

HACKING LIMNOLOGY WORKSHOP AND DSOS22

Creating a Community of Practice for the Nexus of Data Science, Open Science, and the Aquatic Sciences

Michael F. Meyer , Carolina C. Barbosa ,
Robert Ladwig , Jorrit P. Mesman ,
Nahit Soner Börekçi , Kaelin Cawley ,
Johannes Feldbauer , Isabella A. Oleksy ,
Rachel M. Pilla , Patricia Q. Tran, Jacob A. Zwart ,
Nicolas Azaña Schnedler-Meyer , Tobias K. Andersen ,
Matthew R. Brousil , Kate C. Fickas ,
Alessandro Filazzola , Tyler V. King ,
Nuria Sánchez-López , Victoria Stengel , and
Dennis Trolle 

INTRODUCTION

The 2nd Aquatic Ecosystem Modeling-Junior (AEMON-J) Hacking Limnology Workshop and 3rd Virtual Summit: Incorporating Data Science and Open Science in the Aquatic Sciences (DSOS) took place on 25–29 July 2022. These virtual events were developed to bring together researchers from diverse backgrounds to share developments in data-intensive research in the aquatic sciences and train participants in cutting-edge data analysis methods related to remote sensing, data pipelines, and modeling of aquatic ecosystems.

Over 525 people from more than 50 countries and six continents registered for the week-long event. As in previous years, there were no registration costs for the workshop or summit. Countries with the highest number of registrants included the United States (49.2%), Canada (7.5%), Germany (7.1%), Brazil (4.7%), and Nigeria (3.0%). The majority of registrants (61.2%) identified themselves as being in early career positions (i.e., graduate students and post-doctoral researchers; Fig. 1A), and workshop registrants also broadly self-labeled themselves as having more intermediate experience with aquatic ecosystem modeling compared to previous years (Fig. 1B). Regardless of whether

registrants were able to attend synchronously during the time of the workshop and summit, all recordings and materials were made available on the combined Open Science Framework archive for AEMON-J and DSOS (Meyer et al. 2022) for asynchronous viewing and access.

The combined workshop and summit ran over the course of 1 week and expanded on previous years' successes, while also opening new initiatives. As with previous years, the workshop provided participants with diverse talks and opportunity to learn data-intensive skills

through the hands-on coding workshops. While the format, topics, and number of registrants were similar to previous years, AEMON-J and DSOS created a joint social media account (<https://twitter.com/HackingLimno>), which is designed to promote work, opportunities, and news that intersects with data science, open science, and the aquatic sciences. This social media account is intended to increase engagement with the Hacking Limnology Community year-round. Those interested in being involved at the leadership helm can contact the organizational team.

