

THE AEMON-J “HACKING LIMNOLOGY” WORKSHOP SERIES & VIRTUAL SUMMIT

Incorporating Data Science and Open Science in Aquatic Research

Michael F. Meyer , Robert Ladwig , Jorrit P. Mesman , Isabella A. Oleksy , Carolina C. Barbosa , Kaelin M. Cawley , Alli N. Cramer , Johannes Feldbauer , Patricia Q. Tran , Jacob A. Zwart , Gregorio A. López Moreira M. , Muhammed Shikhani , Deviyani Gurung, Robert T. Hensley , Elena Matta , Ryan P. McClure , Thomas Petzoldt , Nuria Sánchez-López , Karline Soetaert , Mridul K. Thomas , Simon N. Topp , and Xiao Yang

INTRODUCTION

Following the 2020 “Virtual Summit: Incorporating Data Science and Open Science in Aquatic Research” (DSOS; Meyer and Zwart 2020), a grassroots group of scientists convened the 2nd Virtual DSOS Summit on 22–23 July 2021. DSOS combined forces with the Aquatic Ecosystem M0deling Network - Junior (AEMON-J; <https://github.com/aemon-j>) to host a 4-d “Hacking Limnology” Workshop Series prior to the summit (13–16 July 2021). The aim was to focus more deeply on skill development and networking among early career researchers (ECRs), both of which are key to growing a workforce of data-intensive aquatic scientists (López Moreira M et al. in press; Meyer et al. 2021a). To support ECRs further, we hosted a virtual job board, where participants could note if they were either looking for employment or hiring for a position. Like the 2020 summit, there was high enthusiasm for both the summit and the workshops. In total, 686 people from over 50 countries registered for the AEMON-J Workshop Series and the DSOS Summit. Countries with the highest number of registrants included the United States (41%), Nigeria (20%), Canada (6%), Brazil (6%), and Germany (5%) (Fig. 1). To increase accessibility, there were no registration costs for the workshops and

summit, and we centralized introductory training materials, coding scripts, and presentation recordings in one community website (<https://aquaticdataciopencsci.github.io/>; Fig. 2), which we hope will continue to support the AEMON-J and DSOS communities over time.

THE AEMON-J “HACKING LIMNOLOGY” WORKSHOP SERIES

The “Hacking Limnology” Workshop Series took place over four consecutive days (13–16 July

2021), with each day having a specific theme: “Remote Sensing,” “Big (Ecological) Data,” “Machine Learning,” and “Numerical Modeling.” For each theme, there was an introductory talk that was available for online viewing 1 week prior to the workshop. Each day consisted of a pre-recorded 25-min keynote presentation that was live-streamed, followed by a live Question and Answer (Q&A) session, a 2-h hands-on workshop, and a 1-h session for brainstorming, technical support, and socializing. The underlying idea of

