Investigating the Drivers of Site-Level Forest Soil Moisture Dynamics

Jake Cohen[1,2], Yushu Xia[2], Jamie Duan[2], Hancheng Guo[2,3]

[1]University of Toronto, [2]Lamont-Doherty Earth Observatory of Columbia University, [3]Zhejiang University

Remote sensing-based measurements of environmental covariates have proven effective at modeling soil moisture at high spatio-temporal resolutions, but often struggle to produce accurate readings in forested areas, where closed canopies and complex topography can interfere with surface reflectance observations. Spatial autocorrelation is another–often neglected–component of digital soil mapping (DSM), which has the potential to inflate model performance and obscure variable contributions. In this study, we integrated *in situ*, 10 m resolution measurements of soil moisture from a 1-hectare plot within the Lamont Sanctuary Forest (LSF) experimental site in the northeastern U.S. with remotely-sensed environmental covariates and a digital elevation model. Our goals were to (1) construct accurate machine-learning models of soil moisture; (2) assess and mitigate autocorrelation to improve our models' generalization ability; (3) understand the relative influence of factors on forest soil moisture. We evaluated random forests (RF), eXtreme Gradient Boosting (XGBoost), and Gaussian Process Regression (GPR), using a buffered block cross-validation strategy. The RF models demonstrated the best performance, with an R^2 of 0.794 and a root mean squared error (RMSE) of 0.046 m^3/m^3. Adjusting for spatial autocorrelation improved our model's ability to handle novel samples by reducing data leakage. Downscaled land surface temperature (LST) was the overall most important predictor, followed by topographical features like slope and biotic indices like the normalized difference moisture index (NDMI). These results highlight key controls on soil moisture that could be used to inform targeted forest management.