

# <sup>1</sup> ROMS-Tools: A Python Package for Preparing and Analyzing ROMS Simulations

<sup>3</sup> **Nora Loose**  <sup>1</sup>, **Tom Nicholas** <sup>2</sup>, **Scott Eilerman** <sup>1</sup>, **Christopher McBride** <sup>1</sup>, **Sam Maticka** <sup>1</sup>, **Dafydd Stephenson** <sup>1</sup>, **Scott Bachman** <sup>1</sup>, **Pierre Damien** <sup>3</sup>, **Ulla Heede** <sup>1</sup>, **Alicia Karspeck** <sup>1</sup>, **Matthew C. Long** <sup>1</sup>, **Jeroen Molemaker** <sup>3</sup>, and **Abigale Wyatt** <sup>1</sup>

<sup>7</sup> **1** [C]Worthy LLC, Boulder, CO, United States **2** Earthmover **3** University of California, Los Angeles, CA, United States

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Published: unpublished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](#)).<sup>33</sup>

## <sup>9</sup> Summary

<sup>10</sup> The ocean shapes Earth's climate and sustains marine ecosystems by circulating and storing vast amounts of heat, oxygen, carbon, and nutrients, while exchanging heat and gases with the atmosphere. To understand these complex dynamics and processes, scientists rely on ocean models, powerful computer simulations of physical circulation and biogeochemical (BGC) dynamics. These models represent the ocean on a grid of cells, where finer grid spacing (more, smaller cells) provides higher fidelity and greater detail but requires significantly more computing power. While global ocean models simulate the entire ocean, **regional ocean models** focus computational resources on a specific area to achieve much finer grid spacing than is computationally feasible over the global domain. This finer grid spacing enables regional ocean models to explicitly resolve fine-scale phenomena, like mesoscale (10-100 km) and submesoscale (0.1-10 km) features, tidal dynamics, coastal currents, upwelling, and detailed BGC processes. Capturing these dynamics and processes at high fidelity is essential for applications in environmental management, fisheries, for assessing regional impacts of climate change, and for evaluating ocean-based carbon dioxide removal (CDR) strategies.

<sup>24</sup> A widely used regional ocean model is the **Regional Ocean Modeling System (ROMS)** ([Shchepetkin & McWilliams, 2005](#)). To connect physical circulation with ecosystem dynamics and the ocean carbon cycle, ROMS has been coupled to a BGC model called the Marine Biogeochemistry Library (MARBL) ([Long et al., 2021](#)). This coupled framework allows researchers to explore a variety of scientific and practical questions. For example, it can be used to investigate the potential of ocean-based carbon removal strategies, such as adding alkaline materials to the ocean to sequester atmospheric carbon dioxide. It can also be used to study how physical processes drive ecosystem dynamics, such as how nutrient-rich waters from upwelling fuel the phytoplankton blooms that form the base of the marine food web ([Gruber et al., 2006](#)).