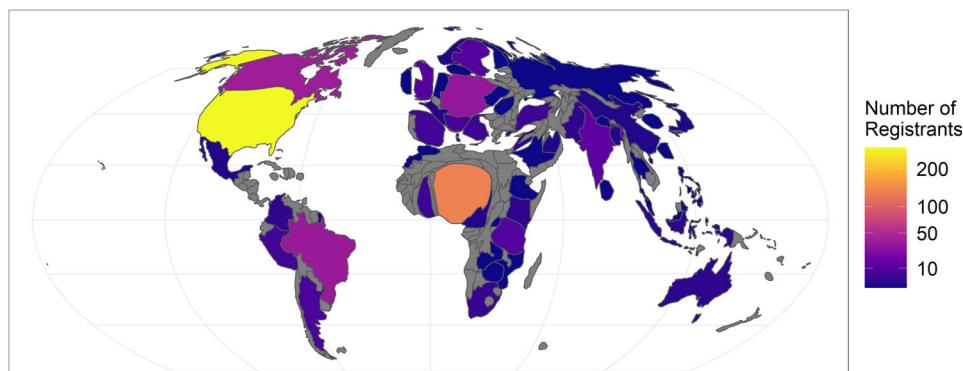


FIG. 1. Cartogram of registrant-reported country of residence for the “Hacking Limnology” workshop and the DSOS Virtual Summit (~95% of registrants). This map was made in the R Statistical Environment (R Core Team 2021) with the tidyverse (Wickham et al. 2019), janitor (Firke 2020), sf (Pebesma 2018), and cartogram (Jeworutzki 2020) packages. Countries are sized and colored by number of registrants. Larger, more yellow countries indicate more registrants. Smaller, more purple countries indicate fewer registrants. Countries colored in gray represent no registrants. All continents except Antarctica were represented. Among the over 50 countries with registrants, countries with the highest number of registrants included the United States (40.8%), Nigeria (20.3%), Canada (5.9%), Brazil (5.6%), and Germany (5.0%).

DSOS/AEMON-J Hacking Limnology



Virtual Workshop and Summit: Incorporating Data Science and Open Science in Aquatic Research

[Home](#) [Registration](#) [AEMON-J Workshops](#) [DSOS Summit](#) [Code of Conduct](#) [Material](#) [Jobs](#) [Info](#)

Advocating Open Data and Open Science in the Aquatic Sciences

The grass-root network DSOS (Data Science and Open Science) and AEMON-J (Aquatic Ecosystem M0deling Network - Junior) are joining forces this year to organize a 2nd ‘Virtual Summit: Incorporating Data Science and Open Science in Aquatic Research’ that includes fascinating presentations about aquatic research including open data and open science approaches, and multiple days of exciting workshops that include state-of-the-art keynote talks with live hands-on coding exercises. We are adapting the [Carpentries Code of Conduct](#) for our meeting.

FIG. 2. Front page of the workshop and summit website aquaticdataciopencsci.github.io which features subpages for registration (interactive form, “Registration”), an overview of each workshop day including on-demand videos and links to the scripts (“AEMON-J Workshops”), an overview of each summit day (“DSOS Summit”), the code of conduct (“Code of Conduct”), prerecorded material to introduce participants into various data science-related topics (“Material”), a job subpage (“Jobs”), and a general information subpage (“Info”).

the workshop program was that many ECRs might not have advanced training in these emerging methods and fields. While a short workshop would not make anyone an expert, it could provide the instruction, material, and confidence to initiate independent study. Each introductory talk provided a general overview about the topic as an aid to novice learners, and the keynote presentation highlighted cutting-edge research and, more broadly, the possibilities that the respective topic could offer. During the workshops, participants engaged in hands-on coding activities, for example using and receiving Google Earth Engine and R scripts that could serve as a starting point for their future work. During the last hour of each day, participants could elect to join virtual breakout rooms, where each breakout room had a dedicated theme. Core themes for each day of the workshop included: (1) informal Q&A and help sessions with workshop organizers, (2) casual discussion with keynote speakers, (3) brainstorming new ideas and collaboration opportunities, or (4) socializing with other participants. Prior to the breakout rooms opening, the organizational team identified facilitators to ensure the meeting's code of conduct was upheld within breakout groups and that participants had equitable time to voice questions and thoughts. Aside from creating an informal setting for participants to interact with each other and presenters, the social hour also provided an excellent opportunity for ECRs around the globe to network professionally and personally.

For participants' reference before, during, and after the workshop, the program and instruction materials were centralized on the event's website (<https://aquaticdatasciopensci.github.io/>). The introductory talks and keynote presentations were available before the workshop and were provided with subtitles in English. The workshop materials were hosted on public GitHub repositories. All the materials will remain on the website to foster re-use and to ensure their availability to people who could not attend the workshop. The workshop largely focused on using the R Statistical Environment (R Core Team 2021) for each day's theme, as it is widely used within the aquatic research community and is the preferred choice of many biologists and ecologists. The workshop materials covered the following topics:

- working with remote sensing data using Google Earth Engine (Gorelick et al. 2017) and the R Statistical Environment (R Core Team 2021);

- accessing, aggregating, and processing data from the National Ecological Observatory Network (NEON; <https://www.neonscience.org/>);
- fitting, visualizing, and interpreting machine learning models, focusing on random forests and artificial neural networks in the R Statistical Environment (R Core Team 2021); and
- solving ordinary differential equations in the R Statistical Environment (R Core Team 2021) using the deSolve package (Soetaert et al. 2010) to develop a nutrient-phytoplankton-zooplankton model.

