Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: optimization of variable selection on optical/LiDAR-derived predictors

Hooman Latifi[1*], Arne Nothdurft[2], Barbara Koch[1]

1- Dept. Of Remote sensing and Landscape information systems, University of Freiburg, Tennenbacherstr. 4, D-79106 Freiburg-Germany.
*- Corresponding Author. Email: hooman.latifi@felis.uni-freiburg.de, Tel: +49 7612033699
2- Dept. of Biometry and Informatics, Forest Research Institute Baden- Württemberg, Wonnhaldestraße 4 79100 Freiburg- Germany

Abstract

Recently, the non-parametric nearest neighbor imputation methods proved to be successful for spatial predictions of standing timber volume in forest landscapes. In the present study this method was applied to estimate total timber volume and biomass using high dimensional covariates derived from spectral and 3D topographic remote sensing. In a mixed temperate forest landscape in southwestern Germany, a combination of multiple remote sensing predictors derived from CIR orthoimages, Thematic Mapper visible/infrared data and small footprint discrete return LiDAR were used for nonparametric predictions of total timber volume and biomass using three different distance measures including the euclidean distance, the Mahalanobis distance and the "Most Similar Neighbor"-distances, the latter being derived by weights obtained from canonical correlation analysis. Additionally, the regression-tree based approach of Breimans Random Forests was tested for spatial predictions. In order to reduce the full set of 113 covariates, a prior selection based on an evolutionary genetic search algorithm (GA search) was applied. Furthermore, the spatial predictions were mapped for 20m grid cells. Prediction errors (RMSD), computed by cross-validations, were compared across all examined methods, as well as data sources. The reduction of covariates by means of the GA search lead to higher precisions of the nearest neighbor predictions based on euclidean and Mahalanobis distance measures, while the RMSD of both MSN and the random forests approaches have increased after covariate selection. . Due to the strong correlations among covariates from the same source of remote sensing, the GA search proved to be unstable in multiple runs, as it suggests differing optimal variable sets for different classifications of the response variable. However, prediction errors with different sets of covariates obtained from multiple GA-runs show only slight differences. Moreover, the LiDAR-based height metrics proved to be the most frequent covariates selected by GA search for both response variables. The predictions based on both, the MSN-approach and the random forests prove to be superior to the other examined approaches. Furthermore, both methods come with an additional benefit as they make usage of the full, unreduced data set.

Keywords: Standing volume and Biomass prediction, nonparametric methods, Optical/LiDAR data, predictor variable, feature selection algorithm