COSIMA Cookbook: a community-driven resource for ocean and sea-ice modelling science

Navid C. Constantinou[3,14,*], Julia Neme[1,*], Wilton Aguiar[1], Matthis Auger[13], Ashley J. Barnes[3,10], Romain Beucher[1,2], Dhruv Bhagtani[11], Christopher Yit Sen Bull[1,2], Hannah Dawson[17], Noah Day[14], Fabio Boeira Dias[17], Edward W. Doddridge[17], Anupiya Ellepola[1], Denisse Fierro-Arcos[17], Angus Gibson[1], Aidan Heerdegen[1,2], Andy McC. Hogg[1,2,3], Ryan M. Holmes[4], Maurice F. Huguenin[8], Wilma Huneke[10], Jemma Jeffree[1,3], Andrew E. Kiss[1], Minghang Li[1,2], Paige Martin[1,2], Josué Martínez-Moreno[12], Jan Jaap Meijer[17], Thomas Moore[6], Ruth Moorman[5], Adele K. Morrison[1], James Munroe[7], Aditya Narayanan[16], Micael Oliveira[1,2], Ellie Q. Y. Ong[3,10], Madelaine G. Rosevear[3,17], Christina Schmidt[12], Taimoor Sohail[14], Paul Spence[3,17], Dougal T. Squire[1,2], Anton Steketee[1,2], Charles Turner[1,2], Felipe Vilela da Silva[9], Marc White[1,2], Luwei Yang[1], Claire Yung[1], Jan Zika[15]

[1]Australian National University, Australia
[2]Australia's Climate Simulator (ACCESS-NRI), Australia
[3]Australian Research Council Centre of Excellence for the Weather of the 21st Century, Australia
[4]Bureau of Meteorology, Australia
[5]California Institute of Technology, USA
[6]Commonwealth Scientific and Industrial Research Organisation, Australia
[7]Earthdata Associates, Canada
[8]GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany
[9]IFREMER - French Institute for Ocean Science, France
[10]Monash University, Australia
[11]Princeton University, USA
[12]British Antarctic Survey, UK
[13]University of Brest, France
[14]University of Melbourne, Australia
[15]University of New South Wales Sydney, Australia
[16]University of Southampton, UK
[17]University of Tasmania, Australia
[*]These authors contributed equally.

Software: github.com/COSIMA/cosima-recipes

Abstract

The COSIMA Cookbook is an open educational resource for analysing ocean and sea-ice model output in Jupyter notebooks. It has been developed by the Consortium for Ocean–Sea Ice Modelling in Australia (COSIMA) as a community resource for researchers, students, and practitioners working with large gridded datasets, especially output from the ACCESS ocean–sea ice model configurations. The Cookbook is organised around a gastronomy theme: "Cooking Tutorials" teach generic, transferable techniques for handling model output, while "Easy Recipes", "Advanced Recipes" and "Regional Specialties" are self-contained notebooks that show how to perform concrete diagnostics and analyses. Along the way, the notebooks demonstrate open-source tools for the scientific analysis of large data, including Intake catalogues, Xarray, Dask, xgcm, xESMF, eofs, Parcels, matplotlib and Cartopy. The Cookbook addresses the steep entry cost of analysing climate model output: it packages expert, often tacit, knowledge into complete, well-documented and reusable workflows, in the same medium in which researchers actually work. It is developed openly on GitHub, is published as a documentation website, and is continuously reviewed and extended by its community through pull requests and regular hackathons. The easy access to data and analysis it provides has contributed to the international adoption of ACCESS ocean and sea-ice model configurations, and has facilitated well over 100 peer-reviewed papers and more than 20 PhD projects.

1 Summary

The COSIMA Cookbook is an open educational resource for analysing ocean and sea-ice model output in Jupyter notebooks (Kluyver et al., 2016). It has been developed by the Consortium for Ocean–Sea Ice Modelling in Australia (COSIMA; https://cosima.org.au) as a community resource for researchers, students, and practitioners working with large gridded datasets, especially output from the ACCESS ocean–sea ice model configurations (e.g. Kiss et al., 2020). The Cookbook is developed openly in a GitHub repository, with all content exposed through a documentation site built with Sphinx (Komiya et al., 2025); see Figure 1. Tagged versions of the repository are uploaded to a citable archived release (Constantinou et al., 2026).