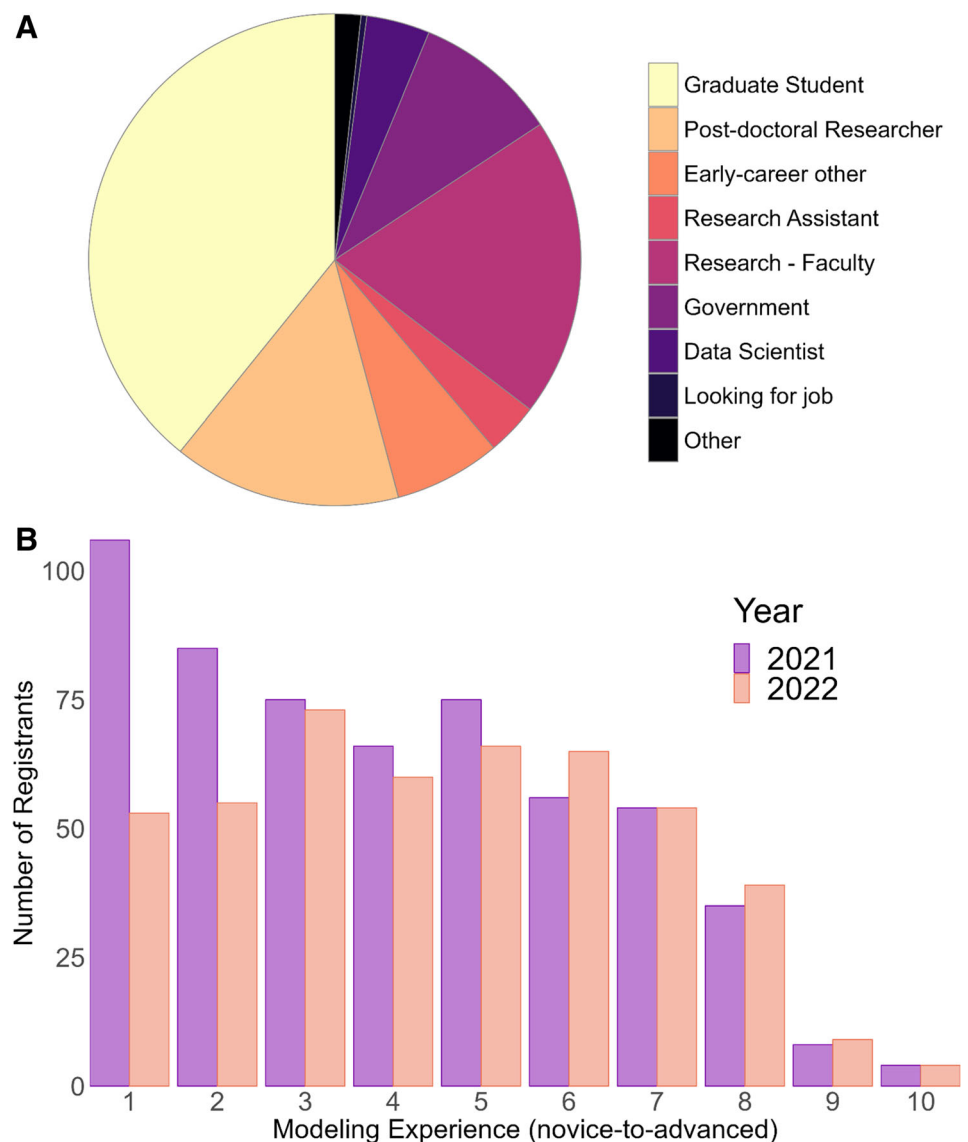


FIG. 1. Reported (A) position and (B) modeling experience for registrants across 2021 and 2022 workshop series. The majority of registrants (61.2%) self-reported as an early career researcher (i.e., graduate student, post-doctoral researcher, or “early career-other”). Faculty and government employees were the second (20%) and third (9.5%) self-reported positions, respectively. Self-reported, average modeling experience among registrants was 4.5 (SD = 2.3), which was slightly higher than self-reported modeling experience in 2021 (mean = 4.0; SD = 2.3). A score of 1 was defined as “Novice” and 10 was defined as “Advanced.”

Michael F. Meyer, Carolina C. Barbosa, Robert Ladwig, and Jorrit P. Mesman contributed equally to this study.

THE “HACKING LIMNOLOGY” WORKSHOP

The 2022 “Hacking Limnology” workshop series took place on 25–27 July 2022, covering topics of “Remote Sensing,” “Data Science,” and “Aquatic Ecosystem Modeling.” Workshops were structured similarly to the previous “Hacking Limnology” series (Meyer et al. 2021b); introductory videos were made available for viewing on the program’s website 7–10 days before the start of the event and were intended to be watched beforehand. At the beginning of each day, the organizational team gave a short introductory presentation, which included guidelines for asking questions, the code of conduct, and introductions to the workshop leaders. A pre-recorded, 15-min keynote presentation was live-streamed, followed by a live “Question and Answer” session, where attendees could submit questions to the introductory and keynote speakers through an online form. Afterward, we hosted a live, hands-on, 2-h workshop, involving coding tutorials and modeling. Each day ended with a “social hour” in which participants could interact with the workshop leaders and with each other to explore the workshop material further, discuss collaborative project ideas, and network with workshop leaders and participants. Approximately 90–100 people attended the workshop series each day, which were the highest numbers recorded for attendance of previous DSOS/AEMON-J workshops (2021) on data science (~50 attendees) and ecosystem modeling (~40 attendees) (Meyer et al. 2021a, b).

