WALTER SCOTT, JR.
COLLEGE OF ENGINEERING
COLORADO STATE UNIVERSITY

AGU Hydrology Days

Capturing the Spatio-Temporal Variability of Groundwater Pumping Leveraging Public Domain Data and Machine Learning: An Assessment of Machine Learning Application for Data Scare Regions

Authors: Dawit W. Asfaw*[1], Ryan Smith*[2], Sayantan Majumdar[3], Katherine Grote[4], Venkataraman Lakshmi[5], Bin Fang[5], James Butler[6], Brownie Wilson[6]

[1]Department of Geoscience, Colorado State University Fort Collins, CO, USA, [2]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, USA, [3]Division of Hydrologic Sciences, Desert Research Institute, Reno, NV 89512, USA, [4]Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO, USA,[5]Department of Engineering Systems and Environment, University of Virginia, Charlottesville, Virginia, USA, [6]University of Kansas, Lawrence, KS, USA

Abstract. Groundwater level decline has increased globally threatening major aquifer systems' sustainable supply of water for irrigated crop production. Anthropogenic driven groundwater depletion is mainly attributed to agriculture. Groundwater pumping data collection is very limited around the world due to lack of policies. Future groundwater sustainability depends on monitoring of groundwater pumping. Machine learning has been used in some areas to estimate withdrawals, but data paucity has limited wider application of these approaches. The reliability of outcomes from machine learning analysis heavily relies on available data quality and quantity. Few studies have used machine learning techniques to study groundwater withdrawals in regions where data is abundant. Nevertheless, the data quality and quantity requirements to produce a robust estimate of groundwater withdrawals are not well studied. In this study, we built point scale groundwater withdrawal prediction machine learning models using a Random Forest algorithm. Data is split into training and testing where the training data is used to build the model and assess the model's ability by comparing the model prediction values with the testing data. The point scale prediction values are aggregated over a 2 km by 2 km grid. We evaluated a combination of different training and testing split to understand model performance variability. We performed the analysis in the Northwestern Kansas Groundwater Management District 4. The model used public domain remote sensing, land surface model output, and hydrogeological variables for the period from 2008 – 2020. We observed that a model trained on 10 % of the total available data showed coefficient of determination (R^2) values of 0.96 and 0.77 for training and testing, respectively. Knowledge of crop irrigation area enabled estimate aggregation over a grid, and we find that aggregation of estimates improved the spatial and temporal groundwater withdrawals estimates. The result of this study has a significant implication for effective groundwater management in regions where there is limited data.

Key words: Groundwater withdrawals, irrigation, remote sensing, machine learning, estimation