|  |
| --- |
| **A Continental-Scale Canopy Height Map at 30-m from ICESat-2 and Ancillary Data** |
| **Introduction/Aim:** Global canopy height maps from spaceborne lidar offer opportunities for affordable estimation and monitoring of canopy height for a multitude of vegetation applications worldwide. The main objective of this study was to use canopy height information from the ICESat-2 (Ice, Cloud, and land Elevation Satellite-2) datasets along with ancillary Landsat, LANDFIRE and topographic variables to produce a 30-m canopy height product across the contiguous United States.**Methods:** To fill in canopy height values between ICESat-2 tracks, we implemented a canopy height regression model using non-parametric gradient boosted regression trees regression implemented in the XGBoost library. The ICESat-2 canopy height was the dependent variable and the Landsat, LANDFIRE and topographic (slope and aspect) variables, were the independent variables. We selected XGBoost based on its demonstrated better performance over other models, Random Forests (RF) and Neural Networks, in our prior work. We fit the canopy height regression model using 85% of the downloaded data for training with the remaining 15% set aside for testing the performance of the model and control the over-fitting during training. The performance metrics used for assessing model fit included the coefficient of determination (R2), mean bias (Bias), mean absolute error (MAE) and their equivalent percent metrics, percent bias (pBias) and percent MAE (pMAE). We analysed variable importance in the canopy height modeling using SHAP (SHapley Additive exPlanations) values.**Results:** Overall, the fitted model achieved an R2 value of 0.76, a mean bias of 0.1 m, a general over-estimation, and MAE (mean absolute error) of 2.5 m on a validation dataset. Model performance varied across biomes (R2 = 0.49 - 0.73, mean biases = -0.3 to 0.4 m, MAE values= 1.4 to 3.3 m). The best performance was observed in Mangrove sites while the worst was observed in Mediterranean forests. Accuracy assessment with independent airborne lidar canopy heights also varied across biomes with predictions moderately correlated with reference data (R2 = 0.36 -0.58) but lower precision (MAE = 1.4 m – 6.1 m) against various height percentiles including the 50th, 75th, 90th,95th, 98th percentile and the maximum. We observed the best agreement between airborne and ICESat-2 canopy height metrics for corresponding 98th percentile estimates across all biomes.**Conclusion:** In this study our goal was to develop and evaluate a canopy height regression model across different biomes using ICESat-2 heights and multisource ancillary datasets at the extent of continental US. Future efforts will expand the current scale to ultimately produce a global product. The continued accumulation of height data by ICESat-2 will allow a densified sample of canopy heights globally, which will ultimately produce a better product of canopy height with future iterations. |

*Please note the headings provided are optional, please remove or change to best suit your content.*