arXiv:2611.30182v1 [physics.ao-ph] 23 Nov 2026

COSIMA Cookbook：海洋与海冰建模科学的社区协作资源

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]澳大利亚国立大学，澳大利亚[2]澳大利亚气候模拟器（ACCESS-NRI），澳大利亚[3]澳大利亚研究理事会二十一世纪天气卓越中心，澳大利亚[4]澳大利亚气象局，澳大利亚[5]加州理工学院，美国[6]澳大利亚联邦科学与工业研究组织，澳大利亚[7]Earthdata Associates，加拿大[8]基尔亥姆霍兹海洋研究中心，德国[9]法国海洋开发研究院，法国[10]莫纳什大学，澳大利亚[11]普林斯顿大学，美国[12]英国南极考察局，英国[13]布雷斯特大学，法国[14]墨尔本大学，澳大利亚[15]新南威尔士大学悉尼分校，澳大利亚[16]南安普顿大学，英国[17]塔斯马尼亚大学，澳大利亚[*]这些作者的贡献相同。

软件： github.com/COSIMA/cosima-recipes

摘要

COSIMA Cookbook 是一项开放教育资源，旨在通过 Jupyter 笔记本分析海洋与海冰模型输出。该资源由澳大利亚海洋—海冰建模联盟（COSIMA）开发，面向使用大型网格数据集（尤其是 ACCESS 海洋—海冰模型配置输出）的研究人员、学生和从业者，作为社区资源使用。Cookbook 以美食为主题进行组织："烹饪教程"讲授处理模型输出的通用可迁移技术，而"简易食谱""进阶食谱"和"区域特色"则是独立的笔记本，展示如何

以开展具体的诊断分析。与此同时，这些笔记本展示了用于大型数据科学分析的开源工具，包括 Intake目录、Xarray、Dask、xgcm、xESMF、eofs、Parcels、matplotlib 和 Cartopy。Cookbook 降低了分析气候模型输出的高门槛：它将专家掌握的、往往不言自明的知识，整理成完整、文档完善且可重复使用的工作流程，并采用研究人员实际工作的媒介形式。该项目在 GitHub 上以开放方式开发，以文档网站形式发布，并由社区通过拉取请求和定期举办的编程马拉松持续审阅和扩展。它让数据和分析更易于获取，推动了 ACCESS 海洋和海冰模型配置在国际上的应用，并促成了 100 多篇同行评审论文和 20 多个博士项目。

1 总结

COSIMA Cookbook 是一项开放教育资源，旨在通过 Jupyter 笔记本分析海洋与海冰模型输出（Kluyver 等人，2016）。它由澳大利亚海洋与海冰建模联盟（COSIMA；https://cosima.org.au）开发，作为一项社区资源，服务于处理大型网格数据集的研究人员、学生和从业者，尤其是使用 ACCESS 海洋—海冰模型配置的人员（例如Kiss 等人，2020 年）。Cookbook 在 GitHub 代码库中以开放方式开发，所有内容均通过使用 Sphinx 构建的文档网站公开（Komiya 等人，2025）；参见图 1。代码库中带标签的版本会上传至可引用的归档版本（Constantinou 等人，2026年）。

Cookbook 源自一个组织较为松散的 Python 脚本和 Jupyter 笔记本集合，其中许多已经无法正常运行。2020 年首次新冠肺炎封锁期间，社区举办了首次编程马拉松（Martínez-Moreno 和 Constantinou，2020 年）：一场为期一天的线上冲刺活动，旨在更新这些笔记本。此后，年度研讨会和编程马拉松逐步将代码库塑造成如今精心整理的样子。



图 1：COSIMA Cookbook 文档网站，展示了用于浏览教程和食谱的基于 Sphinx 的主页。网站地址为 https://cosima-recipes. readthedocs.io。

COSIMA Cookbook 刻意采用美食主题来组织内容，帮助用户快速了解并浏览网站。“烹饪教程”指教授处理海洋—海冰模型输出的通用可迁移技术的教程。这些教程为用户理解示例中的分析奠定基础。接下来是任何 Cookbook 的核心部分——食谱。这里的“食谱”指独立的笔记本，展示如何对海洋与海冰数据集进行具体的诊断分析和分析。首先是“简易