To establish a broad introductory toolkit for early career researchers interested in computational aquatic science, we addressed new topics in the 2022 event, such as building data pipelines (Meyer et al. 2021b). During the “Remote Sensing” day, participants were introduced to a suite of existing remote sensing products, including a tutorial on how to process raw data into more useful environmental data, and the importance of ground-truthing remote sensing data. In the live coding workshop, the participants implemented water quality spectral indices, and extracted a chlorophyll *a* time series for the surface waters of Lake Pontchartrain (Louisiana, USA), using Google Earth Engine (Gorelick et al. 2017). The “Data Science” day focused on workflow management software, including what tools are available, what the potential benefits are, and, during the live coding session, how to set up a reproducible data processing pipeline. The data used for the workshop were provided by the U.S. National

Ecological Observatory Network (NEON; <https://www.neonscience.org/>), and the live coding session introduced the participants to the *targets* R package (Landau 2021). In the final workshop on “Aquatic Ecosystem Modeling,” participants were introduced to the general principles, assumptions, and uses of aquatic ecosystem modeling. Furthermore, there was a live training workshop on the model suite of the General Ocean Turbulence Model and Water Ecosystem Tool (GOTM-WET) (Umlauf et al. 2005; Schnedler-Meyer et al. 2022), using the QGIS (<https://www.qgis.org/en/site/>) graphical user interface QWET (Nielsen et al. 2021) to simulate temperature and water quality dynamics in a Danish lake.

While the structure of the 2021 and 2022 workshop series was similar, there were some key, organizational differences. The 2022 workshop series lasted 3 days, compared to the previous year’s 4 days, which was partially done to strengthen the association between the workshop series and the DSOS summit, promoting it as a single, 1-week event with two parts, rather than separate activities. We note that 75% of participants registered for both events. Additionally, the duration of the keynote presentations was shortened from 25 to 15 min to lessen the fatigue of online presentations and to allow for a longer and more active Question and Answer session. The participants actively made use of the question forms, which allowed the organizers to follow up on questions after the workshop. During the social hour, participants frequently asked for elaboration and feedback related to the workshop content. Collaborative projects regarding the remote sensing applications, data pipelines, and modeling initiatives were lively discussed.

THE VIRTUAL SUMMIT #DSOS22

Following the “Hacking Limnology” workshops, the 3rd annual “Virtual Summit: Incorporating Data Science and Open Science in Aquatic Research” ran on 28–29 July 2022. The third edition of the summit was organized in the same structure as the previous years with pre-recorded talks and a live Question and Answer session with presenters (Meyer et al. 2020, 2021b). The summit consisted of four main sessions each of which hosted approximately five speakers and lasted 1.5 h, organized around the themes of “Big Data,” “Data Intensive Models,” “Tools and Software,” and “Applications of Open Science.” Presentations included a wide range of themes such as global and

regional databases, data-driven and process-based model applications, and open-source tools to help managers, educators, and researchers engage in ecoinformatics and remote sensing. In addition, two panels composed of a diverse team of researchers discussed “Operationalizing Data Interoperability” and “Careers in Data Science and Open Science.” Above all, the main goal of the summit was to bring together diverse early career researchers to share how they bring data science and open science into their research. Approximately 50–80 people attended each day of the summit, with approximately 80% of presenters and panelists at early stages in their respective careers.

Building on previous summits, the 2022 summit was committed to diverse representation of speakers and panelists. In particular, the 2022 summit increased the opportunities for accessible non-English participation. While the 2021 summit had one presentation in Portuguese, the 2022 summit had three non-English presentations in French, Portuguese, and Persian. Each talk was live-streamed with English subtitles, and slides were either completely in English or in a combination of English and the spoken language. The Question and Answer session, however, was all in English. Outside of presentations, participants could interact in 10 language channels on the DSOS virtual workspace which included Spanish, Portuguese, French, German, Russian, Polish, Serbian-Croatian-Bosnian-Montenegrin, Italian, Turkish, and Mandarin. Each channel was moderated by a volunteer, who was responsible for ensuring that each language channel remained in congruence with our community code of conduct.

