

1 **The PACE Data Hackweek launches a new wave of open science: how we did it (twice), and**
2 **why**

3

4 Anna Windle^{1,2*}, Ian Carroll^{1,3}, Carina Poulin^{1,2}, Skye Caplan^{1,2}, P. Jeremy Werdell¹

5 ¹ Ocean Ecology Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA

6 ² Science Systems and Applications, Inc., Greenbelt, MD, USA

7 ³ GESTAR II, University of Maryland Baltimore County, Baltimore, MD, USA

8 *Corresponding author: anna.windledipaola@nasa.gov

9

10 **Abstract**

11 PACE Data Hackweeks bring together student, early-career, and senior scientists to work
12 hands-on with data from NASA’s Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission.
13 During two past events, participants collaborated in small teams to explore scientific questions,
14 develop reproducible workflows, and gain experience working with large, cloud-hosted datasets.
15 Leveraging community supported cloud infrastructure, the hackweeks lowered technical barriers
16 and emphasized modern tools, open-source practices, and FAIR data principles. The events
17 fostered interdisciplinary collaboration and led to tangible outcomes, including new analyses,
18 shared code, and ongoing collaborations. Overall, the PACE Data Hackweeks demonstrate that
19 mission-focused hackweeks are an effective way to build community capacity and support
20 adaptation to evolving data and computing environments.

21

22 **NASA Does Hackweeks**

23 The NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) Data Hackweek builds
24 on a growing ecosystem of geoscience hackweeks that leverage NASA Earth Observing System
25 datasets, including [Geohackweek](#), [OceanHackWeek](#), and mission-specific events like [ICESat-2](#)
26 and [SnowEx hackweeks](#) (Huppenkothen et al., 2018; Mitchell et al., 2026). Unlike traditional
27 workshops or competitive hackathons, hackweeks emphasize participant-driven learning through
28 a mix of coding tutorials and group projects in a collaborative setting. Demand for these events
29 has grown alongside major shifts in Earth science data access, including the migration of
30 petabyte-scale data to NASA’s Earthdata Cloud hosted on Amazon Web Services (AWS) and the

31 launch of new satellite missions that introduce novel, complex data products. Launched in
32 February 2024, the PACE mission carries a three-instrument payload that collects synergistic
33 oceanic, atmospheric, and terrestrial observations, extending Earth system data records for the
34 next generation of scientists. Details on PACE data file formats, types, and volumes can be found
35 at the Ocean Biology Distributed Active Archive System ([OB.DAAC](#)) and [PACE websites](#). The
36 OB.DAAC provides processed data along with publicly available processing and visualization
37 tools, enabling researchers to quickly engage in exploring applications and interpretations of
38 PACE data products. However, the user community still faces the dual challenge of becoming
39 familiar with new data products while adapting workflows to cloud-based computing
40 environments, including the use of community supported software and cloud-infrastructure.

41 Hackweeks provide an effective mechanism for helping researchers adapt to these
42 evolving data environments by promoting open-source tools, FAIR (Findable, Accessible,
43 Interoperable, Reusable) data practices, reproducible workflows, and cloud-based collaboration
44 in support of NASA's open science goals (Scheick et al., 2026). Here, we describe the goals and
45 outcomes of the PACE Data Hackweeks, highlight participant experiences, and identify
46 remaining challenges. We conclude that sustained investment in hackweeks and community-
47 maintained cloud-infrastructure is essential for maximizing scientific and applied value of
48 NASA's evolving Earth science systems.

49 **Goals of a PACE Data Hackweek**

50 From organizing and leading two successful in-person hackweeks, we have gained
51 valuable insights and refined a suite of goals for teams interested in hosting similar events.
52 During early planning stages, our team consulted with the ICESat-2 Hackweek organizers,
53 whose experience and resources proved instrumental in shaping our approach. Members of our
54 team have also participated in and organized a multitude of short-format trainings, some within
55 the [Carpentries](#) or [eScience Institute](#) hackweek frameworks. These experiences allowed us to
56 formulate the goals below as a high-level strategy for achieving a research-focused event that
57 generates tangible scientific outputs.