The Cookbook grew out of a loosely organised collection of Python scripts and Jupyter notebooks—many of which were broken. During the first COVID-19 lockdown in 2020, the community held its first hackathon (Martínez-Moreno and Constantinou, 2020): a full-day online sprint to bring those notebooks up to date. Annual workshops and hackathons since then have gradually shaped the repository into its current curated state.



Figure 1: The COSIMA Cookbook documentation website, showing the browsable Sphinx-based landing page used to navigate tutorials and recipes. The website lives at https://cosima-recipes.readthedocs.io.

The COSIMA Cookbook is deliberately organized using a gastronomy theme to help users quickly understand and navigate the website. "Cooking Tutorials" refers to tutorials that teach generic, transferable techniques for handling ocean–sea ice model output. These tutorials give the users the foundations for understanding the analysis done in the examples. Then comes the main part of any cookbook – its recipes. Here, by "recipes" we mean self-contained notebooks showing how to perform concrete diagnostics and analyses on ocean and sea-ice datasets. First come the "Easy

Recipes" that provide a good starting point after the tutorials; then the "Advanced Recipes" that are more elaborate and demonstrate advanced analysis techniques. Lastly, the "Regional Specialties" contains recipes for regional model configurations (Barnes et al., 2024). This gastronomy framing is shared by other computational workflow collections, such as Project Pythia Cookbooks for the general geoscience community and the FAIR Cookbook for working with life science data.

Pedagogy is central to the project. "Cooking Tutorials" introduces generic skills such as loading model output, working with labelled arrays, plotting, and interacting with shared data catalogues. These tutorials lead into domain-focused recipes giving learners an incremental path from first contact with the data ecosystem to adaptation of full workflows for their own science questions. The recipes provide the starting point for more elaborate analyses. The categories of "easy" and "advanced" recipes provide a lightweight pedagogical cue about expected complexity and scope.

Within this structure, the tutorials leverage and demonstrate open source tools for scientific analysis of large data. This includes examples of loading model output through Intake-based catalogues (Intake developers, 2026), using Xarray (Hoyer and Hamman, 2017) and Dask (Dask Development Team, 2016; Rocklin, 2015) to process very large multi-dimensional datasets in parallel. The recipe collection includes a Lagrangian particle-tracking workflow built with Parcels (Delandmeter and van Sebille, 2019), as well as analyses that use xgcm (Abernathey et al., 2022) to handle variables defined on staggered finite-volume grids, xESMF (Zhuang et al., 2025) for regridding model output, and eofs (Dawson, 2016) for empirical orthogonal function analysis. Figures are generated using matplotlib (Hunter, 2007) incorporating Cartopy (Elson et al., 2024) for map-based visualisation.

The Cookbook grew from a practical need inside the COSIMA community (COSIMA Community, 2026): powerful tools for analysing modern ocean model output already existed, but the gap between package-level documentation and reproducible end-to-end workflows left new users largely on their own. Productive use requires fluency not only in Python and Jupyter (Kluyver et al., 2016), but also in navigating high-dimensional, often very large model output, shared high-performance computing environments, and domain-specific analysis conventions. The Cookbook bridges this gap by packaging reusable workflows in the same medium in which researchers actually work.

2 Statement of Need

Analysing or configuring climate models carries a steep entry cost. New users must learn how to find datasets, load them efficiently, interpret metadata, operate on multi-dimensional arrays, and produce scientifically meaningful diagnostics. General-purpose libraries such as Xarray (Hoyer and Hamman, 2017) provide the foundations for these tasks, but they do not by themselves show a learner how to translate a research question into a robust analysis workflow for a specific model and computing environment. The Cookbook serves as a learning tool for those new to ocean and sea ice science, showing how to interrogate model output and visualise commonly used diagnostics—skills that transfer broadly across models and computing environments.

The gap is not only pedagogical; it directly slows science. Before any analysis can begin, a newcomer typically spends substantial time deciphering the structure of the output, the relevant conventions, and the computational environment. There is also substantial scientific labour embedded in many common diagnostics: some calculations require non-trivial development, understanding of model numerics, and validation before they can be used with confidence—these are the "Advanced Recipes". If every new user had to reconstruct those workflows independently, enormous effort would be spent rebuilding analysis code rather than doing science.