Beyond virtual components, a new initiative for the 2022 summit was encouraging regional “Watch Parties,” where participants could physically meet to watch the live stream of the summit. To organize watch parties, we created a form, where respondents could mark themselves as interested in attending or organizing a local watch party. Once registered, participants would have access to the response document, where they could locate and contact individuals in their region. Ultimately, two “watch parties” at the University of Wyoming and the University of Wisconsin-Madison were organized and helped bring physical presence to an otherwise virtual event. Each watch party consisted of three to eight participants and was primarily organized by individual departments and lab

groups, who could watch the summit's presentations together in one space.

MOVING FORWARD

Much like previous virtual summits and workshops (Meyer and Zwart 2020; Meyer et al. 2021b), the 2022 Hacking Limnology Workshop and Summit received overwhelmingly positive feedback from presenters, panelists, and participants. Unlike previous years, the 2022 workshops and summits occurred at a time when in-person scientific conferences, field sampling campaigns, and lab work were returning. While virtual conferences are likely here to stay (Meyer et al. 2021a), there will be growing pains with learning how virtual conferences attract participants, presenters, and panelists, in times when in-person activities can compete with virtual events. While registration was comparable for the 2022 workshops and summit, synchronous engagement with the summit was lower than previous years, whereas synchronous workshop engagement was higher than the previous year. These differences could stem from any number of reasons, such as limited time to attend both the workshop and summit in the same week, dependent-care responsibilities, time zone differences, field or lab work obligations, prioritization of in-person conferences or training opportunities like the workshop days, or general fatigue of online meetings and listening to online presentations in daily life. Consequently, the Open Science Framework archive (Meyer et al. 2022) is a resource for previous years' recordings, workshop materials, and general information. By aggregating advanced data science and open science training as well as outreach materials in one place, the archive further serves as a resource for asynchronous, extended, and accessible engagement with the larger aquatic sciences community, thereby lowering barriers for inclusion and engagement.

Despite the rapidly evolving challenges of working and living more than 2 years into the COVID-19 pandemic, the success of the second hacking limnology workshop and third virtual summit suggests that the global aquatic science, data science, and open science community continue to engage at this nexus. However, to continue building on our current foundations, the organizational team is interested in creating opportunities for new leadership and trajectories for the larger "Hacking Limnology" community as well as in exploring

potential for external funding that can spur new and sustain current initiatives. While the exact structure is being decided, the organizational team is enthusiastic about the potential to bring in new ideas, backgrounds, disciplines, and skill sets. By expanding the leadership helm, we hope to help even more practitioners from all backgrounds engage in this exciting science.

AUTHOR CONTRIBUTIONS

Authors are listed in the following order: Core author group (MFM, CCB, RL, JPM), DSOS and AEMON-J (NSB, KMC, JF, IAO, RMP, PQT, JAZ) leads in alphabetical order by surname, workshop organizers in alphabetical order by surname (Remote Sensing: KCF, TVK, NSL, VS; Big Data: MRB, AF; and Numerical Modeling: NASM, TKA, DT). All authors contributed to writing and approved the final manuscript.

