A data-driven approach to predicting the impacts of hydrometeorological disturbances on water quality in river corridors

Charuleka Varadharajan,[1,*] , Helen Weierbach,[1] , Michaelle Lubich, [1,2], Mohammed Ombadi,[1] , Aranildo R. Lima, [3], Valerie C. Hendrix, [1], Danielle S. Christianson, [1], Catherine Wong, [1,2]

[1]Lawrence Berkeley National Lab, Berkeley, CA, USA
[2]University of California Berkeley, Berkeley, CA, USA
[3]Aquatic Informatics, Vancouver, Canada

Contact: (cvaradharajan@lbl.gov)

Project Lead Principal Investigator (PI): Charuleka Varadharajan

BER Program: ESS

Project: Early Career Project

Project Website: http://sites.google.com/lbl.gov/inaiads

Project Abstract:

Hydrometeorological disturbances such as floods, droughts, and heatwaves are projected to increase over the next few decades due to climate change. These disturbances can worsen water quality by impacting water temperature and salt, nutrient, contaminant concentrations, which will have direct consequences for human and ecosystem health. Hence, it is important to understand and predict how water quality in streams and rivers will respond to new disturbance regimes. Here we describe results from a U.S. Department of Energy (DOE) early career project that utilizes data-driven approaches to understand how these climate-induced disturbance events impact water quality over time. Our study is focused on water temperature and conductivity (salinity) predictions in three U.S. hydrological regions - the mid-Atlantic, Pacific Northwest and Upper Colorado. First, we present the use of low-complexity machine learning models (Multiple Linear Regression, Support Vector Regression, and Random Forest Regression/XGBoost) to predict monthly stream water temperature at the point to watershed to regional scales. We compare the performance of these models between basins with differing climate, geological, land use and water management attributes. In particular, we focus on improving model performance by separating free-flowing and dammed stream reaches, and by grouping stations by their time-series dynamics. Next, we present results from analysis of change in regional water temperatures during heat waves. Finally, we present the iNAIADS framework (iNtegration, Artificial Intelligence, Analytical Data Services) which comprises a data integration tool BASIN-3D (https://basin3d.readthedocs.io/) that can be used to reproducibly synthesize diverse data from distributed sources, along with reusable analytical and machine learning codes.