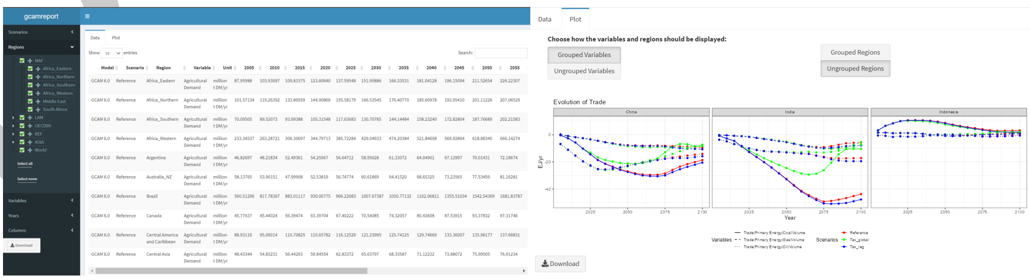

Figure 2: Figure 2: Interactive user interface visualization example.

Lastly, the package will be updated simultaneously with the official GCAM releases, promoting the reusability and transparency. In the same line, some new features are planned for the next releases, such as the reporting of hydrogen or refinery investments, or the generation of other types of plots in the user interface, in order to make the package more flexible and better adaptable to the needs of the scientific community. There are also plans to integrate the package into the GCAM-ecosystem suite of tools. Combining `gcamreport` with other tools in the ecosystem (such as gcamextractor) will allow GCAM users to obtain the model outputs in different formats, depending on their scientific needs.

Acknowledgements

C.R. and D.V. acknowledges financial support from the European Union's Horizon research program under grant agreement 101056306 (IAM COMPACT project).

References

Binsted, M., Iyer, G., Patel, P., Graham, N. T., Ou, Y., Khan, Z., Kholod, N., Narayan, K., Hejazi, M., & Kim, S. (2022). GCAM-USA v5. 3_water_dispatch: Integrated modeling of subnational US energy, water, and land systems within a global framework. *Geoscientific Model Development*, *15*(6), 2533–2559.

Boettiger, C. (2015). An introduction to docker for reproducible research, with examples from the r environment. *ACM SIGOPS Operating Systems Review*, *49*(1), 71–79. https://doi.org/10.1145/2723872.2723882

Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y., Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., Hejazi, M., Horowitz, R., Iyer, G., Kyle, P., Kim, S., Link, R., McJeon, H., Smith, S. J., Snyder, A., … Wise, M. (2019). GCAM v5.1: Representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development*, *12*(2), 677–698. https://doi.org/10.5194/gmd-12-677-2019

Guivarch, C., Le Gallic, T., Bauer, N., Fragkos, P., Huppmann, D., Jaxa-Rozen, M., Keppo, I., Kriegler, E., Krisztin, T., Marangoni, G., Pye, S., Riahi, K., Schaeffer, R., Tavoni, M., Trutnevyte, E., Van Vuuren, D., & Wagner, F. (2022). Using large ensembles of climate change mitigation scenarios for robust insights. *Nature Climate Change*, *12*(5), 428–435. https://doi.org/10.1038/s41558-022-01349-x

Huppmann, D., Wienpahl, L., Hackstock, P., & Castella, L. (2023). *Nomenclature* (Version v0.9.1). Zenodo. https://doi.org/10.5281/zenodo.7956229

Krey, V. (2023). *IAMC Webinar: Standardization of scenario data collection via the IAMC data template: Past developments and future challenges.* https://www.iamconsortium.org/event/iamc-webinar-standardization-of-scenario-data-collection-via-the-iamc-data-template-past-dev

Krey, V., Masera, O., Blanford, G., Bruckner, T., Cooke, R., Fisher-Vanden, K., Haberl, H., Hertwich, E., Kriegler, E., & Mueller, D. (2014). *Annex 2-metrics and methodology.*

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., & Jenkyns, R. (2020). The TRUST principles for digital repositories. *Scientific Data*, *7*(1), 144.

Masson-Delmotte, V., Pörtner, H.-O., Skea, J., Zhai, P., Roberts, D., Shukla, P. R., Pirani, A., Pidcock, R., Chen, Y., Lonnoy, E., Moufouma-Okia, W., Péan, C., Connors, S., Matthews, J. B. R., Zhou, X., Gomis, M. I., Maycock, T., Tignor, M., & Waterfield, T. (2018). *An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.*

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., & Gomis, M. (2021). Climate change 2021: The physical science basis.

Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, *2*.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., & Kram, T. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, *463*(7282), 747–756.

Nikas, A., Gambhir, A., Trutnevyte, E., Koasidis, K., Lund, H., Thellufsen, J. Z., Mayer, D., Zachmann, G., Miguel, L. J., Ferreras-Alonso, N., Sognnaes, I., Peters, G. P., Colombo, E., Howells, M., Hawkes, A., Van Den Broek, M., Van De Ven, D. J., Gonzalez-Eguino, M., Flamos, A., & Doukas, H. (2021). Perspective of comprehensive and comprehensible multi-model energy and climate science in Europe. *Energy*, *215*, 119153. https://doi.org/10.1016/j.energy.2020.119153

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., & Van Vuuren, D. P. (2014). A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Climatic Change*, *122*, 387–400.

Richters, O., Bertram, C., Kriegler, E., Al Khourdajie, A., Cui, R., Edmonds, J., Hackstock, P., Holland, D., Hurst, I., & Kikstra, J. (2022). *NGFS Climate Scenarios Data Set Version 3.0*.

Sampedro, J., Iyer, G., Msangi, S., Waldhoff, S., Hejazi, M., & Edmonds, J. A. (2022). Implications of different income distributions for future residential energy demand in the US. *Environmental Research Letters*, *17*(1), 014031.

Shukla, J., Skea, R., Slade, A., Al Khourdajie, R, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, & et al. (2022). Climate change 2022 mitigation of climate change. *Cambridge University Press, Cambridge, UK and New York, NY, USA*. https://doi.org/10.1017/9781009157926

Skea, J., Shukla, P., Al Khourdajie, A., & McCollum, D. (2021). Intergovernmental panel on climate change: Transparency and integrated assessment modeling. *WIREs Climate Change*, *12*(5). https://doi.org/10.1002/wcc.727

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, *115*(11), 2584–2589.

Van Beek, L., Hajer, M., Pelzer, P., Van Vuuren, D., & Cassen, C. (2020). Anticipating futures through models: The rise of Integrated Assessment Modelling in the climate science-policy interface since 1970. *Global Environmental Change*, *65*, 102191. https://doi.org/10.1016/j.gloenvcha.2020.102191

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, & Bourne, P. E. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18