Verbal and written feedback from participants indicated that early access to the presentations and the workshop material was appreciated, especially by attendees in time zones that made synchronous participation difficult. Participants also reported that they appreciated the topics, and the necessary prerequisite knowledge was amenable to beginners and intermediate/advanced attendees alike. A challenge during the workshops was to assess if everyone managed to follow along; some presenters used polls while others monitored the meeting chat feature. Compared to past in-person AEMON-J workshops, assessing participant comprehension during this year's virtual workshop was more challenging, but we hope to address this facet of virtual learning in future "Hacking Limnology"-type events.

VIRTUAL SUMMIT: INCORPORATING DATA SCIENCE AND OPEN SCIENCE IN AQUATIC RESEARCH

The week after the "Hacking Limnology" Series, workshop organizers collaborated with a grassroots group of ECRs to convene the 2nd annual "Virtual Summit: Incorporating Data Science and Open Science in Aquatic Research" (22–23 July 2021). The summit included four presentation sessions, which focused on "Big Data," "Data Intensive Models," "Tools and Software," and "Applications of Open Science," as well as two expert panels with the themes "Open Data for Open Science" and "Careers in Data Science and Open Science." Presentations included a wide range of topics such as cutting-edge tools for data visualization, approaches for analyzing and aggregating large datasets, and applying large datasets to new modeling frameworks to assist management and monitoring efforts.

Like last year, the 2021 summit included a mix of prerecorded presentations and live panels with presenters. For each presentation

session, 4–5 prerecorded 10-min talks were streamed sequentially over a video conferencing screen and then were immediately followed by a 20-min live Q&A session that was moderated by the organizers. During the Q&A session, speakers were given approximately 4 min to respond to questions or elaborate on their talk. To centralize questions, each session had a specific Q&A form with an option for attendees to leave contact information for follow-up discussions. This format allowed for continued engagement between speakers and members of the audience when questions could not be fully addressed in the scheduled time. Approximately one third of submitted questions contained contact information, and speakers appreciated the opportunity to contact persons directly to answer questions following the summit. Having questions submitted in written form also provided more equitable opportunities for participants to ask a question, especially in instances where non-native English speakers may not feel comfortable voicing a question in front of an audience. Additionally, this format for submitting questions also gave time for presenters to plan or script responses before the live Q&A session began.

The summit also included a panel discussion following each day's sessions. The first panel was themed "Open Data for Open Science" and included guests from seven different data repositories and providers including: the Environmental Data Initiative, National Ecological Observatory Network, Integrated Digitized Biocollections, Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems, Consortium of Universities for the Advancement of Hydrologic Science, Global Lake Ecological Observatory Network, and the Great Rivers Ecological Observatory Network. The second panel focused on careers in data science and open science. Panelists included professionals in freshwater sciences who elaborated on how they currently apply skills gained in academic research careers to government, nonprofit organizations, and industries in the private sector. The Careers Panel attracted particular interest and extended beyond the summit's main schedule into the virtual social hour, where attendees could continue to engage with panelists in a more casual setting. The summit's social hour immediately followed the Career Panel on the second day of the summit. As with the "Hacking Limnology" workshop, the summit's social hour included themed breakout rooms, where attendees could freely move between rooms. Prior to the last day of the summit, we

polled participants about their specific interests for the social hour, which helped us estimate the number of virtual rooms, themes, and facilitators necessary for the social hour. Breakout room themes included: (1) Data Science in Limnology and Oceanography, (2) Natural Resource Applications of Open Science, (3) Careers in Data Science and Open Science, and (4) Socializing. Anecdotally, the Careers and Socializing breakout groups drew the largest attendance, and attendees commented that they appreciated the more informal setting to engage with panelists from industry, academia, and government. As with the workshop, each breakout room contained at least one facilitator, who ensured that the meeting's code of conduct was upheld and that participants had equitable opportunities to ask questions and voice opinions.