The COSIMA Cookbook facilitates knowledge sharing and accelerates research with a domain-specific, openly maintained collection of computational lessons and examples. Its contribution is not a new analysis library; rather, it is a reusable scientific and educational resource that exposes complete, well-documented workflows in the same notebook format in which many researchers actually explore data and communicate methods. By choosing notebooks instead of packaging

these examples only as a software library, the project makes intermediate reasoning, methodological choices, and practical execution details visible to learners while still giving more experienced users analysis patterns they can adapt directly for research. This aligns well with JOSE's emphasis on open educational resources that enable computational learning through authentic practice.

Several design decisions have made the Cookbook particularly useful for adoption by other groups. First, each notebook is intended to be self-contained and well-documented, so learners can read, run, and modify a complete workflow without reconstructing missing context from scattered notes. Second, the repository distinguishes tutorials from recipes. Tutorials teach transferable skills, while recipes demonstrate how those skills combine in realistic analysis tasks. Third, the project includes contributor guidance and notebook review conventions that encourage new analyses to be written in a reusable and pedagogically clear style. A caveat is that most recipes use high-resolution model output with datasets of order terabytes, so the underlying data are not always practical to distribute or mirror in full. However, beyond the data-loading step, which usually happens early on in each recipe, the analysis workflows are generally applicable to most operational ocean–sea ice model outputs that can be loaded with Xarray (Hoyer and Hamman, 2017).

Thus, the design renders the Cookbook valuable beyond the immediate COSIMA context. Many research communities maintain model output on shared infrastructure and face the same challenge: turning expert tacit knowledge into examples that newcomers can adapt. The Cookbook offers a model for how a scientific collaboration can capture that knowledge in version-controlled notebooks, publish it as a living open resource online, and continuously improve it through community contribution.

The rise of AI-assisted coding tools adds a further dimension to the Cookbook's value—and arguably makes it more important, not less. Such tools can rapidly generate analysis code, and they produce far better results when given a well-structured, domain-specific recipe as context rather than starting from scratch. But the deeper point is this: as AI-generated code becomes routine, the critical skill shifts from *writing* code to *understanding and evaluating* it. A researcher who has worked through the Cookbook knows why a workflow is structured the way it is—which parameters matter, where scientific assumptions enter, and what a plausible result looks like. That understanding is what allows AI-generated code to be trusted, corrected, and adapted for specific science questions. Resources that teach the reasoning behind an analysis, not just its mechanics, therefore grow more valuable alongside AI adoption.

3 Educational Design and Experience of Use

As we have elaborated above, the learning experience is organised as a progression. The tutorials and recipes showcase Jupyter-based analysis that can be run directly on Australia's National Computational Infrastructure, and include guidance on running notebooks through JupyterLab sessions and on accessing shared data holdings. This makes the Cookbook more than a static collection of examples: it is operational documentation for a real analysis environment. At the same time, the notebooks demonstrate broadly transferable workflows for working with labelled geophysical data, so individual lessons can be reused or adapted for specific research questions or for use on other computational platforms.

The project also supports community authorship. Contributors are encouraged to submit new notebooks through pull requests, document their methods clearly, and generalise workflows so that they remain useful to others. This is particularly valuable as it is relatively rare for the computer code underlying scientific analyses to be peer reviewed with the same care as the manuscript itself. The COSIMA community provides several examples of the recipe ecosystem being extended into full research projects built on peer-authored software repositories on GitHub, where analysis workflows, methods, and implementation choices are exposed to community scrutiny and reuse. In that sense, the Cookbook functions both as a learning resource for end users and as a framework for teaching reproducible scientific communication through notebook design and peer review. During regularly

organised hackathons, the community reviews, updates, and extends existing recipes, ensuring the Cookbook remains a living resource that keeps pace with evolving tools and community best practices. The easy access to data and analysis provided by the COSIMA Cookbook has enabled rapid adoption of ACCESS ocean and sea-ice model configurations internationally—not only within physical oceanography but also across adjacent disciplines such as marine ecology (Fierro-Arcos et al., 2023), where researchers have used it as an entry point to ACCESS model output. This accessibility has facilitated well over 100 peer-reviewed papers and more than 20 PhD projects to completion to date.

