

EGU21-719

<https://doi.org/10.5194/egusphere-egu21-719>

EGU General Assembly 2021

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Interpreting reach-scale classifications and the role of spatial-morphological variables in river channel mapping using machine learning algorithms

Adeyemi Olusola and Adetoye Faniran

University of Ibadan, Department of Geography, Ibadan, Nigeria (olusolaadeyemi.ao@gmail.com)

Over the years, there has been tremendous growth in the literature as regards river channel classifications, however, very few studies have been able to engage the use of remote sensing products in channel classification at the reach-scale level especially by combining reflections from satellite sensors with channel morphological variables. This study aims to identify discriminating spatio-morphological variables using machine learning algorithms and classify site-specific channel types at the reach scale. Each reach was broadly classified based on valley settings (confined, partly confined and unconfined) and channel types (alluvial or bedrock). However, variations and site observations were recorded for site-specific classification purposes. For each reach, Global Positioning System devices were used to geo-locate their endpoints. Standard field instruments were used for cross-sectional measurements and established hydraulic equations for the derived variables. A total of 249 points across 83 reaches were sampled during the fieldwork. Landsat 8 and Sentinel-1 bands were retrieved for days the fieldwork was carried out/for days close to those dates using Google Earth Engine (GEE) platform. Hierarchical cluster analysis, HCA, using Ward's linkages was used to provide a classification for the channel types. For the identification of important variables in predicting channel unit types, the random-forest - recursive feature elimination (RF-RFE) algorithm was used using the `rfe()` function. To identify the best machine learning algorithm, random-forest (rf), support vector machines (svm), multivariate adaptive regression spline (mars) extreme gradient boosting (xgb) and adaptive boosting (adaboost) were used on the training and test data to identify the best performing algorithm. The `rfe()` feature selection identified five (5) variables that can significantly help in channel unit type identification. The top five variables are dimensionless stream power, slope, width, wetted perimeter and Band 4. Using ROC curve, sensitivity, and specificity, the mars model has the highest ROC curve. Hence, it appears to be the best performing out of the five. However, if the argument is to be based on positive prediction, then any of the models except adaboost will be preferred given their high sensitivity. The HCA using illustrated the clustering structure of the studied reaches by producing five distinct channel classification types distinguished based on width-depth ratio values (high and low). The five distinct channel types are listed as M1e, M5e, B1, E5b, and E. These codings are based partly on Rosgen's classification while, the capital letters (M, B and E) represent mixed channels, bedrock with moderate width-depth ratio and alluvial channels with low width-depth ratio respectively. Numbers 1 and 5 represent bedrocks and sandy beds

based on slope variation respectively. The identified channel unit types are a result of the underlying lithology, process-form dynamics and confinement. As streams are expected to respond differently to shocks and recover from damages, it becomes essential to understand these differences in classification which will go a long way in establishing watershed and streamside management guidelines.