The organizational team actively built off the 2020 summit and kept diversity, accessibility, and inclusion at the forefront. Although we did not conduct an official survey, the 2021 summit hosted a more balanced and diverse gender representation for speakers and panelists relative to the 2020 summit (Meyer and Zwart 2020). To promote accessibility during the summit, all pre-recorded presentations included closed captions in English that were vetted for accuracy. Additionally, speakers were encouraged to submit slides with Alt text (descriptive text embedded within an image) that could be made available to attendees using screen readers. To accommodate a globally distributed audience, we also provided links to all prerecorded presentations, so that talks could be watched asynchronously and questions could still be submitted via the Q&A forms. The 2021 summit also included initiatives to increase non-English participation and engagement. In particular, one streamed presentation was in Portuguese and included English subtitles. Seven language channels on the DSOS virtual workspace were created to facilitate conversation, mentoring, and networking in a non-English-centric space. The represented languages were Spanish, Portuguese, French, German, Russian, Polish, and Mandarin. Moreover, attendees were encouraged to create additional channels, in the event their native language was not one of the seven original channels. This virtual workspace is also active beyond the summit and can serve as a place of continued discussion, networking, and mentoring for the larger DSOS community throughout the year. Finally, to archive presentations and workshop materials, we created an Open Science Framework portal where all content,

including presentations, Alt text slides, and archived GitHub repositories, can be downloaded in a compressed format (Meyer et al. 2021a).

NEXT STEPS

Even with these initiatives, we recognize that there is more we can do to expand representation in the DSOS virtual summits. Yet, the organizational team is hopeful that the combined AEMON-J/DSOS community will continue as an inclusive space for all, regardless of background or experience. In particular, the organizational team is committed to ensuring that the AEMON-J/DSOS community (1) promotes the work and network development of ECRs and (2) encourages ECR leadership at its helm. This mission includes the new virtual Job Board initiative (<https://aquaticdataciopensci.github.io/jobs/>), which is designed to collate employment opportunities as well as contact information for those seeking employment in one space. In tandem with the jobs channel in the DSOS virtual workspace, we hope that the job board can facilitate virtual networking for ECRs in a manner similar to wearing physical ribbons at an in-person conference that reads "I am looking for a PostDoc position" or "I am hiring a PostDoc."

Beyond the virtual workshop and summit, the organizational team intends to sustain the current momentum through continued virtual and potentially in-person components. The past year has encouraged creativity and entrepreneurship among the scientific community, the very same energy and quality that sparked this virtual workshop and summit. As the scientific community eventually returns to in-person meetings or some sort of hybrid model, the organizational team is beginning to envision formats where regional hubs could congregate for in-person gatherings that all virtually connect to a main meeting. Alternatively, in-person meetings could take place in coordination with larger conferences, as was done with previous AEMON-J workshops. However, these previous in-person AEMON-J workshops tended to be smaller, with 15–25 participants, whereas 2021's virtual Workshop Series had a peak attendance of nearly 100 attendees. Such formats could enable more cost-effective travel, visa, and lodging fees relative to those normally incurred. When international travel may be more feasible, an alternative format could include virtual summits and workshops similar to the 2021 summit, with an asynchronous in-person component that is more geared toward product development and working groups. As suggested in Meyer et al. (2021b), these hybrid formats could facilitate aspects of the

in-person conference (e.g., hallway chats, side conversations, late-night brainstorming) but also enable widened participation through virtual connections. Regardless of the exact form that these summits or working groups may take, we are excited about these prospects and encourage individuals interested in co-convening a future virtual summit or workshop to contact us as we continue to grow this community.

ACKNOWLEDGMENTS

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. 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. RL is funded through an NSF ABI development grant (#DBI 1759865).