58 *Goal 1: Kickstart Novel Research and New Collaborations*

59 PACE Data Hackweek participants arrive with diverse research ideas and are expected to
60 depart with well-defined projects and teams prepared for sustained remote collaboration.
61 Developing effective group projects is one of the primary organizational challenges, requiring
62 early alignment of research goals and active guidance from both scientific and technical mentors.
63 Daily science lectures and tutorials, modeled after the ICESat-2 Hackweek, helped shape
64 projects while providing exposure to current PACE research across oceanic, atmospheric, and
65 terrestrial domains, and a break from intensive coding.

66 To promote continuity beyond the hackweek, teams were encouraged to develop projects
67 into conference presentations, manuscripts, or proposal concepts, with continued support through
68 cloud computing resources, Slack communication, technical guidance, and funding and
69 publication advice. Structured social activities were also integral to fostering collaboration,
70 helping participants build networks, exchange feedback across teams, and sustain a balanced,
71 collaborative environment.

72 *Goal 2: Demonstrate the Value and Limits of Cloud Computing*

73 The transition of NASA’s Earthdata archive to the commercial cloud encourages
74 researchers to adopt cloud-based data analysis (Gentemann et al., 2021). Earthdata Cloud
75 provides free access to datasets stored in AWS and discoverable through NASA’s Common
76 Metadata Repository, but it does not currently offer compute resources; however, efforts are
77 underway to expand access to integrated cloud-based computing environments allowing users to
78 analyze data without needing to manage cloud infrastructure. Cloud-based analysis requires
79 separate compute infrastructure, most commonly AWS Elastic Compute Cloud (EC2). This
80 architecture supports high-throughput, data-intensive workflows by enabling direct access to
81 large datasets within the same AWS region, reducing data movement and speeding up
82 processing.

83 An emerging model for accessing EC2 instances is through community-maintained
84 cloud-based [JupyterHubs](#). Co-locating hackweek compute and data in the cloud lowers barriers
85 to working with large datasets without duplication, supports data streaming and reproducible
86 workflows, enables real-time collaboration, and allows users to scale seamlessly from simplified
87 tutorials to full-mission workflows (Box 1, Martin et al., 2025). For the PACE Data Hackweek,

88 we used CryoCloud, a JupyterHub maintained by a broad Earth science community, and
89 collected participation feedback on usability during the event (Box S2).

90

91 *Goal 3: Grow a Persistent Online Resource*

92 To extend the reach of the hackweeks, we created a public website using an open-source
93 [website template](#) and the GitHub Pages hosting service. The [PACE Data Hackweek website](#)
94 includes a single-page landing site for logistics and a detailed JupyterBook with recorded
95 lectures, coding tutorials, schedules, and preparation materials. Participants found it intuitive and
96 helpful, reducing pre-event questions.

97 Hackweek content is preserved as static archives and incorporated into NASA's Ocean
98 Ecology Lab [Help Hub](#), which provides reproducible Jupyter Notebook tutorials combining
99 code, text, and visualizations. Tutorials are collaboratively developed, peer-reviewed, and
100 regularly updated using hackweek feedback, ensuring relevance and usability. Hackweeks play a
101 critical role in motivating the continual updating of Help Hub tutorials, serving as development
102 sandboxes where project activities and participant feedback feeds directly into new or improved
103 notebooks. Since its launch, the Help Hub has grown from 48 to over 6,200 weekly page views,
104 with usage spikes following trainings and events (Figure S1).

105 *Goal 4: Leverage Prior Expertise*

106 Hackweeks, workshops, and short courses are common in academic and open-source
107 communities, often implicitly adopting the [Openscapes](#) program's flywheel approach: leveraging
108 momentum toward greater outcomes (Robinson & Lowndes, 2022). Organizing a hackweek
109 raises many questions about costs and execution, and planners are encouraged to consult multiple
110 past event managers for guidance, as it is unlikely that any one approach can be exactly
111 replicated. The PACE Data Hackweeks draw on two key sources of momentum: the CryoCloud
112 community and the Ocean Carbon and Biogeochemistry (OCB) Project Office.

113 Participants receive access to [CryoCloud](#) (Snow et al., 2022), a cloud-based JupyterHub
114 managed by the [International Interactive Computing Collaboration](#) (2i2c), originally developed
115 for NASA cryosphere research but now serving a broader Earth science community. To reflect

116 this expansion, the platform will soon be rebranded (T. Snow, personal communication). Access
117 during and after the hackweek is supported by NASA funding and AWS cloud credits.

