Global patterns of tree wood density

Hui Yang¹, Vitus Benson¹, Yahai Zhang¹, Rackhun Son¹, Siyuan Wang¹, Weijie Zhang¹, Yuzhen Zhang¹, Claire Robin¹, Dmitry Schepaschenko², Zbigniew Karaszewski³, Sterenczak Krzysztof⁴, Álvaro Moreno-Martínez⁵, Cristina Nabais⁶, Thomas Ibanez⁷, Ghislain Vieilledent⁷, Ulrich Weber¹, and Nuno Carvalhais¹

¹Max-Planck-Institute for Biogeochemistry, Biogeochemical Integration, Germany (huiyang@bgc-jena.mpg.de)
²International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
³Department of Wood Investigation and Application, Łukasiewicz Research Network-Wood Technology Institute, Winiarska 1, 60-654 Pozna´n, Poland
⁴Department of Geomatics, Forest Research Institute, Sękocin Stary, Braci Leśnej 3 Street, 05-090, Raszyn, Poland
⁵Image Processing Laboratory (IPL), Universitat de València, València, Spain
⁶Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal
⁷AMAP, Univ Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France

Wood density is an emergent property resultant of tree growth strategies modulated by local edapho-climatic and stand development conditions. It is associated with the biomechanical support of trees and hydraulic conductivity or safety, directly and indirectly influencing a range of ecological processes, including, among others, tree growth, tree resistance to disturbances, and mortality. Tree wood density is also crucial for assessing vegetation carbon stocks by supporting the link between a volumetric retrieval and a mass estimate. Earlier studies based on tree-level wood density measurements have reported significant relationships between wood density, environmental conditions, and tree growth strategies. However, these were either regionally focused or suffering from data availability, lacking a representative large-scale and spatially explicit representation of factors influencing tree wood density. This study collects and collates information from several sources to construct a global database of 28,822 tree-level wood density measurements alongside with a wide set of climate, soils, topography, and Earth observation covariates to support the development of statistical models for wood density. The dataset, consisting of more than 3,000 global covariates, is used for training different machine learning models, including random forest model (RF), light gradient boosting model (LGBM), extreme gradient boosting model (XGBoost), and bagged trees models. The experimental design considers six cross-validation approaches: either random 5-fold; according to two sets of climate classifications, land cover types, ecozones, or latitudinal ranges. Model performances are assessed with the coefficient of determination ($R^2$) and Root-mean-square errors (RMSE) when predicting an independent test subset of wood density. The top ten models show a prominent performance ($R^2 > 0.67$ and RMSE < 0.09), and their ensemble mean, and standard deviation are considered the best estimation and uncertainty in wood density predictions, respectively. Systematic underestimation biases are observed around the low northern latitudes (0°-20°N), primarily due to the lack of wood density measurements. Further analysis of sources of uncertainties and their
quantification support the generation of a global quantitative and spatially explicit representation of wood density. The ecological interpretation and quantitative assessment of global wood density, and associated uncertainties aim to contribute to improving predictions of vegetation biomass and inferring ecosystem resistance under current and future climate scenarios.