EGU21-6395
https://doi.org/10.5194/egusphere-egu21-6395
EGU General Assembly 2021

Estimating phytoplankton species populations in irrigation ponds from drone-based imagery and *in situ* water quality sensing and sampling

Jaclyn E. Smith[1,2,3], Billie J. Griffith[2], Matthew D. Stocker[1,2,3], Moon S. Kim[2], and Yakov A. Pachepsky[2]
[1]University of Maryland, Environmental Science and Technology Department, College Park, MD, United States of America (jesmith@umd.edu)
[2]Environmental Microbial Food Safety Laboratory, United States Department of Agriculture, Beltsville, MD, United States of America
[3]Oak Ridge Institute of Science and Education, Oak Ridge, TN, United States of America

Phytoplankton is known to affect freshwater habitats of pathogenic and indicator organisms in irrigation water sources. Cyanobacteria are associated with producing harmful toxins which can be transferred to crops, and the gene transfer between phytoplankton and pathogens is of interest particularly in connection with the antibiotic resistance in microorganisms. The objective of this work was to evaluate the possibilities of estimating phytoplankton populations in irrigation ponds by using separate and combined in situ water quality sensing/sampling and sUAS imagery. The study was conducted during a 5-month summer-early fall period at a working irrigation pond on Maryland's Eastern shore, USA. In situ physical, biochemical, and nutrient measurements were taken at 34 locations in the pond with a total of 21 parameters. Phytoplankton species were enumerated using a modified Ütermohl method and then grouped into green algae, diatoms, and cyanobacteria. The imagery was obtained from an altitude of 120 meters using three modified GoPro cameras and a MicaSense camera. It was then clipped to represent the area around locations of sensing/sampling. The measured parameters were grouped into physical, biochemical, nutrient, and imagery datasets as inputs. Various combinations of these inputs constituted 17 different datasets used with machine learning algorithms. The target variables were the three groups of phytoplankton and the proportion of cyanobacteria in the total count of observed phytoplankton cells. The regression tree (RT) algorithm was applied to research the structure of the dataset and to determine the major influential variables. The random forest (RF) algorithm was applied to estimate the target variables for each of the 15 total datasets. With the RT analysis, nutrient concentrations appear to be influential for green algae and cyanobacteria proportion. After the nutrients were added to the physical and biochemical parameters in the RT analysis for these specific variables, the R^2 went from 0.782 to 0.869 and from 0.678 to 0.758, respectively. The imagery alone provided moderate RT accuracy for green algae (R^2=0.661) and cyanobacteria (R^2=0.586), but less for diatoms (R^2=0.483). The RT analysis provided good estimates for green algae with the R^2 of 0.756 but was not efficient for diatoms (avg. R^2=0.524), cyanobacteria (avg. R^2=0.284), nor the proportion of cyanobacteria (avg. R^2=0.524). In the random forest study,

the most important predictors for green algae and cyanobacterial proportion were nutrient concentrations of potassium and calcium, respectively. MicaSense imagery at the red edge and near-infrared parts of the spectrum were among the most important predictors. The drone-based imagery provided information useful for the estimation and prediction of green algae. Influential input variables were different amongst phytoplankton groups. For the 17 input datasets, the overall accuracy increased in the sequence imagery < physical < biochemical < nutrient water quality parameters as inputs.