ACKNOWLEDGMENTS

Michael F. Meyer and Kate C. Fickas were funded by a Mendenhall Fellowship from the U.S. Geological Survey. Carolina C. Barbosa and Isabella A. Oleksy were supported through an NSF EPSCoR Track 2 RII grant (EPS-2019528). Robert Ladwig was supported through an NSF ABI development grant (#DBI 1759865), a UW-Madison Data Science Initiative grant, and an NSF HDR grant (#1934633). Jorrit P. Mesman was funded by the European Union's Horizon 2020 research and innovation program, grant agreement number 101017861 (SMARTLAGOON). Nahit Soner Börekçi was supported by YÖK 100/2000 PhD Scholarship from the Council of Higher Education of Turkey. Rachel M. Pilla was supported by the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, and Water Power Technologies Office. Tobias K. Andersen was supported by the Poul Due Jensen Foundation. Johannes Feldbauer received funding from the BMBF project 437FKZ 01LR 2005A—Fördermaßnahme "Regionale Informationen zum Klimahandeln" (RegIKlim). The National Ecological Observatory Network is a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle. This material is based in part upon work supported by the National Science Foundation through the NEON Program. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Gorelick, N., and others. 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**: 18–27. doi:10.1016/j.rse.2017.06.031.
- Landau, W. M. 2021. The targets R package: a dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *J. Open Source Softw.* **6**: 2959. doi:10.21105/joss.02959.
- Meyer, M. F., and J. A. Zwart. 2020. Virtual summit: incorporating data science and open science in aquatic research. *Limnol. Oceanogr. Bull.* **29**: 144–146. doi:10.1002/lob.10411.
- Meyer, M. F., and others. 2021a. Virtual growing pains: Initial lessons learned from organizing virtual workshops, summits, conferences, and networking events during a global pandemic. *Limnol. Oceanogr. Bull.* **30**: 1–11. doi:10.1002/lob.10431.
- Meyer, M. F., and others. 2021b. The AEMON-J "Hacking Limnology" Workshop Series & Virtual Summit: Incorporating data science and open science in aquatic research. *Limnol. Oceanogr. Bull.* **30**: 140–143. doi:10.1002/lob.10475.
- Meyer, M. F., and others. 2022. AEMON-J/DSOS archive: "Hacking Limnology" Workshop + Virtual Summit in Data Science & Open Science in Aquatic Research. doi:10.17605/OSF.IO/682V5
- Nielsen, A., and others. 2021. Introducing QWET – a QGIS-plugin for application, evaluation and experimentation with the WET model. *Environ. Model. Softw.* **135**: 104886. doi:10.1016/j.envsoft.2020.104886.
- Schnedler-Meyer, N. A., and others. 2022. Water ecosystems tool (WET) 1.0 – a new generation of flexible aquatic ecosystem model. *Geosci. Model Dev.* **15**: 3861–3878. doi:10.5194/gmd-15-3861-2022.
- Umlauf, L., H. Burchard, and K. Bolding 2005. GOTM: Sourcecode and test case documentation. Version 4.0. Available from <https://gotm.net/manual/stable/pdf/a4.pdf>.

Michael F. Meyer, U.S. Geological Survey, Madison, WI, USA; mfmeyer@usgs.gov

Carolina C. Barbosa, University of Wyoming, Laramie, WY, USA

Robert Ladwig, University of Wisconsin, Madison, WI, USA

Jorrit P. Mesman, Uppsala University, Uppsala, Sweden

Nahit Soner Börekçi, Mersin University, Mersin, Turkey

Kaelin Cawley, National Ecological Observatory Network, Boulder, CO, USA

Johannes Feldbauer, Technische Universität Dresden, Dresden, Germany

Isabella A. Oleksy, University of Wyoming, Laramie, WY, USA

Rachel M. Pilla, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Patricia Q. Tran, University of Wisconsin, Madison, WI, USA

Jacob A. Zwart, U.S. Geological Survey, Pittsburgh, PA, USA

Nicolas Azaña Schnedler-Meyer, Technical University of Denmark, Kongens Lyngby, Denmark

Tobias K. Andersen, Aarhus University, Aarhus, Denmark

Matthew R. Brouil, Washington State University, Pullman, WA, USA

Kate C. Fickas, U.S. Geological Survey, Sioux Falls, SD, USA

Alessandro Filazzola, ApexRMS, Ottawa, Ontario, Canada

Tyler V. King, U.S. Geological Survey, Boise, ID, USA

Nuria Sánchez-López, University of Idaho, Moscow, ID, USA

Victoria Stengel, U.S. Geological Survey, Austin, TX, USA

Dennis Trolle, Aarhus University, Aarhus, Denmark