Poster #21-83

Data Management and Assimilation for the Berkeley Lab Watershed Function SFA

Charuleka Varadharajan[1*], Roelof Versteeg[2], Boris Faybishenko[1], Valarie Hendrix[1], Matthew Henderson[1], and Deb Agarwal[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA;
[2]Subsurface Insights, Hanover, NH

Contact: cvaradharajan@lbl.gov

BER Program: SBR
Project: Berkeley Lab Watershed Function SFA
Project Website: watershed.lbl.gov

The Watershed Function SFA project generates heterogeneous datasets at its East River field site that include a variety of data types such as hydrological, geochemical, geophysical, microbiological and remote sensing data. The data are collected from various sources, including data generated by the project team (e.g. sensor data, geochemical sampling and remote sensing products), data from collaborators, and data from external sources (e.g. USGS and NRCS).

The objective of the Data Management and Assimilation (DMA) component of the SFA is to enable science by: (1) managing and archiving the data collected by the project, and releasing those data publicly with appropriate citation information, (2) enabling the project team and the broader community to find where, when and what types of data are being collected through an interactive portal, (3) performing quality assurance and quality control of priority datasets and (4) creating an data integration engine and search portal that can help retrieve, fuse and visualize the diverse data for further synthesis and analysis. The DMA team has developed a number of tools for data management and preservation, QA/QC, data discovery, search and visualization.

Data preservation and distribution are being enabled by a web portal that allows authorized users to upload and download data files as packages. The tool requires users to enter metadata needed to enable web searches and obtain DOIs for citation.

A public interactive map of data collection sites run by the SFA and its collaborators is available at https://wfsfa-data.lbl.gov/watershed/, to inform the broader community about SFA field activities. Sites can be filtered by their key measurements and other metadata, leading to detailed site landing pages.

Automated data QA/QC is performed using statistical methods to identify and flag issues in meteorological, water level and geochemical data. These data are used for example in estimating evapotranspiration estimates or building a set of model drivers (temperature, precipitation, solar radiation, etc.) from the network of meteorological stations, and the river discharge from field observations.

Data integration is achieved via a brokering service BASIN-3D that dynamically integrates data from distributed databases via web services, based on user queries. The integrated results are presented to users in a web portal that enables intuitive search, interactive visualization and download of integrated datasets.

These tools are used for building crosscutting data products needed for the hypothesis testing and numerical modeling of hydrological and biogeochemical conditions of the East River watershed by the SFA project teams.