食谱”，为用户在学习教程后提供良好的起点；接着是“进阶食谱”，内容更为详尽，并展示进阶分析技术。最后，“区域特色”包含针对区域模型配置的食谱（Barnes 等人，2024年）。这种美食主题的组织方式也见于其他计算工作流程合集，例如面向地球科学领域的 Project Pythia 食谱，以及用于处理生命科学数据的 FAIR 食谱 。

教学是本项目的核心。“烹饪教程”介绍一些通用技能，例如加载模型输出、处理带标签数组、绘图以及使用共享数据目录。这些教程进一步引出面向特定领域的食谱，为学习者提供循序渐进的学习路径，帮助他们从初步接触数据生态系统，逐步掌握如何调整完整工作流程，以解答自己的科学问题。这些食谱为更深入复杂的分析提供了起点。“简单”和“进阶”食谱的分类则以简洁的方式提示其预期的复杂程度和范围。

在这一架构中，教程借助并展示了用于大规模数据科学分析的开源工具。其中包括通过基于 Intake 的目录加载模型输出的示例（ Intake 开发团队，2026 ），以及使用 Xarray （ Hoyer 和 Hamman，2017）和 Dask（Dask 开发团队，2016；Rocklin，2015）并行处理超大型多维数据集的示例。食谱集包含一个使用 Parcels（Delandmeter 和 van Sebille，2019）构建的拉格朗日粒子追踪工作流，也包含使用 xgcm（Abernathey 等，2022）处理定义在交错有限体积网格上的变量、使用 xESMF（Zhuang 等，2025）对模型输出进行网格重映射，以及使用 eofs（Dawson，2016）进行经验正交函数分析的内容。图表使用 matplotlib（Hunter，2007）生成，并结合 Cartopy（Elson 等，2024）进行地图可视化。

这本 Cookbook 源于 COSIMA 社区内部的实际需求（COSIMA 社区，2026年）：分析现代海洋模型输出的强大工具早已存在，但软件包级文档与可复现的端到端工作流程之间存在缺口，使新用户在很大程度上只能自己摸索。要有效使用这些工具，不仅需要熟练掌握 Python 和 Jupyter（Kluyver 等人，2016），还需要懂得如何处理高维且往往规模庞大的模型输出、如何使用共享的高性能计算环境，以及如何遵循特定领域的分析惯例。Cookbook 将可复用的工作流程整合到研究人员实际工作的同一平台中，从而弥合这一缺口。

2 需求说明

分析或配置气候模型的入门门槛很高。新用户必须学会如何查找数据集、高效加载数据、解读元数据、处理多维数组，以及生成具有科学意义的诊断分析。Xarray（Hoyer 和 Hamman，2017年）等通用库为这些任务提供了基础，但仅凭这些库，学习者仍无法了解如何针对特定模型和计算环境，将研究问题转化为稳健的分析工作流程。对于刚接触海洋与海冰科学的人来说，Cookbook 是一种学习工具，展示了如何探查模型输出并将常用诊断分析可视化；这些技能广泛适用于不同的模型和计算环境。

这一差距不仅关乎教学，也会直接拖慢科学研究。在开始任何分析之前，新手通常需要花费大量时间来弄清输出的结构、相关约定以及计算环境。许多常见的诊断分析也蕴含着大量科学劳动：有些计算需要经过大量开发、理解模型数值计算并完成验证，才能让人放心使用——这些就是“进阶食谱”。如果每位新用户都必须独立重建这些工作流程，那么大量精力就会花在重写分析代码上，而非开展科学研究。

COSIMA Cookbook 通过领域专属、开放维护的计算课程与示例集，促进知识共享并加速研究。它的贡献并非开发一个新的分析库，而是提供一种可重复使用的科学与教育资源，以许多研究人员实际用于探索数据和交流方法的同一种笔记本格式，展示完整且文档齐备的工作流程。项目选择使用笔记本，而不是仅将这些示例打包成软件库，

这样一来，学习者可以了解中间推理、方法选择和实际执行细节，同时更有经验的用户也能获得可直接用于研究并加以调整的分析模式。这与 JOSE 强调开放教育资源的理念十分契合：通过真实实践促进计算学习。