## <sup>34</sup> Input Data and Preprocessing

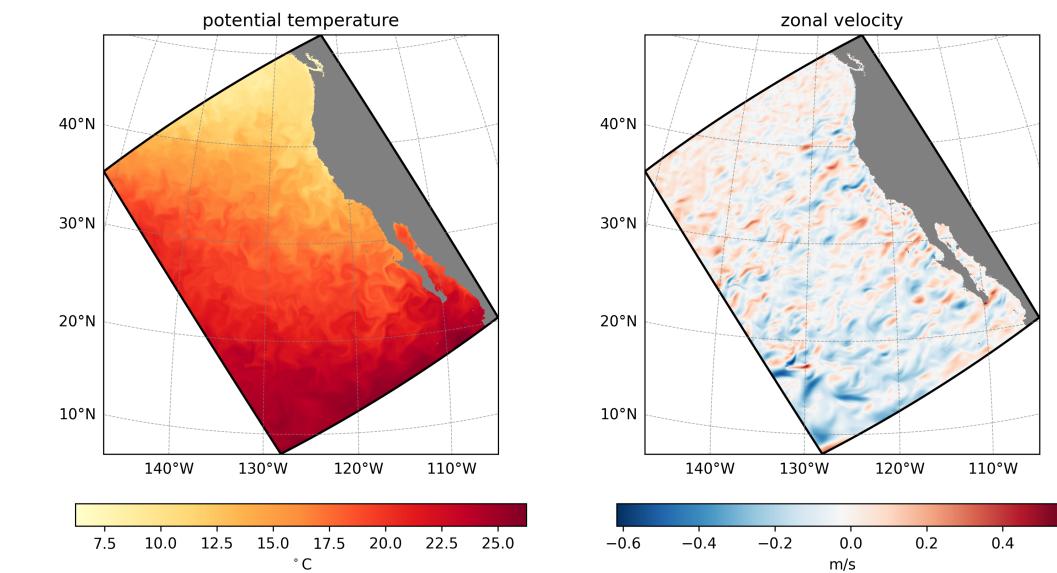
<sup>35</sup> Whether for research or industrial-focused applications, configuring a regional ocean model like ROMS-MARBL remains a major technical challenge. Generating the required input files is time-consuming, error-prone, and difficult to reproduce, creating a bottleneck for both new and experienced model users. The Python package ROMS-Tools addresses this challenge by providing an efficient, user-friendly, and reproducible workflow to generate all required model input files. Its user interface and underlying data model are based on xarray ([Hoyer & Hamman, 2017](#)), enabling seamless handling of multidimensional datasets with rich metadata and optional parallelization via a dask ([Dask Development Team, 2016](#)) backend.

43 ROMS-Tools can automatically process commonly used datasets or incorporate custom user  
44 data and routines. Currently, it can generate the following inputs:

- 45 1. **Model Grid:** Customizable, curvilinear grid, rotatable to align with coastlines, with a  
46 terrain-following vertical coordinate.
- 47 2. **Bathymetry:** Derived from **SRTM15** ([Tozer et al., 2019](#)).
- 48 3. **Land Mask:** Inferred from **Natural Earth** coastlines.
- 49 4. **Physical Ocean Conditions:** Initial and open boundary conditions for sea surface height,  
50 temperature, salinity, and velocities derived from GLORYS ([Lellouche et al., 2021](#)).
- 51 5. **BGC Ocean Conditions:** Initial and open boundary conditions for dissolved inorganic car-  
52 bon, alkalinity, and other biogeochemical tracers from CESM output (?-multiyear\_2022)  
53 or hybrid observational-model sources.
- 54 6. **Meteorological forcing:** Wind, radiation, precipitation, and air temperature/humidity  
55 processed from ERA5 ([Hersbach et al., 2020](#)) with optional radiation bias and coastal  
56 wind corrections.
- 57 7. **BGC surface forcing:** pCO<sub>2</sub>, iron, dust, nitrogen deposition from CESM output (?-mul-  
58 tiyear\_2022) or hybrid observational-model sources.
- 59 8. **Tidal Forcing:** Tidal potential, elevation, and velocities derived from **TPXO** ([Egbert &](#)  
60 [Erofeeva, 2002](#)) including self-attraction and loading (SAL) corrections.
- 61 9. **River Forcing:** Freshwater runoff derived from **Dai & Trenberth** ([Dai & Trenberth, 2002](#))  
62 or custom files.
- 63 10. **CDR Forcing:** User-defined interventions that inject BGC tracers at point sources or as  
64 larger-scale Gaussian perturbations, suitable for the simulation of field- or large-scale  
65 CDR experiments.

66 While the source datasets listed above are the ones currently supported, the package's modular  
67 design makes it straightforward to add new data sources or custom routines in the future. To  
68 generate the model inputs listed above, ROMS-Tools automates several intermediate processing  
69 steps, including:

70 **Bathymetry processing:** The bathymetry is smoothed in two stages, first across the entire  
71 domain and then along the shelf, to ensure local steepness ratios are not exceeded and  
72 to reduce pressure-gradient errors. A minimum depth is enforced to prevent water levels  
73 from becoming negative during large tidal excursions. **Mask definition:** The land-sea  
74 mask is generated by comparing the ROMS grid's horizontal coordinates with a coastline  
75 dataset using `regionmask` ([Hauser et al., 2024](#)). Enclosed basins are subsequently filled  
76 with land. **Land value handling:** Land values are filled via an algebraic multigrid method  
77 using `pyamg` ([Bell et al., 2023](#)) prior to horizontal regridding. This extends ocean values  
78 into land areas to resolve discrepancies between source data and ROMS land masks that  
79 could otherwise produce artificial values in ocean cells. **Regridding:** Ocean and atmospheric  
80 fields are horizontally and vertically regridded from standard lat-lon-depth grids to the  
81 model's curvilinear grid with a terrain-following vertical coordinate using `xarray` ([Hoyer &](#)  
82 [Hamman, 2017](#)). Optional sea surface height corrections can be applied, and velocities  
83 are rotated to align with the rotated ROMS grid. **Longitude conventions:** ROMS-Tools  
84 handles differences in longitude conventions, converting between -180°–180° and 0°–360° as  
85 needed. **River forcing:** Relevant rivers are automatically selected and relocated to the nearest  
86 coastal cell, while multi-cell or moving rivers can be managed manually. **Atmospheric data**  
87 **streaming:** ERA5 atmospheric data can be accessed directly from the cloud, removing the  
88 need for users to pre-download large datasets locally. Users can quickly design and visualize  
89 regional grids and inspect all input fields with built-in plotting utilities. An example of sur-  
90 face initial conditions generated for a California Current System simulation is shown in Figure ??.



91  
92 ROMS-Tools also includes features that facilitate simulation and output management. It  
93 supports partitioning and recombining input and output files to enable parallelized ROMS  
94 simulations across multiple nodes, and writes NetCDF outputs with metadata fully compatible  
95 with ROMS-MARBL. Currently, UCLA-ROMS ([Molemaker & contributors, 2025](#)) is fully  
96 supported, with the potential to add other ROMS versions, such as Rutgers ROMS ([Arango &](#)  
97 [contributors, 2024](#)), in the future.

## 98 Analysis Layer

99 ROMS-Tools includes an analysis layer for postprocessing ROMS-MARBL output. It provides  
100 utilities for general-purpose tasks, such as loading model output directly into an Xarray dataset  
101 with additional metadata, enabling seamless use of the Pangeo scientific Python ecosystem for  
102 further analysis and visualization. The analysis layer also supports regridding from the native  
103 curvilinear ROMS grid with terrain-following coordinate to a standard latitude-longitude-depth  
104 grid using xesmf ([Zhuang et al., 2023](#)). Beyond these general-purpose features, the analysis  
105 layer offers a suite of targeted tools for evaluating CDR interventions. These include utilities  
106 for generating standard plots, such as CDR uptake efficiency curves, and performing specialized  
107 tasks essential for CDR monitoring, reporting, and verification.

## 108 Workflow, Reproducibility, and Performance

109 ROMS-Tools is designed to support modern, reproducible workflows. It is easily installable via  
110 Conda or PyPI and can be run interactively from Jupyter Notebooks. To ensure reproducibility  
111 and facilitate collaboration, each workflow is defined in a simple YAML configuration file.  
112 These compact, text-based YAML files can be version-controlled and easily shared, eliminating  
113 the need to transfer large NetCDF files between researchers, as source data like GLORYS and  
114 ERA5 are accessible in the cloud.

115 For performance, the package is integrated with dask ([Dask Development Team, 2016](#)) to  
116 enable efficient, out-of-core computations on large datasets. Finally, to ensure reliability, the  
117 software is rigorously tested with continuous integration (CI) and supported by comprehensive  
118 documentation.

## 119 Statement of need

120 Setting up a regional ocean model is a major undertaking. It requires generating a wide range  
121 of complex input files, including the model grid, initial and boundary conditions, and forcing

122 from the atmosphere, tides, and rivers. Traditionally, this work has depended on a patchwork  
123 of custom scripts and lab-specific workflows, which can be time-consuming, error-prone, and  
124 difficult to reproduce. These challenges slow down science, create a steep barrier to entry for  
125 new researchers, and limit collaboration across groups.

