Random Forest predicts river substrate characteristics using airborne Lidar data and imagery

Abstract

River ecosystems involve complex interactions in several spatiotemporal scales and commonly require large field survey data collection to allow meaningful investigation. Insufficient comprehended natural phenomena and adequate data availability encourage the application of machine learning algorithms. This study aimed to investigate if Random Forest (RF) can predict riverbed clogging and grain size on a stretch of the Inn river (Germany) through topographic variables computed from given airborne lidar data and imagery. Gravel bars overlaying sieved sediment and hierarchical clustering analysis provided grain size classes, and expert visual inspection provided classified regions through a standardized clogging levels system. The performance measurements of the RF models have shown: (1) accurate predictions for low and high clogging levels (2) accurate prediction for one (out of four) grain size class; (3) that the predictors computed from detrended digital elevation model, flow velocity, distance from the water channel, and meandering morphology have major influence over the model's classification. These results indicate a possibility to map clogging severity through airborne data and illustrate how limited data affects an RF model regarding grain size predictions.