

Development and management of tropical forest datasets for model simulation and benchmarking

Charuleka Varadharajan, Gilberto Pastorello, Danielle S. Christianson, Boris Faybishenko, Ping Hu, Ranjeet Devarakonda, Michael Crow, Terri Killeffer, Les Hook, Tom Boden, and Deb Agarwal

New and existing ecological, hydrological, and meteorological datasets from tropical forests in Central and South America are being collected and utilized in the U.S. Department of Energy's NGEE Tropics project. One of the primary goals of the NGEE Tropics data team is to enable efficient generation of data synthesis and analysis products for the project in a repeatable and scalable manner, by developing an agile infrastructure. Currently components of the infrastructure have been developed to standardize data and metadata collection activities, archive NGEE Tropics datasets, and provide data for model parameterization.

We have developed standardized templates to report metadata for various field measurements by working with the data collectors. The objective was to minimize effort required to archive a dataset, while maximizing the utility of search and synthesis across datasets. Metadata specifications were based on project data collection protocols, example data files submitted, and other related protocols, e.g., Smithsonian Tropical Research Institute (STRI) and AmeriFlux. We have built the NGEE Tropics Archive, an accessible and searchable repository, to host all data generated in the project, including field measurements, simulation results, and data products. The Archive leverages the NGEE Arctic data management system and complies with DOE Office of Science Digital Data Management Requirements. Data files generated in a common theme, such as an experiment, field campaign, data synthesis product, or a publication, will be grouped into collections for submission to the archive. Data collections will be archived following basic QA/QC checks, documentation, assignment of a DOI, and will be made available to the project team and public.

High priority datasets from the Archive will receive additional QA/QC, processing, and transformations. For example, in partnership with the modeling team and the data provider—STRI, we conducted QA/QC processing of meteorological data from Barro Colorado Island, Panama, to provide input parameters for numerical simulations using climate models. This involved gap-filling of missing data, removal of outliers, and flagging of suspicious segments of the time series, which allowed us to create a consistent dataset of meteorological drivers.

In the future, access to NGEE Tropics and other community data will be available through a data broker and web portal that will provide an integrated data search, visualization, analysis, and access interface. The underlying infrastructure is being developed to meet project needs to perform specific data queries, create synthesis products, initiate automated QA/QC, and transform data for model parameterization and benchmarking tasks.