126 Within the ROMS community, the preprocessing landscape has been shaped by tools like  
127 pyroms ([Hedstrom & contributors, 2023](#)). While pyroms has long provided valuable low-level  
128 utilities, it also presents challenges for new users. Installation can be cumbersome due to its  
129 Python and Fortran dependencies, and its inconsistent API and limited documentation make it  
130 hard to learn. The package was not designed with reproducible workflows in mind, and it lacks  
131 tests, continuous integration, and support for modern Python tools such as xarray and dask.  
132 Since development of pyroms has largely ceased, its suitability for new projects is increasingly  
133 limited. Importantly, tools from other modeling communities cannot simply be adapted, since  
134 each ocean model has distinct structural requirements. For example, the new regional-mom6  
135 package ([Barnes et al., 2024](#)), developed for MOM6 ([Adcroft et al., 2019](#)), cannot be used  
136 to generate ROMS inputs, because ROMS employs a terrain-following vertical coordinate  
137 system that requires a fundamentally different regridding approach, whereas MOM6 accepts  
138 inputs on arbitrary depth levels. Several other differences further prevent cross-compatibility.  
139 Together, these limitations underscored the need for a modern, maintainable, and reproducible  
140 tool designed specifically for ROMS.<sup>1</sup>

141 ROMS-Tools was developed to meet this need. It draws on the legacy of MATLAB preprocessing  
142 scripts developed at UCLA ([Molemaker, 2024](#)), which encapsulate decades of expertise in  
143 configuring regional ocean model inputs. While many of the core algorithms and design principles  
144 are retained, ROMS-Tools provides an open-source Python implementation using an object-  
145 oriented programming paradigm. This implementation enables a modernized workflow driven  
146 by high-level user Application Programming Interface (API) calls, enhancing reproducibility,  
147 reducing the potential for user errors, and supporting extensibility for additional features,  
148 forcing datasets, and use cases. In some cases, ROMS Tools diverges from the MATLAB  
149 implementation to take advantage of new methods or better integration with the modern  
150 Python ecosystem. By streamlining input generation and analysis, ROMS-Tools reduces technical  
151 overhead, lowers the barrier to entry, and enables scientists to focus on research rather than  
152 data preparation. The primary users of the package include ocean modelers developing new  
153 domains and researchers in the CDR community, who use it to test climate intervention  
154 scenarios.

## 155 Acknowledgements

156 Acknowledgement of any financial support.

## 157 References

- 158 Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., Dunne, J.  
159 P., Griffies, S. M., Hallberg, R., Harrison, M. J., Held, I. M., Jansen, M. F., John, J. G.,  
160 Krasting, J. P., Langenhorst, A. R., Legg, S., Liang, Z., McHugh, C., Radhakrishnan,  
161 A., ... Zhang, R. (2019). The GFDL Global Ocean and Sea Ice Model OM4.0: Model  
162 Description and Simulation Features. *Journal of Advances in Modeling Earth Systems*,  
163 11(10), 3167–3211. <https://doi.org/10.1029/2019MS001726>
- 164 Arango, H., & contributors. (2024). Rutgers ROMS. In *Github repository*. GitHub. <https://github.com/myroms/roms>

<sup>1</sup>In the future, packages like 'ROMS-Tools' and 'regional-mom6' could share a common backbone, with model-specific adaptations layered on top.