若干设计决策使 Cookbook 特别适合其他团队采用。首先，每本笔记本都旨在做到独立完整且文档齐备，让学习者无需从零散笔记中拼凑缺失的背景信息，就能阅读、运行并修改完整的工作流程。其次，代码库将教程与食谱区分开来。教程讲授可迁移的技能，而食谱则展示如何将这些技能结合起来，用于真实的分析任务。第三，项目提供了贡献者指南和笔记本审查规范，鼓励大家以可复用且易于教学的方式编写新的分析内容。需要注意的是，大多数食谱使用高分辨率模型输出，所用数据集的规模通常达到太字节量级，因此底层数据并不总是适合完整分发或镜像。不过，除了通常在每份食谱开头进行的数据加载步骤之外，这些分析工作流通常适用于大多数可通过 Xarray 加载的业务化海洋—海冰模型输出（Hoyer 和 Hamman，2017 年）。

因此，这些设计使 Cookbook 的价值超越了 COSIMA 本身的范畴。许多科研社区都在共享基础设施上维护模型输出，并面临同样的挑战：如何将专家的隐性知识转化为新手可以借鉴的示例。Cookbook 展示了一种模式，说明科研协作如何将这些知识记录在版本控制笔记本中，将其作为持续更新的开放资源在线发布，并通过社区贡献不断完善。

人工智能辅助编程工具的兴起，为 Cookbook 的价值增添了一个维度——可以说，这让它变得更加重要，而非不那么重要。这类工具能够快速生成分析代码；相比从零开始，如果将结构完善、领域专用的配方作为上下文提供给它们，生成的结果会好得多。但更深层的要点在于：随着人工智能生成的代码日益成为常态，关键技能正从编写代码转向理解和评估代码。研读过 Cookbook 的研究人员知道，工作流程为何要采用这样的结构——哪些参数很重要，科学假设会在哪些环节引入，以及合理的结果应该是什么样子。正是这种理解，使人工智能生成的代码能够得到信任、修正，并针对具体的科学问题进行调整。因此，随着人工智能的普及，传授分析背后的推理过程、而不只是操作方法的资源，其价值也会日益提升。

3 教学设计与使用体验

如上文所述，学习体验是循序渐进地组织起来的。这些教程和配方展示了基于 Jupyter 的分析，可直接在澳大利亚国家计算基础设施上运行，并提供了通过 JupyterLab 会话运行笔记本以及访问共享数据资源的指导。这使 Cookbook 不再只是一个静态的示例集：它还是面向真实分析环境的操作文档。同时，这些笔记本也展示了处理带标签的地球物理数据时具有广泛可迁移性的工作流程，因此，个别课程可以重复使用，也可以针对具体研究问题或其他计算平台进行调整。

该项目也支持社区协作创作。我们鼓励贡献者通过拉取请求提交新的笔记本，清晰记录所用方法，并将工作流程加以概括，使其对他人也有用。这一点尤其重要，因为科学分析所依据的计算机代码相对少有机会像论文稿件本身那样接受严格的同行评审。COSIMA 社区提供了多个案例，展示了配方生态系统如何拓展为完整的研究项目，并基于 GitHub 上同行协作编写的软件仓库开展研究；在这些项目中，分析工作流、方法和实现选择都公开接受社区审查，并可供他人复用。从这个意义上说，Cookbook 既是面向终端用户的学习资源，也通过笔记本设计和同行评审，构建了一个传授可复现科学传播的框架。在定期举办的

编程马拉松期间，社区会审查、更新并扩展现有配方，确保 Cookbook 始终是一项与不断发展的工具和社区最佳实践同步更新的活态资源。COSIMA Cookbook 让数据和分析易于获取，推动 ACCESS 海洋和海冰模型配置在国际范围内迅速得到采用；其应用领域不仅包括物理海洋学，也涵盖海洋生态学等相关学科（Fierro-Arcos 等人，2023），这些领域的研究人员将其用作获取 ACCESS 模型输出的入口。这种易用性迄今已助力完成 100 多篇同行评审论文，以及 20 多个博士项目。

