Fostering collaboration through improved software development practices for the ONEFlux eddy covariance data processing pipeline

Gilberto Pastorello\textsuperscript{1}, Carlo Trotta\textsuperscript{2,3}, Alessio Ribeca\textsuperscript{3}, Keith Beattie\textsuperscript{1}, Sy-Toan Ngo\textsuperscript{1}, Housen Chu\textsuperscript{1}, You-Wei Cheah\textsuperscript{1}, Danielle Christianson\textsuperscript{1}, Giacomo Nicollin\textsuperscript{2,3}, Sigrid Dengel\textsuperscript{1}, Diego Polidori\textsuperscript{3}, Peter Isaac\textsuperscript{4}, Matthew Archer\textsuperscript{5}, Dominic Orchard\textsuperscript{5}, Deb Agarwal\textsuperscript{1}, Sebastien Biraud\textsuperscript{1}, Margaret Torn\textsuperscript{1,6}, and Dario Papale\textsuperscript{3}

\textsuperscript{1}Lawrence Berkeley National Laboratory, Berkeley, United States of America
\textsuperscript{2}Euro-Mediterranean Centre on Climate Change (CMCC), Viterbo, Italy
\textsuperscript{3}University of Tuscia, Viterbo, Italy
\textsuperscript{4}TERN-OzFlux, Melbourne, Australia
\textsuperscript{5}University of Cambridge, Cambridge, United Kingdom
\textsuperscript{6}University of California, Berkeley, United States of America

Standardized processing of eddy covariance data is important for studies combining data from multiple sites, for validating remote sensing measurements as well as runs of ecosystem and climate models, and for applications relying on these flux data to create derived products like upscaled fluxes, among other examples. However, maintaining consistency within the software used for this processing while allowing for evolution of this code across research networks presents novel challenges in software development. The introduction of the ONEFlux (Open Network-Enabled Flux) eddy covariance data processing pipeline, originally developed within a collaboration of the AmeriFlux Management Project, the European Fluxes Database, and the ICOS Ecosystem Thematic Centre, supported the creation of consistently processed global eddy covariance data products. In particular, ONEFlux codes were used to generate the FLUXNET2015 dataset, which is widely adopted by thousands of eddy covariance data users in their work in research, ranging from soil microbiology to large scale drought effects, and also education, from basic plant biology all the way to global climate change. We are now more thoroughly instrumenting the code, and the code development process, to better address these challenges, efforts which we will describe in this presentation. In particular, we are seeking to improve software development practices to allow for more streamlined collaboration on expanding and contributing to the codebase. For instance, we are adopting planned release cycles for code updates, designing more detailed ways to incorporate and evaluate new modules, introducing data-centric testing and continuous integration, improving code performance, and adopting several other software engineering best practices more widely in the development workflows. The main goal of these changes is to lower the barriers for running ONEFlux by regional networks processing their data, while at the same time better supporting contributions from the community into the codebase. This will be critical to continue the current use of ONEFlux to generate updated versions of flux datasets by regional networks, the components of new global products.