**SDP Vegetation Structure and Radiation Data Products – Descriptions of Draft Processing Pipelines**

**Ian Breckheimer**

**4-27-2020**

## Source Datasets:

1. NEON AOP 1m LiDAR Digital Elevation Model (DEM).
2. NEON AOP 1m LiDAR Digital Surface Model (DSM).
3. NEON AOP 1m NDVI, NDSI, NDWI, and RGB maps (NDVI, NDSI, NDWI).
4. Airborne Snow Observatory (ASO) 3m bare-earth DEM.
5. A basic deciduous vs evergreen vegetation classification derived from the NEON AOP spectrometer and lidar data (Falco et al. 2020).
6. Building Footprints from the 2018 Bing Maps Imagery-derived building footprints data, supplemented with additional heads-up-digitized buildings.

## Processing for Draft Products:

### Topography:

#### Corrected NEON Digital Elevation Model

The original NEON DEM had five types of significant errors: (1) Artificial flattening of sharp ridges associated with the algorithm used to select ground returns in the LiDAR data. (2) Positive errors for low dense vegetation resulting in mounds on floodplains that are actually trees and shrubs. (3) Positive errors in deep snowfields. (4) Large errors in small swaths with low densities of returns, most notable in the Copper Creek drainage. (5) Incomplete masking of buildings, resulting in some building elevation included in the DEM. All DEM processing was done to address these errors.

1. Most of the ridges affected by flattening errors were relatively bare of vegetation, and so were corrected by substituting in values from the AOP DSM in pixels with low NDVI values.
2. High-NDVI pixels on ridges were removed, and the resulting gaps were filled using cubic spline interpolation using the r.fillnulls module is GRASS GIS.
3. Errors associated with floodplain vegetation were addressed by removing pixels which had high NDVI, were higher than surrounding cells (local maxima), and were within 4m of the elevation of the nearest stream channel. Elevation relative to stream was calculated with the “Vertical Distance to Stream” module in SAGA GIS.
4. Resulting small gaps in the DEM (after removing trees and shrubs) were filled using cubic spline interpolation using the r.fillnulls module in GRASS GIS.
5. Errors associated with snowfields at high elevation were addressed by substituting values interpolated from the 3m Airborne Snow Observatory DEM collected in mid-September 2018. ASO values were used for all pixels which were (1) classified as snow in the landcover map and (2) had ASO DEM values lower than the AOP DEM.
6. Errors associated with low point density had obvious triangulation artifacts and were manually identified in the hillshaded DEM. Values from the 2016 Lidar DEM were substituted in the problematic areas. Areas beyond the extent of the 2016 DEM were patched with values from the National Elevation Dataset, resampled to the 1m grid using bilinear interpolation.
7. Pixels that intersected known building footprints were removed and filled using cubic spline interpolation.

#### Boundary-filled DEM

To ensure that solar radiation products have realistic topographic shading around the boundaries of the map, we created a version of the NEON DEM that substituted coarser-scale elevation data from the ASO DEM for areas within 500m of the boundary of the NEON DEM. To improve alignment with the NEON DEM, the following steps were taken with the ASO DEM before compositing:

1. The ASO DEM elevation values were presented as Height Above Ellipsoid, were not ground-controlled and had a datum that was incorrectly specified, resulting in systemic bias in elevation values relative to the AOP DEM. These errors were corrected by regressing the AOP and ASO DEM’s in the area of overlap, and applying the regression coefficients to the ASO DEM. The resulting map agreed well with the AOP DEM in the area of overlap, and had an RMSD of approximately 0.35m.
2. The harmonized ASO DEM had erroneous small pits and peaks which were filtered and gap-filled. All areas less than 9m^2 that were greater than 1m above or below the mean of the adjacent pixels in the ASO DEM were removed, and missing pixels were filled using cubic spline interpolation.

#### Digital Surface Model

The NEON-provided Digital Surface Model had relatively few identified problems, but suffered from triangulation errors in areas of low point density. Where available, values in these areas were patched using the 2016 Lidar DSM. Where these estimates were not available, pixels were removed and filled using cubic spline interpolation. DEM values were substituted where interpolated values in the DSM fell below the elevation of the DEM

#### DEM with Structures

This map was produced by substituting DSM values in the DEM for pixels that intersect the building footprints dataset.

