R for marine ecologists: wrangling Earth System Model outputs

Authors: Jessica A. Bolin[1,2,3,*], Mercedes Pozo Buil[4], Nerea Lezama-Ochoa[4], Mikaela M. Provost[1,2]

[1] Coastal and Marine Sciences Institute, UC Davis, CA
[2] Department of Wildlife, Fish and Conservation Biology, UC Davis, CA
[3] School of Science, Technology and Engineering, University of the Sunshine Coast, QLD
[4] Institute of Marine Sciences, UC Santa Cruz, CA
[*] Corresponding author: jessica.anne.bolin@gmail.com

Github software repository: https://github.com/JessicaBolin/esmRworkshop_notes
Workshop notes: https://jessicabolin.quarto.pub/esmrworkshop_notes/

This manuscript is a non-peer reviewed preprint submitted to EcoEvoRxiv.

Summary

This one-day, hands-on workshop introduces marine ecologists to the use and application of Earth System Model (ESM) outputs using R–a free, open-source programming language widely used in marine ecology. Designed for participants with basic or no prior experience in climate modeling, our workshop: R for marine ecologists: wrangling Earth System Model outputs provides a comprehensive foundation in accessing, processing, and applying ESM outputs for marine ecology applications, by combining theoretical instruction with practical, live-coded sessions. Participants learn to download, regrid, statistically downscale, bias-correct, and visualize sea-surface temperature projections from two Coupled Model Intercomparison Project 6 (CMIP6) models across two climate scenarios. All course materials—including a fully documented Quarto website, annotated R scripts, datasets, and a public GitHub repository—are openly available and reusable for both workshop and individual use. Our ultimate goal is to make ESMs more accessible to marine ecologists, and empower the community to incorporate climate projections into their own ecological analyses.

Statement of Need

As the impacts of climate change on marine ecosystems intensify, there is a growing demand within the marine science community for tools that can integrate climate projections into ecological research. Earth System Models (ESMs), particularly those from the CMIP6 archive, provide essential insights into simulating future climate conditions. The

physical oceanography and climate modeling communities have developed robust coding infrastructure and workflows for working with ESMs and gridded ocean data (e.g., typically using MATLAB, Python, and Climate Data Operators (CDO; Irving 2019). However, these coding platforms are not commonly used within the marine ecology community, where R remains the dominant programming language (Lai et al. 2019), resulting in a critical skills gap (Asch et al. 2016; Halpin et al. 2014). ESM data, while highly relevant, often remain inaccessible to many marine ecologists due to the technical challenges involved in accessing and difficulty of processing model outputs.

Our workshop is designed to directly address this gap. Our goal is to make ESM data more accessible to marine ecologists by providing the technical expertise needed to work with these complex datasets in R. As such, this workshop is designed around Step 8, "Pre-process the data", of the proposed workflow for working with ESM outputs described in Schoeman et al. 2023. The workshop format combines guided theoretical instruction with hands-on, live-coded tutorials that walk participants through downloading, and statistically downscaling (i.e., increasing the spatial and/or temporal resolution of modelled data), bias-correcting (i.e., adjusting model outputs to reduce systematic differences from observed conditions) sea-surface temperature projections from two ESMs within CMIP6, across two climate scenarios. These preprocessing steps are crucial to ensure that model outputs are not only scientifically rigorous, but also ecologically interpretable and appropriate for decision-making at local to regional scales. Our workshop aims to reduce technical barriers and foster capacity-building within the marine ecology community by providing practical and reproducible tools and methods for working with ESM output using R (Kellie & Westgate 2023).

Learning objectives

The learning objectives span seven modules, each targeting a key concept:

1. Introduction to ESMs (theory)
    a. What is an ESM, how do they work, what is CMIP6, and how are they used in marine ecological studies?
    b. Understanding climate scenarios and how to select appropriate models.
2. Downloading ESM output (theory and live coding)
    a. Navigating the Earth System Grid Federation (ESGF) website and using shell scripts for downloading ESM data.
    b. Parallelizing downloads in R.
3. Regridding and statistically downscaling (live coding)
    a. Challenges of handling raw ESM spatial data in R.

    b. Tools and techniques for regridding and downscaling.
4. Observations (theory and live coding)
    a. Selecting observational data for bias correction.
    b. Regridding and downscaling observational datasets.
5. Bias correction (theory and live coding)
    a. Importance and methods of bias correction: the delta method.
    b. Consequences of uncorrected ESM outputs.
    c. Implementing bias correction with the delta method in R.
6. Evaluating ESM accuracy (live coding)
    a. Assessing model accuracy during the observational period.
    b. Common metrics and visualization approaches.
7. Making projections (live coding)
    a. Creating time series of projected SST to 2100.
    b. Mapping projections across multiple climate scenarios and time periods.