AUTHOR CONTRIBUTIONS

Authors are listed in the following order: Core author group (MFM, RL, JM, IAO), DSOS (CCB, KMC, ANC, JAZ) and AEMON-J (JF, PQT) leads in alphabetical order by surname, AEMON-J helpers in alphabetical order by surname (GALMM, MS), workshop organizers in alphabetical order by surname (Remote Sensing: NSL, SNT, XY; Big Ecological Data: KMC, RTH, RPM; Machine Learning: MKT, EM; and Numerical Modeling: TP, KS, DG). Website materials and design by: MFM, JM, JF, RL, CCB and PQT.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

References

- Firke, S. 2020. *janitor*: Simple tools for examining and cleaning dirty data.
- Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**: 18–27. doi:10.1016/j.rse.2017.06.031
- Jeworutzki, S. 2020. *cartogram*: Create cartograms with R.
- López Moreira M, G. A., M. Toffolon, and F. Höller. in press. Hitting the sweet spot of complexity: Reasons why the development of new custom-tailored models is still warranted

- and should be encouraged in aquatic sciences. *J. Limnol.* doi:10.4081/jlimnol.2021.2035
- 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., and others. 2021a. AEMON-J/DSOS archive: “Hacking Limnology” Workshop + virtual summit in Data Science & Open Science In Aquatic Research. doi:10.17605/OSF.IO/682V5
- Meyer, M. F., and others. 2021b. 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
- Pebesma, E. 2018. Simple features for R: standardized support for spatial vector data. *R J.* **10**: 439–446. doi:10.32614/RJ-2018-009
- R Core Team. 2021. R: A language and environment for statistical computing. Available from <https://www.r-project.org/>
- Soetaert, K., T. Petzoldt, and R. W. Setzer. 2010. Solving differential equations in R: Package deSolve. *J. Stat. Softw.* **33**: 1–25. doi:10.18637/jss.v033.i09
- Wickham, H., and others. 2019. Welcome to the tidyverse. *J. Open Source Softw.* **4**: 1686. doi:10.21105/joss.01686
- Jorrit P. Mesman**, Department F.-A. Forel for Environmental and Aquatic Sciences, University of Geneva, Geneva, Switzerland; Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden
- Isabella A. Oleksy**, Department of Zoology and Physiology, University of Wyoming, Laramie, WY, USA
- Carolina C. Barbosa**, São Carlos School of Engineering, Hydraulics and Sanitation Department, University of São Paulo, São Carlos, SP, Brazil
- Kaelin M. Cawley**, National Ecological Observatory Network Project — Operated by Battelle, Boulder, CO, USA
- Alli N. Cramer**, Institute of Marine Sciences, University of California Santa Cruz, Santa Cruz, CA, USA
- Johannes Feldbauer**, Institute of Hydrobiology, Technische Universität Dresden, Dresden, Germany
- Patricia Q. Tran**, Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, USA; Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA
- Jacob A. Zwart**, U.S. Geological Survey, Integrated Information Dissemination Division, South Bend, IN, USA
- Gregorio A. López Moreira M.**, Department of Ecohydrology and Biogeochemistry, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany; Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy
- Muhammed Shikhani**, Lake Research, Helmholtz Centre for Environmental Research — UFZ, Magdeburg, Germany
- Devijani Gurung**, Department of Civil and Environmental Engineering, Washington State University, Pullman, WA, USA
- Robert T. Hensley**, National Ecological Observatory Network Project — Operated by Battelle, Boulder, CO, USA
- Elena Matta**, Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy
- Ryan P. McClure**, Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA
- Thomas Petzoldt**, Institute of Hydrobiology, Technische Universität Dresden, Dresden, Germany
- Nuria Sánchez-López**, Department of Forest, Rangeland and Fire Sciences, University of Idaho, Moscow, ID, USA
- Karline Soetaert**, Department of Estuarine and Delta Systems, Royal Netherlands Institute of Sea Research (NIOZ), Yerseke, The Netherlands
- Mridul K. Thomas**, Department F.-A. Forel for Environmental and Aquatic Sciences, University of Geneva, Geneva, Switzerland; Institute for Environmental Sciences, University of Geneva, Geneva, Switzerland
- Simon N. Topp**, U.S. Geological Survey, Integrated Information Dissemination Division, Chapel Hill, NC, USA
- Xiao Yang**, Department of Earth, Marine and Environmental Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Michael F. Meyer, School of the Environment, Washington State University, Pullman, WA, USA; michael.f.meyer@wsu.edu

Robert Ladwig, Center for Limnology, University of Wisconsin-Madison, Madison, WI, USA