作者排序

前两位作者贡献相同；从第三位起，作者按姓氏字母顺序排列。

致谢

本研究使用的计算资源由澳大利亚政府通过国家计算基础设施（NCI）提供，资助渠道为国家计算优异成果分配计划和澳大利亚国立大学分配计划。我们感谢澳大利亚海洋与海冰建模联盟（COSIMA；cosima.org.au）和澳大利亚气候模拟器（ACCESS-NRI；access-nri.org.au）活跃的社区通过 NCI 提供 ACCESS-OM2 和 ACCESS-OM3 的输出数据及分析工具。我们还感谢澳大利亚研究委员会通过气候极端卓越中心项目 CE170100023 和 21世纪天气项目 CE230100012、联结基础设施、设备与设施计划 LP160100073 和 LP200100406，以及探索项目 DP240101274 提供资助。

参考文献

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 和 Jon Thielen。xgcm，2022年11月。网址：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 和 Luwei Yang。regional-mom6：一个用于自动生成模块化海洋模型 6 区域配置的 Python 软件包。《开放源代码软件期刊》，9(100):6857，2024年。doi：10.21105/joss.06857。网址：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 和 Aditya Narayanan。COSIMA Cookbook：用于分析海洋和海冰模型输出的食谱，2026年。网址 https://doi.org/10.5281/zenodo.14353852。

COSIMA 社区。COSIMA：电影。https://youtu.be/bmUYRoPKtIA，2026年。YouTube 视频。

Dask 开发团队。*Dask*：动态任务调度库，2016年。网址 http://dask. pydata.org。

Andrew Dawson。eofs：用于气象、海洋学和气候数据 EOF 分析的库。《开放研究软件期刊》，4(1):e14，2016年。doi: 10.5334/jors.122。

Philippe Delandmeter 和 Eric Vanderbilt。Pacific-Arctic 2 0 拉格朗日框架和新的场插值方案。《地球科学模型开发》，12(8):3571–3584，2019年。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 和 Daniel Kirkham。SciTools/cartopy：v0.24.1，2024年10月。网址 https://doi.org/10.5281/zenodo . 13905945.

Denisse Fierro-Arcos、Stuart Corney、Amelie Meyer、Hakase Hayashida、Andrew E. Kiss 和 Petra Heil。利用高分辨率模型分析南大洋中与生态相关的海冰和海洋变量，以为生态系统研究提供依据。《海洋学进展》，215:103049，2023年。doi: 10.1016/j.pocean.2023.103049。

Stephan Hoyer 和 Joseph J. Hamman。xarray：Python 中的 N 维带标签数组和数据集。《开放研究软件期刊》，5(1):10，2017年。doi: 10.5334/jors.148。网址 https: //doi.org/10.5334/jors.148。

John D. Hunter。Matplotlib：二维图形环境。《计算机科学与工程》，9 (3):90–95，2007年。doi: 10.1109/MCSE.2007.55。

Intake 开发者。*Intake*：用于查找、调查、加载和分发数据的轻量级软件包，2026 年。网址：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 和 Xihan Zhang。ACCESS-OM2 v1.0：一个具有三种分辨率的全球海洋-海冰模型。《地球科学模型开发》，13(2):401–442，2020 年。doi：10.5194/gmd-13-401-2020。网址：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 和 Carol Willing。《Jupyter 笔记本——一种用于可复现计算工作流的出版格式》。收录于 Fernando Loizides 和 Birgit Schmidt 编，《学术出版中的定位与权力：参与者、代理人与议程》，第 87–90 页。IOS Press，阿姆斯特丹，2016 年。doi：10.3233/978-1-61499-649-1-87。网址：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 Ma- son、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 和 François Freitag。《sphinx-doc/sphinx：Sphinx 9.1.0》，2025 年。网址：https://doi.org/10.5281/zenodo.18109073。

Josué Martínez-Moreno 和 Navid C. Constantinou。COSIMA 黑客松宣传视频，2020年4月。https://youtu.be/JO-_jaPW8S8，2020 年。YouTube 视频。

Matthew Rocklin。Dask：采用分块算法和任务调度的并行计算。载于 Kathryn Huff 和 James Bergstra 编，第14届 Python 科学会议论文集，第 130–136 页，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 和 Xianxiang Li。pangeo-data/xESMF：v0.9.2，2025 年。网址 https://doi.org/10.5281/zenodo.4294774。