The identification of meaningful variables from Sentinel-2 time series data for effective tree species classification

Tiziana L. Koch, Marius Rüetschi, and Lars T. Waser
Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Remote Sensing, Birmensdorf, Switzerland
(tiziana.koch@wsl.ch)

Sentinel-2 time series provide large amounts of data and information which can be easily used to classify tree species with machine learning algorithms. In addition to the original Sentinel-2 bands, further data such as indices, phenological metrics or even synthetical images can be derived. While tree species classifications highly benefit from such additional data resulting in improved prediction accuracy, severe drawbacks have to be considered - For large data sets, large storage is needed, the computation time expands and a linkage to ecological or phenological reasons behind the usage of these variables can hardly be drawn. Therefore, the implemented variables should be limited to the ones, which are meaningful and on the same time providing the best prediction accuracy. To identify meaningful variables from original Sentinel-2 images and the additionally calculated data first we used basic correlation analyses and subsequently feature selection methods in combination with the commonly used Random Forest algorithm. We classified the most common forest tree species in the Swiss canton of Grisons, which is mountainous and characterized by diverse landscapes. The presented approach will lead to higher efficiency for classifying tree species and additionally provides potential conclusions regarding ecological patterns beyond the distinction of tree species by remote sensing data. Moreover, the proposed approach can also be used to improve classifications or predictions of other outcome variables for vegetated areas with Sentinel-2.