118 Professional logistics support from the OCB Project Office assisted with funding,
119 contracts, and travel, allowing the core organizing team to focus on participant recruitment,
120 instructional design, and tutorial development. Other hackweeks have successfully partnered
121 with the University of Washington eScience Institute for similar logistical coordination.

122 We also built on prior expertise by empowering past participants to take on leadership
123 roles in subsequent events. These mentors brought valuable participant perspectives and offered
124 practical feedback that helped improve the overall participant experience.

125 **Experience at a PACE Data Hackweek**

126 The PACE Data Hackweeks, held in August 2024 and 2025, aimed to broaden access to
127 PACE data, enhance data-access tools, promote open-source workflows, and support integration
128 with other Earth observation datasets. Each five-day in-person workshop brought together 40-50
129 participants, with all instructional materials released and new code and analyses shared across
130 the community. Project teams were organized around participants' applications, which
131 organizers grouped into broad themes. Mentors then helped refine project ideas and ensure
132 feasibility within the week-long program. Through these collaborative efforts, participants
133 explored novel applications of PACE data, producing preliminary analyses that seeded new
134 collaborations, research directions, and early-stage proposal development.

135 Of the 88 total participants across both years, 36 completed a post-event feedback survey
136 and a separate demographic survey (41% response rate). Highlights from the surveys are
137 included in the following summary of the hackweek experience and in supplementary material.
138 We also show data on cloud computing resource usage and costs that were derived from EC2
139 activity logs during the 2025 event.

140 *Logistics*

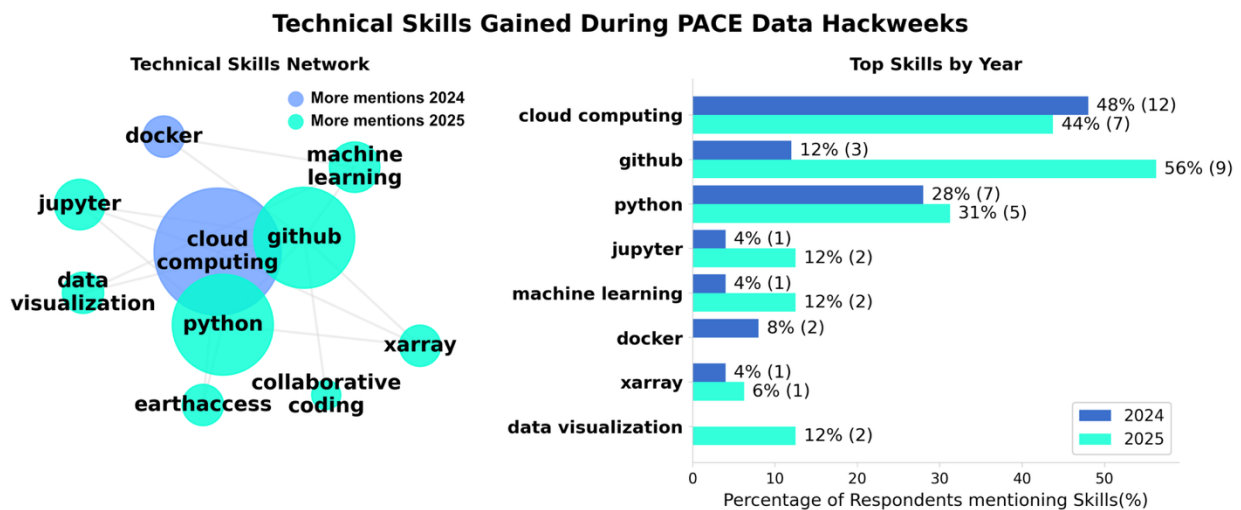
141 Each five-day, in-person event was held at the University of Maryland Baltimore County
142 (UMBC) and funded through OCB Training Activity grants. For U.S.-based participants, OCB
143 covered travel, lodging, and meals, while international participants received a partial travel
144 stipend, with remaining expenses supported by the participant or their home institution.

145 Accommodations were provided in UMBC dormitories, and meals were served at the university
 146 cafeteria. To support participants with family or dependent care responsibilities, additional
 147 funding was available to offset caregiver costs or to allow for customized travel arrangements.

