



SiDroForest Siberian Drone-mapped Forest inventory

Femke van Geffen¹, Birgit Heim¹, Ulrike Herzschuh¹, Luidmila Pestryakova², Evgenii Zakharov^{2,3}, Ronny Hänsch⁴, Begüm Demir⁵, Birgit Kleinschmit⁶, Michael Förster⁶, and Stefan Kruse¹

¹Alfred-Wegener-Institut für Polar- und Meeresforschung, Polar Terrestrial Environmental Systems Research Group, Berlin, Germany

²Institute of Natural Sciences, North-Eastern Federal University of Yakutsk, Yakutsk, 677000, Russia 10

³Institute for Biological Problems of the Cryolithozone, Russian Academy of Sciences, Siberian branch, Yakutsk, 677000, Russia

⁴German Aerospace Center (DLR), Microwaves and Radar Institute, SAR Technology, Germany

⁵Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, Germany

⁶Institute for Landscape Architecture and Environmental Planning, Geoinformation in Environmental Planning Lab, Technische Universität Berlin, Germany

To gain a better understanding of global carbon storage and albedo feedback mechanisms it is important to have insights into high latitude vegetation change. Boreal forest compositions are changing in response to changes in climate, which in turn can lead to feedbacks in regional and global climate through altered carbon cycles and albedo dynamics. Circumpolar boreal forests represent close to 30% of all forested area on the planet, between 900 and 1,200 million ha. These forests are located primarily in Alaska, Canada, and Russia. Due to the remote location of these forests and the short seasons without snow, data collected on the boreal vegetation is limited.

The proposed dataset is an attempt to remedy data scarcity whilst providing adjusted data for machine learning practices. We present a dataset containing diverse formats of forest structure information that covers two important vegetation transition zones in Siberia: the Evergreen - Summergreen transition zone in Central Yakutia and the northern treeline in Chukotka (NE Siberia).

This dataset contains data from the locations covered by fieldwork was performed by the Alfred Wegener Institute for Polar and Marine research, (AWI) and the North-Eastern Federal University of Yakutsk (NEFU). The fieldwork upscaled through the addition of Red Green Blue (RGB) UAV (Unmanned Aerial Vehicle) camera data and Sentinel-2 satellite data cropped to a 5 km radius around the fieldwork sites. The dataset is created with the aim of providing ground truth validation and training data to be used in various vegetation related machine learning tasks .

The dataset contains:

1. Labelled individual trees per 30x30 m plot assigned in field work with additional data on species, height, crown width, and biomass.
2. Structure from Motion (SfM) point clouds that provide 3D information about the forest structure,

included generated Canopy Height Model (CHM), Digital Elevation Model (DEM) and a Digital Surface Model (DSM) per 50x50 m.

3. Multispectral Sentinel-2 satellite data (10 m) cropped to a 5km radius with generated a NDVI(normalized difference vegetation index), available in three seasons: Early Summer, Peak Summer and Late Summer.

4. Extracted tree crowns with species information and a synthetically generated large (10.000 samples) dataset for training machine learning algorithms.

The dataset will be made publicly available on the data repository PANGAEA.