#### DSM leaf-off

This map was produced by substituting DSM values into the bare-earth DEM for areas that intersect building footprints or were classified as evergreen trees and shrubs in a vegetation map derived from the Falco et al. vegetation classification. The Falco et al. map was reclassified into a binary deciduous / evergreen map, and missing pixels (mostly shadows) were imputed by segmenting individual tree canopies using a watershed algorithm applied to a map of canopy heights (derived from subtracting the DEM from the DSM). Missing values in each canopy polygon were filled with the most common deciduous / evergreen classification in each canopy polygon. Canopies with zero classified pixels (~8% of canopies) were assigned a class according to the most common class within 10m of the focal pixel. All calculations were done with the GRASS GIS modules r.mapcalc, r.neighbors, r.watershed, r.clump, and r.stats.zonal.

#### Gross Landcover

This map divides the UER domain into 7 classes depending on their dominant land cover type, and was generated by compositing a set of binary classifications each generated using methods tailored to the cover type.

1. The domain was first divided into vegetated and non-vegetated cover types using an NDVI threshold.
2. Non-vegetated cover types (bare ground, snow) were manually classified using a rule-based classifier. Maps used for this classification were three band ratios derived from the NEON AOP spectrometer data: NDVI, NDSI, NDWI, as well as a single map representing average reflectance in the visible spectrum.
3. Vegetated cover types were further subdivided using the AOP Lidar-derived canopy height model. Vegetated pixels with canopy heights less than 1m in height were designated as Meadow / Subshrub unless they were classified as evergreen in the Falco et al. vegetation map (described below).
4. Pixels with canopy heights greater than 1m were classified as deciduous or evergreen using the Falco et al. vegetation classification map (a supervised classification based on an SVM ensemble). The resulting pixel-wise classification had considerable variability (salt-and-pepper) as well as a high proportion of missing pixels in areas of shadow in the AOP data.
5. To improve the deciduous vs evergreen forest classification, individual tree canopies were delineated using a least-cost segmentation algorithm applied to the canopy height map (described below), and the modal deciduous vs. evergreen class for each canopy was computed. A small proportion of canopies (~8%) did not have any classified pixels, and were assigned the class that was most common within 10m of the canopy centroid.
6. A map of surface water was produced by combining a pixel-level water classification from the NEON spectrometer data with a thresholded flow accumulation map. The resulting map provides a more complete picture of surface water than a pixel-based classification alone, since water is obscured by shadow in large streams flowing through forested landscapes.
7. Buildings were delineated by rasterizing manually-corrected building footprints derived from the 2018 Microsoft Bing building footprints dataset. Data from this computer-vision based classifier was visually checked, and footprints of buildings constructed after the Bing imagery was collected were added.
8. The final landcover map was composited using the following ruleset: Deciduous Forest > Conifer Forest > Buildings > Water > Meadow > Snow > Bare, where classifications for the first class listed take priority over those listed later.

#### Canopy Height

We generated a map of vegetation canopy height (a canopy height model, CHM) by differencing the DEM and DSM, and then filtering the resulting map to remove non-vegetated pixels. We used an NDVI threshold to remove surface objects (such as towers and other structures) that were not associated with vegetation.

#### Canopy Segment ID

We delineated putative tree and shrub canopies and assigned them a unique integer value. Canopy segmentation was done using a cost distance algorithm is GRASS GIS. The processing steps are described below:

1. The peaks of canopy crowns were found using a computer-vision based terrain classifier (“geomorphon”, Stepinski et al. 2011) applied to the canopy height model, implemented in the GRASS GIS module “r.geomorphon” with a maximum radius of 5m. This algorithm assigns a pixel to a peak if its elevation is greater than all 8 pixels along each cardinal direction that are in focal pixel’s viewshed (up to the maximum radius).
2. The resulting pixel-based classification of crown peaks was then then segmented to assign a unique identifier for each geographically distinct peak using the GRASS GIS module “r.clump”.
3. Crowns were then delineated from the CHM using a cost-distance algorithm. Crown peaks were considered the starting points, and the cost surface was taken as an additive combination of slope and the inverse of canopy height. This means that costs grow more slowly with distance for taller and more rounded (less steep) crowns. The GRASS GIS module “r.cost” was then used to assign each pixel with a height greater than 0.5m to the crown peak that was nearest by the cost distance metric.