148 *Schedule*

149 Each day combined scientific lectures with hands-on coding tutorials and demonstrations,
 150 building skills in data access, Python programming, and cloud computing (Figure 1). Both
 151 science lectures and coding tutorials were highly rated (based on 1-5 ranking scores collected in
 152 feedback surveys) and 100% of survey respondents reported increased confidence in working
 153 with PACE data (Figure S2). Daily check-ins supported collaboration and progress, and on the
 154 final day teams presented their projects and shared code via public GitHub repositories, with
 155 most planning continued collaboration (Box S1). Feedback was overwhelmingly positive, with
 156 participants describing the hackweek as fun, engaging, and collaborative (Figure S3). Social
 157 events, including a Bob Ross paint night, fireside chat with NASA Program Managers, local crab
 158 feast, PACE Nerd Night, and themed trivia, were frequently cited as essential for community
 159 building and sustained engagement:

160 *“Best part of the day! Every night was so great and the energy was great.”*
 161 *“I loved the intentional inclusion of social activities. It helped people meet and have*
 162 *fun while maintaining work-life balance and camaraderie.”*



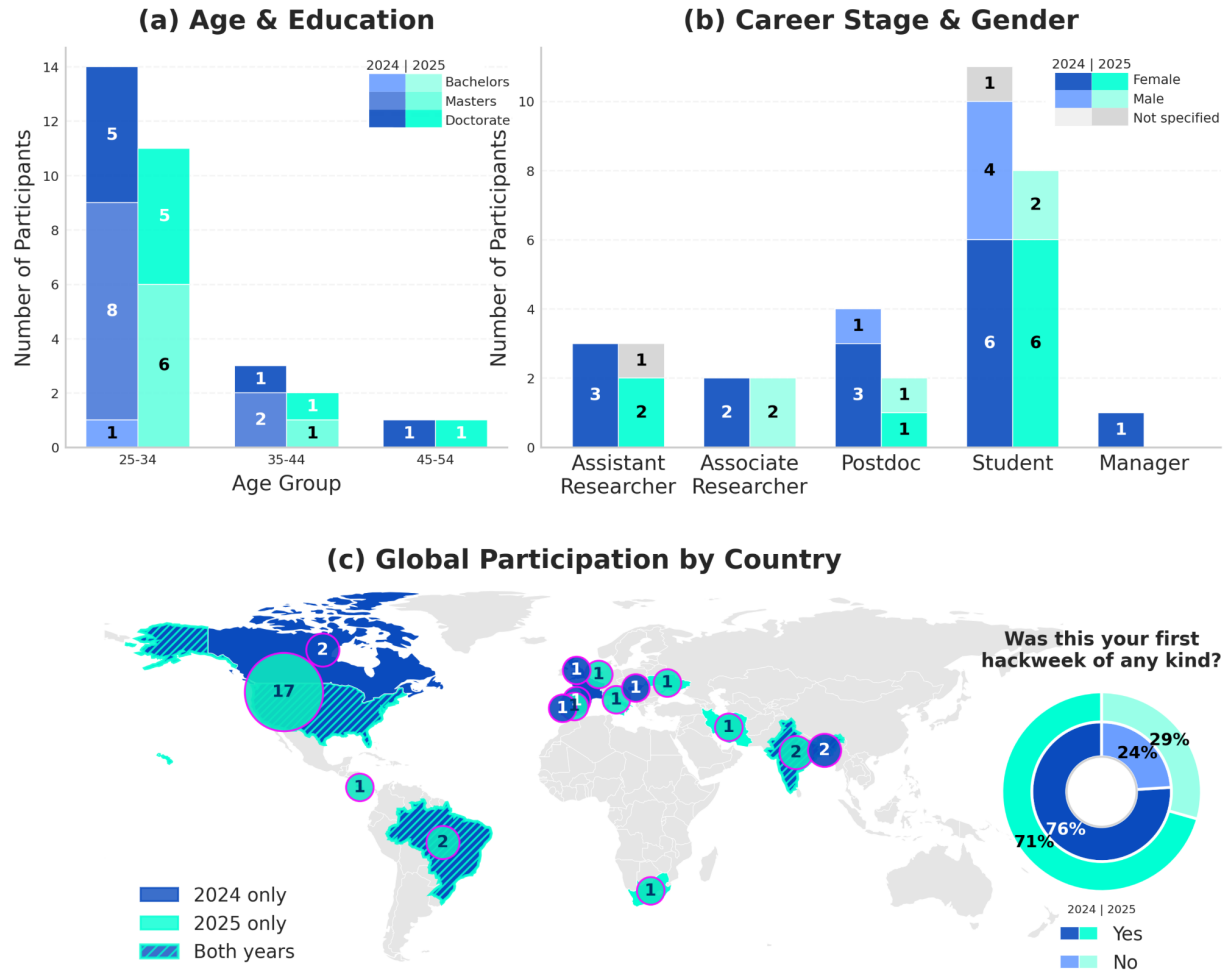
163
 164 Figure 1. Most common technical skills, and associations among technical skills, mentioned in