Author Order

The first two authors contributed equally; authors from the third position onward are listed alphabetically by surname.

Acknowledgements

This work was supported by computational resources provided by the Australian Government through the National Computational Infrastructure (NCI) under the National Computational Merit Allocation Scheme and the Australian National University Allocation Scheme. We thank the vibrant communities of the Consortium for Ocean–Sea Ice Modelling in Australia (COSIMA; cosima.org.au) and Australia's climate simulator (ACCESS-NRI; access-nri.org.au) for making the ACCESS-OM2 and ACCESS-OM3 outputs and analysis tools available through the NCI. We further acknowledge funding from the Australian Research Council under the Centres of Excellence for Climate Extremes CE170100023 and for the Weather of the 21st Century CE230100012, the Linkage Infrastructure, Equipment and Facilities LP160100073 and LP200100406, and the Discovery Project DP240101274.

References

Ryan P. Abernathey, Julius J. M. Busecke, Timothy Andrew Smith, Josephine Dianne Deauna, Anderson Banihirwe, Thomas Nicholas, Filipe Fernandes, Bourbeau James, Raphael Dussin, Deepak A. Cherian, Romain Caneill, Anirban Sinha, Leonardo Uieda, Willi Rath, Dhruv Balwada, Navid C. Constantinou, Aurélien Ponte, Yuxin Zhou, Takaya Uchida, and Jon Thielen. xgcm, November 2022. URL https://doi.org/10.5281/zenodo.7348619.

Ashley J. Barnes, Navid C. Constantinou, Angus H. Gibson, Andrew E. Kiss, Chris Chapman, John Reilly, Dhruv Bhagtani, and Luwei Yang. regional-mom6: A Python package for automatic generation of regional configurations for the Modular Ocean Model 6. *Journal of Open Source Software*, 9(100):6857, 2024. doi: 10.21105/joss.06857. URL https://doi.org/10.21105/joss.06857.

Navid C. Constantinou, Andy McC. Hogg, Angus Gibson, Anton Steketee, Romain Beucher, Max Proft, Aidan Heerdegen, Julia Neme, Ryan M. Holmes, Adele Morrison, Denisse Fierro-Arcos, Andrew E. Kiss, Micael Oliveira, Claire Yung, Edward Doddridge, Wilma Huneke, Dhruv Bhagtani, Ellie Q. Y. Ong, James Munroe, Christina Schmidt, Hannah Dawson, Dougal T. Squire, Fabio Boeira Dias, Jemma Jeffree, Ruth Moorman, Josué Martínez-Moreno, Jan Zika, Jan Jaap Meijer, Matthis Auger, Wilton Aguiar, Thomas Moore, Felipe Vilela da Silva, Luwei Yang, Marc White, Taimoor Sohail, Ashley J. Barnes, Paul Spence, Charles Turner, Madelaine G. Rosevear, Christopher Yit Sen Bull, Noah Day, Minghang Li, Anupiya Ellepola, and Aditya Narayanan. COSIMA Cookbook: A cookbook of recipes for analysing ocean and sea ice model output, 2026. URL https://doi.org/10.5281/zenodo.14353852.

COSIMA Community. COSIMA: The Movie. https://youtu.be/bmUYRoPKtIA, 2026. YouTube video.

Dask Development Team. *Dask: Library for dynamic task scheduling*, 2016. URL http://dask.pydata.org.

Andrew Dawson. eofs: A library for EOF analysis of meteorological, oceanographic, and climate data. *Journal of Open Research Software*, 4(1):e14, 2016. doi: 10.5334/jors.122.

Philippe Delandmeter and Erik van Sebille. The Parcels v2.0 Lagrangian framework: new field interpolation schemes. *Geoscientific Model Development*, 12(8):3571–3584, 2019. doi: 10.5194/gmd-12-3571-2019.

Phil Elson, Elliott Sales de Andrade, Greg Lucas, Ryan May, Richard Hattersley, Ed Campbell, Ruth Comer, Andrew Dawson, Bill Little, Stephane Raynaud, scmc72, Alan D. Snow, lgolston, Byron Blay, Peter Killick, lbdreyer, Patrick Peglar, Nat Wilson, Andrew, Jon Szymaniak, Adrien Berchet, Corinne Bosley, Luke Davis, Filipe, John Krasting, Matthew Bradbury, stephenworsley, and Daniel Kirkham. SciTools/cartopy: v0.24.1, October 2024. URL https://doi.org/10.5281/zenodo.13905945.