Experience of use in teaching and learning situations

The initial pilot of the workshop was delivered both in-person and virtually at Bodega Marine Laboratory, California, to a diverse audience including undergraduates, graduate students, postdocs, and researchers. A team of four instructors facilitated teaching, answered questions, and supported technical troubleshooting. Our design leverages worked examples and live coding to facilitate active learning, reduce cognitive load, and mirror real-world workflows. Learners alternate between conceptual grounding and applied exercises, enhancing retention and transfer.

Since the pilot, we have delivered the workshop to other university lab groups and adapted the content for online platforms such as R-Ladies webinars. Participant feedback based on online surveys were positive and have guided continuous improvements to the material. The workshop materials are now also used independently by individuals seeking to build expertise in working with ESM data in R.

Future directions

Moving forward, we aim to expand the reach of this workshop by (i) translating materials into additional languages, (ii) incorporating more climate variables and modeling scenarios, (iii) partnering with ecological and climate modeling groups globally to broaden access and training opportunities, and (iv) exploring the application of ESMs in marine ecology, such as through species distribution models, using guided examples. We hope that the materials herein can be collaboratively updated and improved (via the associated Github repository) by the marine ecology community.

How the workshop came to be

Jessica Bolin (Postdoctoral Scholar, University of California Davis) initiated the workshop following a research stay in Monterey, California in collaboration with the UCSC and NOAA Ecosystem group. There, in collaboration with Mercedes Pozo Buil (Associate Project Scientist, University of California Santa Cruz), Mikaela Provost (Assistant Professor, University of California Davis) and Nerea Lezama-Ochoa (Project Scientist, University of California Santa Cruz), conversations highlighted the steep learning curve marine ecologists face when attempting to work with ESM data. These discussions catalyzed the design of our dedicated workshop aimed at closing this skills gap. With financial support from the UC Davis Coastal and Marine Science Institute, we collaboratively developed the pilot workshop and delivered it in June 2025 at Bodega Marine Laboratory, California.

Acknowledgements

We thank the Coastal and Marine Science Institute (CMSI) of the University of California, Davis for awarding funding to deliver the pilot workshop via their internal grant scheme, in addition to Bodega Marine Laboratory for providing space and facilities. We also thank Barb Muhling, Alice Pidd and Mary Fisher for reviewing workshop code.

References

1. Asch, R., Pilcher, D., Rivero-Calle, S., & Holding, J. (2016). Demystifying Models: Answers to Ten Common Questions That Ecologists Have About Earth System Models. Limnology and Oceanography Bulletin, 25. https://doi.org/10.1002/lob.10113
2. Halpin, P., Boustany, A., Dunn, D., & Donnelly, J., B. Clearly. (2014). Incorporating climate data into the decision making process for living marine resources. A Survey Report Submitted to the NASA Earth Science Division.
3. Irving, D. (2019). Python for Atmosphere and Ocean Scientists. Journal of Open Source Education. https://doi.org/10.21105/jose.00037
4. Kellie, D., & Westgate, M. (2023). Improving code reproducibility: Small steps with big impacts. Ecology and Evolution. https://communities.springernature.com/posts/improving-code-reproducibility-small-steps-with-big-impacts
5. Lai, J., Lortie, C., Muenchen, R., Yang, J., & Ma, K. (2019). Evaluating the popularity of R in ecology. Ecosphere, 10. https://doi.org/10.1002/ecs2.2567
6. Schoeman, D., Gupta, A., Harrison, C., Everett, J., Brito-Morales, I., Hannah, L., Bopp, L., Roehrdanz, P., & Richardson, A. (2023). Demystifying Global Climate Models for Use in the Life Sciences. Trends in Ecology and Evolution, 38. https: //doi.org/10.1016/j.tree.2023.04.005