165 responses to the feedback survey question, “*What technical skills did you gain this week?*” Node
166 size in the technical skills network represents the number of times a skill was mentioned, while
167 proximity indicates how frequently skills were co-mentioned by the same participant. Nodes are
168 colored based on the year in which the skill was mentioned most often.

169 *Demographics*

170 The respondents represented a diverse international community spanning fifteen
171 countries (Figure 2). Most were graduate students or early career researchers, with the majority
172 holding master’s or doctoral degrees (Figure 2). Notably, 74% reported that the PACE Data
173 Hackweek was their first hackweek experience.

174



Total participants: 2024 (n=22), 2025 (n=15)

175
 176 Figure 2. Responses to the demographic survey showing (a) age and education level, (b) career
 177 stage and gender, and (c) country of residence and first-time hackweek participation. Providing
 178 gender was optional and open-ended, with a non-response tagged as “Not specified”.

179 *Coding environment*

180 PACE Data Hackweek participants onboarded to CryoCloud benefited from a fully
 181 configured, cloud-based software environment with direct access to all PACE datasets and no
 182 local setup required. The hub launched Docker containers, running on adjustable hardware
 183 configurations (2 - 29 GB RAM and 0.2 - 4 CPUs), that included required Python packages and
 184 the OBPg’s Ocean Color Science SoftWare (OCSSW); this container is available on CryoCloud

185 as the “Ocean Biology Python Image” (Figure S4). Participants were able to begin running
 186 tutorials and regenerating PACE data products in the commercial cloud almost immediately, all
 187 through a web browser.

188 Using the community-maintained JupyterHub, participants could execute tutorial and
 189 group-project notebooks, while experiencing both the benefits and limitations of cloud-based
 190 workflows (Box 1). Overall, CryoCloud played a central role in reducing technical barriers,
 191 fostering collaboration, and supporting open, reproducible workflows (Box S2). Feedback
 192 indicated that the majority of participants intend to continue using cloud computing in their
 193 research; although some concerns remain (Box S3, Figure S5).

194

Box 1. Summary of the value and limitations of using a cloud-based community-maintained JupyterHub for the PACE Data Hackweek	
+ Value	- Limits
Managed Compute Environment	
<ul style="list-style-type: none"> • Eliminates the need for individual software configuration, allowing participants to begin productive work immediately • Can be reused on other local or remote platforms beyond the hackweek • Reduces technical barriers and inequalities among participants • Fosters a community of support and collaboration 	<ul style="list-style-type: none"> • Restricted control over resources and system configuration • Learning to manage limited memory (Box S3) • Unfamiliar development environment and operating system (i.e., Jupyter Lab, Linux, S3 bucket storage and access, commonly owned "shared" filesystem)
Direct Data Access	
<ul style="list-style-type: none"> • Storage-adjacent compute resources reduce latency • High-bandwidth connections and unlimited concurrency support collaborative workflows 	<ul style="list-style-type: none"> • Lack of cloud-optimization for data products stored as NetCDF4/HDF5 reduces performance • Data products from different agencies (e.g., NASA, USGS, NOAA, ESA) are not all co-located in the commercial cloud

<ul style="list-style-type: none"> ● Removes dependencies on local storage or network limitations 	
Facilitated Collaboration	
<ul style="list-style-type: none"> ● Encourages reproducible research by standardizing the computing environment and access to datasets ● Supports generation of preliminary analyses and hypothesis testing that can be built into future research proposals ● Promotes sharing of code, workflows, and results with the broader community, accelerating scientific discovery 	<ul style="list-style-type: none"> ● Reproducing the compute environment outside a managed JupyterHub requires additional training ● Limits and “best effort” restrictions on per-user and shared file storage (Figure S5)
Commercial Cloud Scalability	
<ul style="list-style-type: none"> ● Supports both simple tutorials and computationally intensive projects ● Enables seamless scaling for larger datasets or complex analyses 	<ul style="list-style-type: none"> ● Reliance on S3 over local or network filesystems for persistent storage is unfamiliar and less widely supported