Denisse Fierro-Arcos, Stuart Corney, Amelie Meyer, Hakase Hayashida, Andrew E. Kiss, and Petra Heil. Analysis of ecologically relevant sea ice and ocean variables for the Southern Ocean using a high-resolution model to inform ecosystem studies. *Progress in Oceanography*, 215:103049, 2023. doi: 10.1016/j.pocean.2023.103049.

Stephan Hoyer and Joseph J. Hamman. xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5(1):10, 2017. doi: 10.5334/jors.148. URL https://doi.org/10.5334/jors.148.

John D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9 (3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Intake developers. *Intake: a lightweight package for finding, investigating, loading and disseminating data*, 2026. URL https://intake.readthedocs.io/en/reader/.

Andrew E. Kiss, Andrew McC. Hogg, Nicholas Hannah, Fabio Boeira Dias, Gary B. Brassington, Matthew A. Chamberlain, Christopher Chapman, Peter Dobrohotoff, Catia M. Domingues, Earl R. Duran, Matthew H. England, Russell Fiedler, Stephen M. Griffies, Aidan Heerdegen, Petra Heil, Ryan M. Holmes, Andreas Klocker, Simon J. Marsland, Adele K. Morrison, James Munroe, Maxim Nikurashin, Peter R. Oke, Gabriela S. Pilo, Océane Richet, Abhishek Savita, Paul Spence, Kial D. Stewart, Marshall L. Ward, Fanghua Wu, and Xihan Zhang. ACCESS-OM2 v1.0: a global ocean-sea ice model at three resolutions. *Geoscientific Model Development*, 13(2):401–442, 2020. doi: 10.5194/gmd-13-401-2020. URL https://doi.org/10.5194/gmd-13-401-2020.

Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing. Jupyter notebooks—a publishing format for reproducible computational workflows. In Fernando Loizides and Birgit Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90. IOS Press, Amsterdam, 2016. doi: 10.3233/978-1-61499-649-1-87. URL https://doi.org/10.3233/978-1-61499-649-1-87.

Takeshi Komiya, Georg Brandl, Adam Turner, Jean-Francois Bakosi, Takayuki Shimizukawa, Jakob Lykke Andersen, Daniel Neuhäuser, Stephen Finucane, Robert Lehmann, Jacob Mason, Timotheus Kampik, James Addison, Justus Magin, Jon Dufresne, Jonathan Waltman,

Daniel Eades, Juan Luis Cano Rodríguez, Armin Ronacher, Matthias Geier, Dmitry Shachnev, Yoshiki Shibukawa, Bénédikt Tran, Adam Dangoor, Chris Sewell, Rob Ruana, Pauli Virtanen, Tim Hoffmann, Martin Liška, and François Freitag. sphinx-doc/sphinx: Sphinx 9.1.0, 2025. URL https://doi.org/10.5281/zenodo.18109073.

Josué Martínez-Moreno and Navid C. Constantinou. Promotional video for COSIMA Hackathon April 2020. https://youtu.be/JO-_jaPW8S8, 2020. YouTube video.

Matthew Rocklin. Dask: Parallel computation with blocked algorithms and task scheduling. In Kathryn Huff and James Bergstra, editors, *Proceedings of the 14th Python in Science Conference*, pages 130–136, 2015. doi: 10.25080/Majora-7b98e3ed-013.

Jiawei Zhuang, Raphael Dussin, Pascal Bourgault, David Huard, Anderson Banihirwe, Stephane Raynaud, Brewster Malevich, Martin Schupfner, Charles Gauthier, Filipe, Sam Levang, Mattia Almansi, André Jüling, Richard Scott, RondeauG, Stephan Rasp, Trevor James Smith, Aaron G. Meyer, Ben Mares, Bill Sacks, Jemma Stachelek, Matthew Plough, Pierre, Ray Bell, Romain Caneill, and Xianxiang Li. pangeo-data/xESMF: v0.9.2, 2025. URL https://doi.org/10.5281/zenodo.4294774.