A Novel Random Forest Approach to Revealing Interactions and Disentangling Inland and Oceanic Controls Over Coastal Phytoplankton Productivity

YIWEI CHENG[1], VED BHOOT[1], KARL KUMBIER[2], MARILOU SISON-MANGUS[3], MICHELLE NEWCOMER[1],

[1] Lawrence Berkeley National Laboratory, Berkeley, USA
[2] University of California, Berkeley, USA
[3] University of California, Santa Cruz, USA

Coastal zones play a large role in the global carbon cycle and primary productivity. Exports from coastal watersheds and seasonal upwellings represent terrestrial and oceanic nutrient sources respectively, that when combined with favourable environmental conditions, lead to enhanced phytoplankton productivity (blooms). The increasing occurrence of blooms across the land-ocean interface impacts biogeochemical cycling and poses risks to ecosystems.

A novel iterative Random Forests (iRF) machine-learning model was developed to identify key governing factors and stable interactions between significant drivers. We applied iRF to coastal bloom datasets along the Californian coast to (1) predict temporal patterns of remotely sensed chlorophyll a, and (2) assess stable interactions between bloom drivers in response to coastal upwelling and inland watershed exports. Datasets combined *in-situ* measurements of coastal environmental conditions and inland nutrient fluxes calculated from the California Environmental Data Exchange Network (CEDEN), a collective environmental database. The analysis differentiates key ocean versus inland watershed controls on blooms. We reveal for the first time, interactions between these marine and inland controls that affect phytoplankton growth. Our framework can help to identify coastal regions vulnerable to perturbation due to changes in land use, temperature and precipitation patterns.