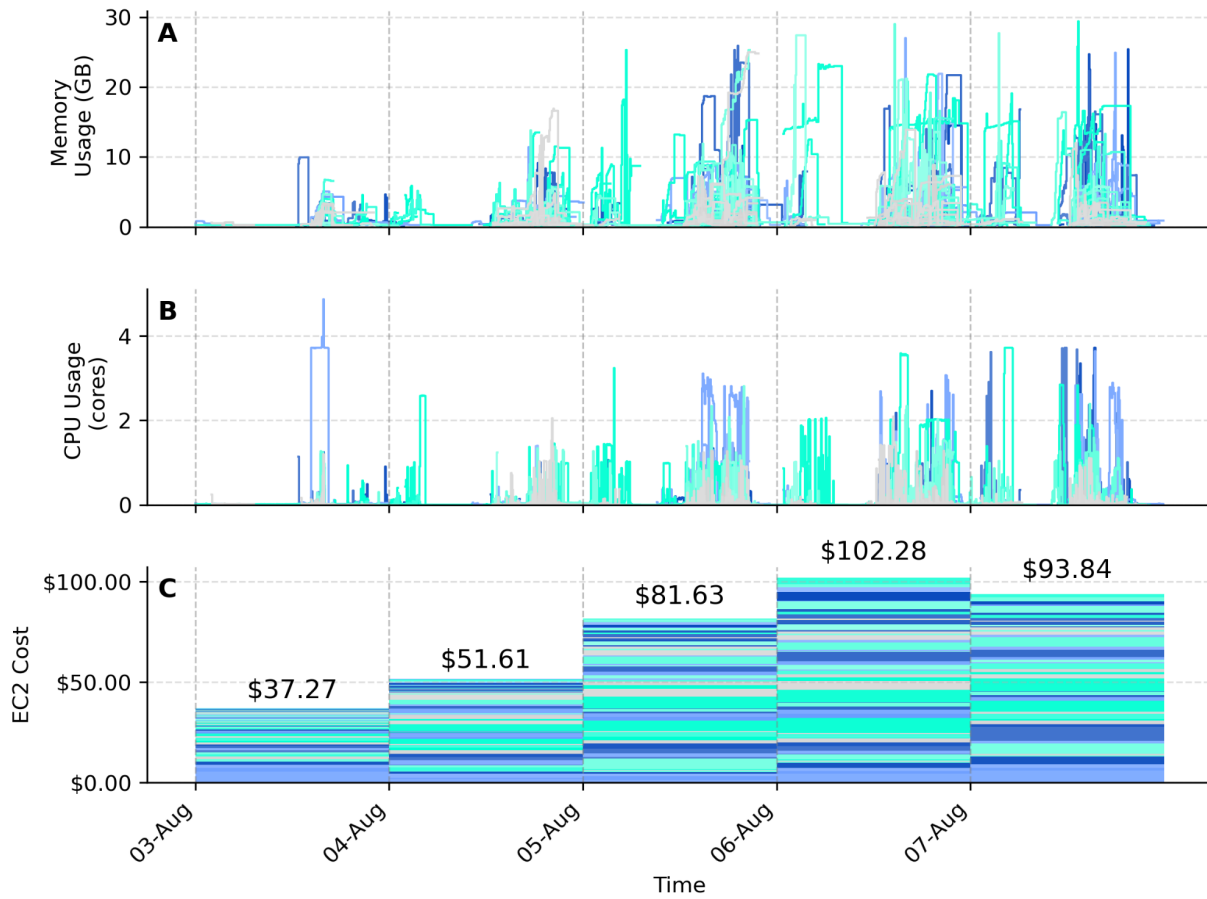
195

196 *Commercial cloud usage and costs*

197 Using Grafana dashboards developed by 2i2c, we charted EC2 usage and costs for all
198 participants during the 2025 PACE Data Hackweek (Figure 3). Daily memory usage per user
199 ranged from 155 KB to 29.4 GB, with an average of 1.79 GB, while CPU usage peaked at ~3.7
200 CPUs per user and averaged 0.6 CPUs per day (Figure 3). Peak memory and CPU demand
201 occurred late in the week during intensive group project development.