- 166 Barnes, A. J., Constantinou, N. C., Gibson, A. H., Kiss, A. E., Chapman, C., Reilly, J.,  
167 Bhagtani, D., & Yang, L. (2024). Regional-mom6: A Python package for automatic  
168 generation of regional configurations for the Modular Ocean Model 6. *Journal of Open  
169 Source Software*, 9(100), 6857. <https://doi.org/10.21105/joss.06857>
- 170 Bell, N., Olson, L. N., Schroder, J., & Southworth, B. (2023). PyAMG: Algebraic multigrid  
171 solvers in python. *Journal of Open Source Software*, 8(87), 5495. <https://doi.org/10.21105/joss.05495>
- 173 Dai, A., & Trenberth, K. E. (2002). *Estimates of Freshwater Discharge from Continents:  
174 Latitudinal and Seasonal Variations*. [https://journals.ametsoc.org/view/journals/hydr/3/6/1525-7541\\_2002\\_003\\_0660\\_eofdfc\\_2\\_0\\_co\\_2.xml](https://journals.ametsoc.org/view/journals/hydr/3/6/1525-7541_2002_003_0660_eofdfc_2_0_co_2.xml)
- 176 Dask Development Team. (2016). *Dask: Library for dynamic task scheduling*. <http://dask.pydata.org>
- 178 Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient Inverse Modeling of Barotropic Ocean  
179 Tides. *Journal of Atmospheric and Oceanic Technology*, 19(2), 183–204. [https://doi.org/10.1175/1520-0426\(2002\)019%3C0183:EIMOBO%3E2.0.CO;2](https://doi.org/10.1175/1520-0426(2002)019%3C0183:EIMOBO%3E2.0.CO;2)
- 181 Gruber, N., Frenzel, H., Doney, S. C., Marchesiello, P., McWilliams, J. C., Moisan, J. R.,  
182 Oram, J. J., Plattner, G.-K., & Stolzenbach, K. D. (2006). Eddy-resolving simulation of  
183 plankton ecosystem dynamics in the California Current System. *Deep Sea Research Part I:  
184 Oceanographic Research Papers*, 53(9), 1483–1516. <https://doi.org/10.1016/j.dsr.2006.06.005>
- 186 Hauser, M., Spring, A., Busecke, J., Driels, M. van, Lorenz, R., & readthedocs-assistant.  
187 (2024). *Regionmask/regionmask: Version 0.12.1*. Zenodo. <https://doi.org/10.5281/zenodo.10849860>
- 189 Hedstrom, K., & contributors. (2023). Pyroms. In *GitHub repository*. GitHub. <https://github.com/ESMG/pyroms>
- 191 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas,  
192 J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan,  
193 X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J.-N.  
194 (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*,  
195 146(730), 1999–2049. <https://doi.org/https://doi.org/10.1002/qj.3803>
- 196 Hoyer, S., & Hamman, J. (2017). Xarray: ND labeled arrays and datasets in python. *Journal  
197 of Open Research Software*, 5(1). <https://doi.org/10.5334/jors.148>
- 198 Lellouche, E., Jean-Michel Greiner, Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon,  
199 M., Bricaud, C., Hamon, M., Le Galloudec, O., Regnier, C., Candela, T., Testut, C.-E.,  
200 Gasparin, F., Ruggiero, G., Benkiran, M., Drillet, Y., & Le Traon, P.-Y. (2021). The  
201 Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. *Frontiers in Earth  
202 Science*, 9. <https://doi.org/10.3389/feart.2021.698876>
- 203 Long, M. C., Moore, J. K., Lindsay, K., Levy, M., Doney, S. C., Luo, J. Y., Krumhardt, K.  
204 M., Letscher, R. T., Grover, M., & Sylvester, Z. T. (2021). *Simulations With the Marine  
205 Biogeochemistry Library (MARBL)*. <https://doi.org/10.1029/2021MS002647>
- 206 Molemaker, J. (2024). UCLA MATLAB tools. In *GitHub repository*. GitHub. <https://github.com/nmolem/ucla-tools>
- 208 Molemaker, J., & contributors. (2025). UCLA-ROMS. In *GitHub repository*. GitHub. <https://github.com/CESR-lab/ucla-roms>
- 210 Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system  
211 (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model.  
212 *Ocean Modelling*, 9(4), 347–404. <https://doi.org/10.1016/j.ocemod.2004.08.002>

- 213 Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., & Wessel, P. (2019).  
214 Global Bathymetry and Topography at 15 Arc Sec: SRTM15+. *Earth and Space Science*,  
215 6(10), 1847–1864. <https://doi.org/10.1029/2019EA000658>
- 216 Zhuang, J., Dussin, R., Huard, D., Bourgault, P., Banhirwe, A., Raynaud, S., Malevich,  
217 B., Schupfner, M., Filipe, Levang, S., Gauthier, C., Jüling, A., Almansi, M., Scott,  
218 R., RondeauG, Rasp, S., Smith, T. J., Stachelek, J., Plough, M., & Li, X. (2023).  
219 xESMF: Universal regridder for geospatial data. In *GitHub repository*. Zenodo. <https://doi.org/10.5281/zenodo.4294774>  
220

DRAFT