

# gcamreport: An R tool to process and standardize GCAM outputs

Clàudia Rodés-Bachs<sup>1</sup>, Jon Sampedro<sup>1</sup>, Dirk-Jan Van de Ven<sup>1</sup>, Ryna Yiyun Cui<sup>2</sup>, Alicia Zhao<sup>2</sup>, Matthew Zwerling<sup>2</sup>, and Zarrar Khan<sup>3</sup>

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

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [Open Journals](#)

## Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#))

## Summary

There is an urgent need to perform multi-model studies to characterize uncertainty arising from models' heterogeneity 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-regarded model that has been extensively used for different international and national scenario analysis, the harmonization code has never been documented nor standardized, making it difficult to reproduce outputs and hindering the transparency of 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).

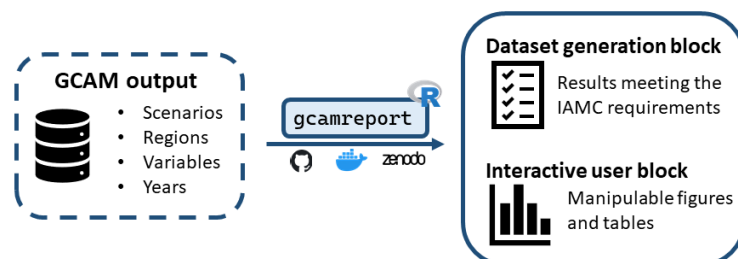


Figure 1: Structure of the gcamreport package.

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 a detailed [step-by-step tutorial](#) or this simplified code:

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

# run the main function to produce the dataset report meeting the IAMC
# requirements and launch the user interface
db_path <- "path/to/my-gcam-database"
db_name <- "my-gcam-database"
prj_name <- "my-project.dat"
scen_name <- "name of the GCAM scenarios"
generate_report(db_path = db_path, db_name = db_name, prj_name = prj_name,
scenarios = scen_name, final_year = 2100, save_output = TRUE, launch_ui = TRUE)
```

The additional instructions to run the package through a Docker image and using further functionalities can be found in the [online documentation](#) for all the release versions.

## 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 prominent 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 1.5°C (Masson-Delmotte et al., 2018) and the recent climate scenarios of the Network for Greening the Financial System (NGFS) (Richters et al., 2022). For these 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 by 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).

Here we present gcamreport, 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

70 results directly applicable to multi-IAM studies (including upcoming IPCC reports). Moreover,  
 71 to overcome the limitations to reproduce the execution environment (e.g., packages or libraries),  
 72 gcamreport is also available within a [Docker](#) image, which already contains all the required  
 73 packages and facilitates the usage, ensuring interoperability and reproducibility ([Boettiger,](#)  
 74 [2015](#)).

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

## 81 **Functionality**

82 The gcamreport package consists of a set of functions divided into two different blocks. All  
 83 functions are described in a specific [documentation](#) supplemented by detailed tutorials.

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

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

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

102 The outputs generated by the functions within the *dataset generation block* consist of a RData  
 103 dataset, a Comma-Separated Values (CSV) file, and a Microsoft Excel Open XML Spreadsheet  
 104 (XLSX), all of which are automatically saved unless the user specifies otherwise with the  
 105 `save_output` parameter. This ensures portability and format compatibility of the outputs  
 106 ([Krey, 2023](#)). Besides, the outputs generated by the *interactive user block* are both CSV files  
 107 and Portable Network Graphic (PNG) files, corresponding to the processed tabular data and  
 108 the generated figures, which are only saved if the user clicks on the corresponding *download*  
 109 button. This last block can be directly run with the RData dataset generated by the former  
 110 block, avoiding to create again the reporting dataset as detailed in this [example](#).

111 (a)

gcamreport

Scenarios

Regions

Variables

Years

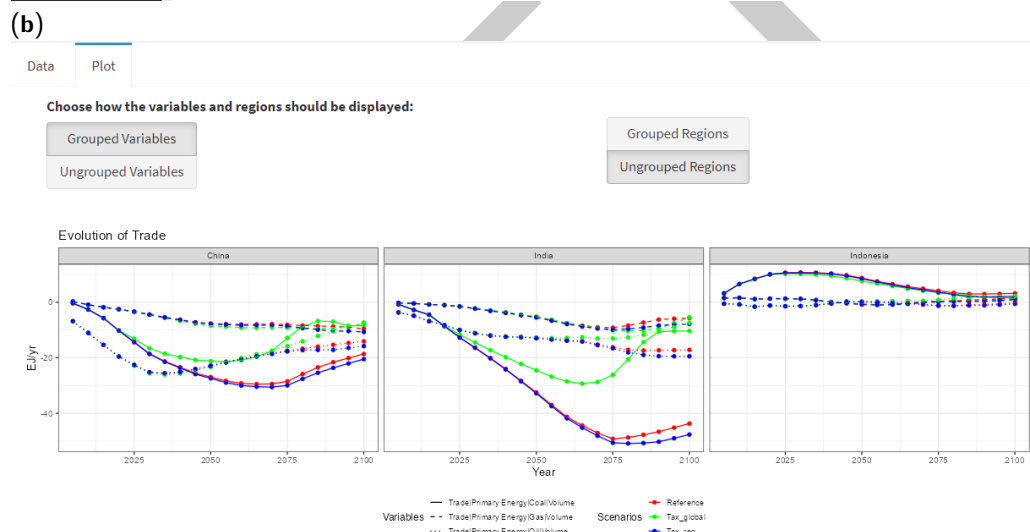
Columns

Download

Show 10 entries

Search:

Model	Scenario	Region	Variable	Unit	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050	2055
GCAM 6.0	Reference	Africa_Eastern	Agricultural Demand	million t DM/yr	87.95988	103.93097	109.81575	123.00640	137.59548	151.90886	166.33531	181.04128	196.15094	211.52654	226.22307
GCAM 6.0	Reference	Africa_Northern	Agricultural Demand	million t DM/yr	101.57134	119.26392	133.89559	144.90869	155.58179	166.53545	176.40770	185.60978	193.95410	201.11226	207.06529
GCAM 6.0	Reference	Africa_Southern	Agricultural Demand	million t DM/yr	70.09595	89.52073	93.99388	105.31548	117.63663	130.70785	144.14484	158.23240	172.82804	187.76689	202.21583
GCAM 6.0	Reference	Africa_Western	Agricultural Demand	million t DM/yr	233.34537	263.28721	306.30697	344.79713	385.72284	429.04933	474.20384	521.84698	569.92864	618.88340	666.14274
GCAM 6.0	Reference	Argentina	Agricultural Demand	million t DM/yr	46.82697	48.21834	52.49361	54.25967	56.64712	58.95628	61.33072	64.04901	67.12997	70.01431	72.18674
GCAM 6.0	Reference	Australia_NZ	Agricultural Demand	million t DM/yr	58.13765	53.90151	47.99908	52.53819	56.74774	60.61869	64.41520	68.65325	73.23565	77.53459	81.16281
GCAM 6.0	Reference	Brazil	Agricultural Demand	million t DM/yr	560.51296	817.78367	883.01117	930.00775	966.22683	1007.67387	1050.77132	1162.06811	1355.51034	1542.54369	1681.83787
GCAM 6.0	Reference	Canada	Agricultural Demand	million t DM/yr	45.77637	45.44024	56.39474	63.39704	67.40222	70.54085	74.32057	80.43608	87.53915	93.37832	97.31746
GCAM 6.0	Reference	Central America and Caribbean	Agricultural Demand	million t DM/yr	88.93116	95.00014	110.70825	110.65782	116.12529	121.23995	125.74125	129.74669	133.36807	135.96177	137.66831
GCAM 6.0	Reference	Central Asia	Agricultural Demand	million t DM/yr	48.43344	54.83231	56.44293	59.84554	62.83372	65.63797	68.33587	71.12232	73.88072	75.99505	76.91234



**Figure 2:** Interactive user interface visualization example. In a), see a tabular example and in b) see a plotting example.

The package has been optimized but still needs some time to run, basically to create the project file from the given database. In case the project file is already given, it requires less than a minute to produce the reporting dataset.

Lastly, the package will be updated simultaneously with the official GCAM releases, promoting the reusability and transparency. To keep track of the versions, there will be a gcamreport release by each GCAM version, tagged and named as documented [here](#). In the same line, some new features are planned for 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., Blandford, 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.

- 175 Nikas, A., Gambhir, A., Trutnevyte, E., Koasidis, K., Lund, H., Thellufsen, J. Z., Mayer, D.,  
176 Zachmann, G., Miguel, L. J., Ferreras-Alonso, N., Sognaes, I., Peters, G. P., Colombo,  
177 E., Howells, M., Hawkes, A., Van Den Broek, M., Van De Ven, D. J., Gonzalez-Eguino,  
178 M., Flamos, A., & Doukas, H. (2021). Perspective of comprehensive and comprehensible  
179 multi-model energy and climate science in Europe. *Energy*, 215, 119153. <https://doi.org/10.1016/j.energy.2020.119153>  
180
- 181 O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., &  
182 Van Vuuren, D. P. (2014). A new scenario framework for climate change research: The  
183 concept of shared socioeconomic pathways. *Climatic Change*, 122, 387–400.
- 184 Richters, O., Bertram, C., Kriegler, E., Al Kourdajie, A., Cui, R., Edmonds, J., Hackstock,  
185 P., Holland, D., Hurst, I., & Kikstra, J. (2022). *NGFS Climate Scenarios Data Set Version*  
186 *3.0*.
- 187 Sampedro, J., Iyer, G., Msangi, S., Waldhoff, S., Hejazi, M., & Edmonds, J. A. (2022).  
188 Implications of different income distributions for future residential energy demand in the  
189 US. *Environmental Research Letters*, 17(1), 014031.
- 190 Shukla, J., Skea, R., Slade, A., Al Kourdajie, R., van Diemen, D., McCollum, M., Pathak, S.,  
191 Some, P., Vyas, R., Fradera, M., Belkacemi, A., Hasija, G., Lisboa, S., Luz, J., Malley, & et al.  
192 (2022). Climate change 2022 mitigation of climate change. *Cambridge University Press*,  
193 *Cambridge, UK and New York, NY, USA*. <https://doi.org/10.1017/9781009157926>
- 194 Skea, J., Shukla, P., Al Kourdajie, A., & McCollum, D. (2021). Intergovernmental panel  
195 on climate change: Transparency and integrated assessment modeling. *WIREs Climate*  
196 *Change*, 12(5). <https://doi.org/10.1002/wcc.727>
- 197 Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for  
198 computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11),  
199 2584–2589.
- 200 Van Beek, L., Hajer, M., Pelzer, P., Van Vuuren, D., & Cassen, C. (2020). Anticipating  
201 futures through models: The rise of Integrated Assessment Modelling in the climate  
202 science-policy interface since 1970. *Global Environmental Change*, 65, 102191. <https://doi.org/10.1016/j.gloenvcha.2020.102191>  
203
- 204 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A.,  
205 Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, & Bourne, P. E. (2016). The FAIR  
206 guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1),  
207 160018. <https://doi.org/10.1038/sdata.2016.18>