202 CryoCloud costs include user-dependent expenses (e.g., compute and personal storage)
203 and user-independent infrastructure costs that run continuously. Compared to the previous model
204 of launching a dedicated hackweek JupyterHub, incurring ~\$2,250 in startup and ~\$2,250 in
205 monthly infrastructure costs (T. Snow, personal communication), maintaining a shared hub is

206 substantially more cost-effective. For the full week, total EC2 compute costs were \$367.99, with
 207 \$36.36 in individual storage and \$19.49 in shared storage, for a combined total of \$423.84.
 208 Individual daily costs peaked on Day 4, with the highest single-user daily cost at \$7.31 and a
 209 maximum single-user weekly cost of \$22.61 (Figure 3). Uncertainty around these costs was a
 210 commonly raised concern among participants (Box S3), who did not have real-time access to the
 211 Grafana dashboards.



212
 213 Figure 3. A) Memory usage (GB) in CryoCloud for each PACE Data Hackweek participant over
 214 the five-day 2025 event, B) central processing unit (CPU) usage (cores) in CryoCloud for each
 215 participant, and C) Elastic Compute Cloud (EC2) cost incurred by each participant over the five-
 216 day event. Colors denote individual participants.

217 Opportunities for Improvement

218 While cloud computing is essential for processing large PACE datasets and enabling
 219 hackweeks several challenges remain. (1) CryoCloud provides user home directories backed by

220 AWS Elastic Block Storage (EBS) but limits storage to 50 GB due to per-user costs, creating
221 constraints for users and the community. (2) Configuring and maintaining compute environments
222 remains a technical challenge. Guidance on defining and running reproducible workflows, e.g.
223 [binder-ready repositories](#), is often perceived as technically advanced by hackweek participants.
224 (3) Data-adjacent computing requires co-located datasets, which is often not the case across
225 agencies, and user-contributed data archives, making temporary data transfer, storage and
226 remote execution necessary even when working in a commercial cloud. (4) Performance
227 limitations occur when accessing data not optimized for S3 (McNally et al. 2026). (5) Growing
228 the CryoCloud community requires additional funding, but clear mechanisms for contributions
229 from grants, institutions, or agencies are not yet established. Future collaboration between
230 hackweek organizers, data engineers, and cloud system administrators would help prioritize and
231 address these challenges, with benefits extending beyond future hackweeks.

232 While PACE Data Hackweeks have been highly effective in building skills and
233 community, several opportunities exist to improve their reach, structure, and long-term impact.
234 (1) PACE Data Hackweeks have primarily attracted early-career scientists, but they also offer
235 valuable opportunities for late-career researchers to update their skills in modern, cloud-based
236 workflows. Future hackweeks will aim to increase participants from this group to broaden impact
237 across the community. (2) Participant feedback also points to challenges with project
238 development and completion. To address this, future PACE Data Hackweeks may adopt a more
239 structured approach, with pre-defined project topics, participant matching, and dedicated mentors
240 to support progress during and after the event.

241 **Concluding Remarks**

242 Hackweeks have become an effective mechanism for advancing scientific work, in large
243 part because they bring together researchers who might not otherwise collaborate. Interactions
244 often continue after the event, leading to new ideas and follow-on projects. By combining
245 different areas of expertise over an intense but short research sprint, hackweeks provide space to
246 experiment with new approaches and to take risks on novel or uncertain research directions.

247 Participants of the PACE Data Hackweeks not only explored new data products but also
248 developed familiarity with cloud-based workflows and joined a larger community of Earth
249 science researchers working to maximize the benefits of NASA's Earthdata Cloud. Leveraging

250 an established, cloud-based JupyterHub for hackweeks is practical, cost-effective, and provides a
251 nexus for continued interactions. Continued investment in hackweeks and shared cloud resources
252 will be critical for sustaining these benefits.

253 As NASA's Earth Observing System evolves, hackweeks are well positioned to serve as
254 catalysts for skill-building and community formation The oceanography community faces a mix
255 of opportunities and challenges as it advances open science. Participants in the PACE Data
256 Hackweeks are empowered to translate broad shifts in the data and computational landscape into
257 meaningful, lasting impacts for the community.