#### Canopy Segment Maximum Height & NDVI

We used the GRASS GIS module “r.stats.zonal” to compute the maximum height and NDVI value for each crown delineated in the canopy segmentation.

#### Bare-earth Potential Solar Radiation

We used the GRASS GIS module “r.sun” to estimate total clear-sky potential solar radiation for four days of the year (spring equinox, summer solstice, fall equinox, and winter solstice), based on the bare-earth DEM resampled to 3m resolution. The r.sun algorithm uses a ray-tracing approach to estimate topographic shading effects on direct, diffuse, and reflected components of incoming radiation given calculated solar geometry and prescribed atmospheric effects. The algorithm was run at 0.5-hour time-steps, and the resulting values were numerically integrated to derive total accumulated radiation for a given day. Other than calendar date and location, the algorithm has two other parameters, the Linke atmospheric turbidity coefficient which affects the proportions of direct and diffuse radiation and ground surface albedo, which influences the reflected radiation component. Seasonal mean values for each of these were taken from the literature (Remund et al. 2003 for the Linke coefficient, Schaaf et al. 2015 for albedo).

#### Subcanopy Solar Radiation

We estimated clear-sky potential solar radiation below the forest canopy (at a height of approximately 1m) using an approach adapted from Bode et al. (2014).

1. The Bode et al. model modifies bare-earth direct and diffuse radiation estimates based on a light penetration factor (LPF) related to canopy density.
2. We assumed that the light penetration factor scaled with the natural log of canopy height, dropping to a minimum of 0.1 for canopies greater than 20 m tall.
3. A seasonally varying asymmetric neighborhood (kernel) function was then used to integrate the LPF over parts of the sky through which sunlight passes, accounting for differences in sun angle over the year, and the bare-earth radiation estimate was then multiplied by the smoothed LPF to derived understory radiation estimates.
4. These subcanopy measurements were then combined with estimates of solar radiation derived from the DSM (which estimates tree crown and building shadows for non-forested areas.
5. In line with Bode et al., the final subcanopy maps were taken as the maximum of the modified bare-earth radiation estimates, and the estimates derived from the DSM.
6. For estimates on the summer solstice and fall equinox, the canopy height model was used to derive the LPF. For estimates on the spring equinox and winter solstice (when deciduous trees are bare), we used the land-cover map to generate a “leaf-off” DSM and canopy height model that totally removes canopy effects for deciduous vegetation in winter.

#### References

Bode, C. A., Limm, M. P., Power, M. E., & Finlay, J. C. (2014). Subcanopy Solar Radiation model: Predicting solar radiation across a heavily vegetated landscape using LiDAR and GIS solar radiation models. Remote sensing of environment, 154, 387-397.

Falco N ; Balde A ; Breckheimer I ; Brodie E ; G. Brodrick P ; Chadwick K D ; Chen J ; Dafflon B ; Henderson A ; Lamb J ; Maher K ; Kueppers L ; Steltzer H ; Wainwright H ; Williams K ; S. Hubbard S (2020): Plant species distribution within the Upper Colorado River Basin estimated by using hyperspectral and LiDAR airborne data. Watershed Function SFA. doi:10.15485/1602034

Remund, J., Wald, L., Lefèvre, M., Ranchin, T., & Page, J. (2003). Worldwide Linke turbidity information.

Stepinski, T. F., & Jasiewicz, J. (2011). Geomorphons-a new approach to classification of landforms. Proceedings of geomorphometry, 2011, 109-112.

Schaaf, C & Wang, C. (2015). MCD43A1 MODIS/Terra+Aqua BRDF/Albedo Model Parameters Daily L3 Global - 500m V006. NASA EOSDIS Land Processes DAAC.