Breakout Workshop: Scientific computing and the Open Source software revolution: how can ocean colour science benefit

Chair: Joaquín E. Chaves (joaquin.chaves@nasa.gov)
Co-chair: Erdem M. Karaköylü, (erdem.m.karakoylu@nasa.gov)
Co-chair: Joel P. Scott (joel.scott@nasa.gov)

Description of Breakout Workshop:

Ocean color practitioners have principally relied on commercial off-the-shelf software (COTS) for data analysis (e.g., IDL, Matlab). Use, support, and maintenance of COTS requires paid licenses, and often come with proprietary, black-box features. This framework hinders task-oriented modification and is an obstacle to transparency, code sharing, and reproducibility. While COTS were instrumental in propelling a revolution in our understanding of the oceans, the restrictions associated with their use have now become an obstacle to innovation and the establishment of ocean color as a truly global discipline in terms of the diversity of data users and producers (i.e., mission-sponsoring nations).

However, during the past few years there has been an explosive growth in computational power, data availability, and the open source software movement. In combination, these factors have resulted in the democratization of advanced computational tools and platforms for diverse commercial and scientific applications. There is now a rich ecosystem of easily-accessible, open source (i.e. freely available and modifiable) software. This ecosystem includes programming languages, such as R, Python, and Octave, and their derivative software packages and libraries developed for domain-specific applications. These tools are now easily accessible via the internet and their use is reinforced with online resources such as GitHub and Bitbucket, as well as project management platforms such as the Open Science Framework. This shift in paradigm, has lowered the threshold for entry, expanded the user pool, and increased opportunity for collaboration, while promoting scientific innovation, transparency, and reproducibility. The rise of open source software, new data analysis methods, and greater computational power enables new approaches for answering ocean color research questions, conducting instrument calibration and algorithm validation, and streamlining data access, use, and availability.

Objectives:

- Highlight recent advances in ocean color science that have leveraged open source software (e.g., Python, R), modern computational methods (e.g., machine learning, artificial intelligence, neural networks, Google Earth Engine), and online software and knowledge repositories.

- Present Integrated Development Environments (IDEs) and how they are facilitating the scientific-computing open source revolution, making software development and data analysis more accessible across the ocean color community (e.g., Anaconda, RStudio, Jupyter-Notebooks)

- Show the benefit of modern data structures that enable easier access to ocean color data, leveraging the benefits of structured, self-describing data (e.g., NetCDF4, SeaBASS) in computational environments (e.g., object-oriented programming, Pandas data-frames, NumPy arrays) and relational databases (e.g., MySQL, NoSQL)

- Demonstrate the advantages of open source scripting languages (e.g., Python, Perl) for both *in-situ* data collection, as well as multi-level data processing, analysis, and visualization.

- Discuss novel academic and research deliverables, such as publishing processing and analysis code for reproducibility alongside peer-reviewed publications, presenting an overview of current platforms for publication and distribution (e.g., GitHub/GitLab, Journal of Open Source Software, Open Science Framework)

- Emphasize the potential for open source technologies to lower barriers and promote diversity within the ocean color community by increasing collaboration across geographic regions and stakeholders of differing national origins and formal educations and improving gender parity and engagement by underrepresented groups.