258 **References**

- 259 Gentemann, C. L., Holdgraf, C., Abernathey, R., Crichton, D., Colliander, J., Kearns, E. J.,
260 Panda, Y. & Signell, R. P. (2021). Science storms the cloud. *AGU Advances*, 2(2),
261 e2020AV000354.
262
- 263 Huppenkothen, D., Arendt, A., Hogg, D. W., Ram, K., VanderPlas, J. T., & Rokem, A. (2018).
264 Hack weeks as a model for data science education and collaboration. *Proceedings of the*
265 *National Academy of Sciences*, 115(36), 8872-8877.
266 <https://doi.org/10.1073/pnas.1717196115>
267
- 268 Martin, P.E., E. Eli Holmes, E. Mayorga, J.K. Ansong, U. Bhaskar, J. Cornejo-Donoso, D.
269 Correa-Chilón, R. Damoah, D. Fierro-Arcos, L. Gómez-Navarro, N. Kumar, A. Lawal-
270 Are, S. Maity, S. Majumder, D. Menemenlis, A. Modi, E. Nyadjro, O. Oghenechovwen,
271 A. Oikonomou, M. Oladipo, M. Peña, D.T. Quaye, Y. Santana-Falcón, BR Smitha, C.
272 Troupin, G. Vagenas, H. Villalobos, and G.L. Wagner. (2025). Harnessing marine open
273 data science for ocean sustainability in Africa, South Asia, and Latin America.
274 *Oceanography* 38(1):59–66, <https://doi.org/10.5670/oceanog.2025.121>.
275
- 276 McNally, A., Sterzinger, L., & Carroll, I. (2026). *Producing Earth science data for impact:*
277 *Improved commercial cloud usability of archive model data*. Environmental Modelling &
278 Software: 197 (106846). <https://doi.org/10.1016/j.envsoft.2025.106846>
279
- 280 Mitchell, C., W.-J. Lee, F. Fernandes, J. Gum, A. Kerney, E. Mayorga, T. Moore, N. Mortimer,
281 N. Ribeiro, and V. Staneva. 2026. OceanHackWeek: An inclusive, collaborative
282 approach to developing oceanography data science skills. *Oceanography* 39(1),
283 <https://doi.org/10.5670/oceanog.2026.e104>.
284
- 285 Robinson, E., & Stewart Lowndes, J. (2022). The Openscapes Flywheel: A framework for

286 managers to facilitate and scale inclusive Open science practices. California Digital
287 Library (CDL). <https://doi.org/10.31223/x5cq02>

288
289 Scheick, J., Barrett, A., Fair, Z., Katz, Z., Lopez, L., Neeley, A., Roberts, C., Smith, B., Snow,
290 T., Wegener, R. and Arendt, A. (2026). Ten Simple Rules to Increase the Usability of
291 Earth Observation Mission Datasets. *ESS Open Archive*. January 20, 2026.
292 <https://doi.org/10.22541/essoar.176892205.56393151/v1>

293
294 Snow, T., Millstein, J., Scheick, J., Sauthoff, W., Leong, W. J., Colliander, J., Pérez, F.,
295 Munroe, J., Felikson, D., Sutterley, T., & Siegfried, M. (2023). *CryoCloud JupyterBook*.
296 Zenodo. [10.5281/ZENODO.7576601](https://doi.org/10.5281/ZENODO.7576601)

297
298 **Acknowledgments**

299 We gratefully acknowledge the PACE Data Hackweek mentors, tutorial contributors, and
300 science presenters as well as Tasha Snow for guidance on CryoCloud and Eli Holmes for advice
301 on hackweek organization. We also thank the ICESat-2 Hackweek team for early scoping
302 support, the OCB Project Office, NASA Ocean Biology Biogeochemistry Program Managers,
303 and the UMBC staff for logistical and A/V assistance. And lastly, we thank all hackweek
304 participants for their enthusiastic engagement.

305
306 **Funding Support**

307 The PACE Data Hackweek was established with the support of the Ocean Carbon &
308 Biogeochemistry (OCB, us-ocb.org) Project Office, which receives funding from the National
309 Aeronautics & Space Administration (NASA 80NSSC25K7397) and the National Science
310 Foundation (NSF OCE-2445578).