Mapping Vegetation Canopy Structure and Distribution for Great Smoky Mountains National Park Using LiDAR

Jon Weiner[1], Jitendra Kumar[2], Steve P. Norman[3], William W. Hargrove[3], Nathaniel Collier[2], Forrest M. Hoffman[2]

[1]University of California Berkeley, CA, [2]Oak Ridge National Laboratory, Oak Ridge, TN, [3]USDA Forest Service, Southern Research Station, Asheville, NC

Introduction

Objective: Utilize high resolution LiDAR to map vertical canopy structure and distribution for Great Smoky Mountains National Park (GSMNP)

- Multiple-return LiDAR (Light Detection and Ranging) is a remote sensing tool that gathers high resolution 3D point cloud data.
- We processed and analyzed multiple-return LiDAR to investigate vertical canopy structures and their spatial distribution in GSMNP.
- We want to correlate vertical canopy structure with vegetation and validate with existing vegetation maps.
- *Big question*: Can LiDAR-based canopy structure improve vegetation mapping and monitoring efforts?

Data

- High resolution LiDAR point cloud data sets were obtained from the Great Smoky Mountains National Park Service.
- 1,658 flight miles were flown in February–April 2011 covering 1,400 sq. km, including the Tennessee section of GSMNP and the Foothills Parkway.
- The data sets consisted of 724 tiles in "*las*" format, each 1,500m x 1,500m and non-overlapping, which totaled 94 GB in size (Figure 1).
- A Python–based workflow was developed to process the files in an embarrassingly parallel fashion on a multi-core machine.

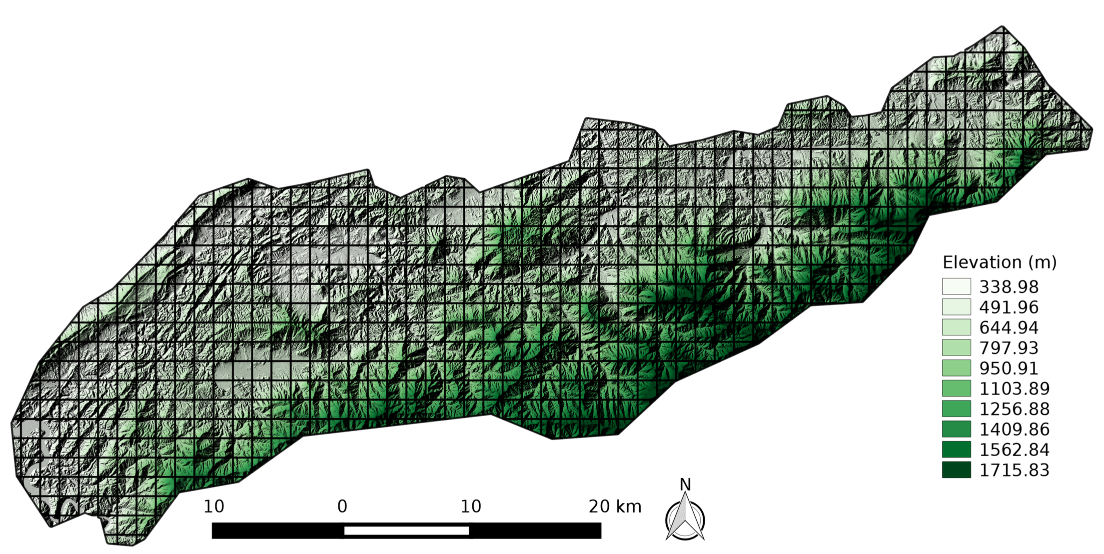

Figure 1: *Tiled LiDAR files overlain on a 1.5m resolution digital elevation map*

Methods

- To match LANDSAT and NLCD resolution and to obtain a reasonable sample size per sample cell, we gridded the park at 30m × 30m horizontal resolution.
- So we could compare results across sites at different elevations, we detrended the data from the underlying topography to only retain the above ground elevation, seen in Figures 2(a) and 2(b).
- The data had some significant anomalies, such as extremely high elevations (>75m) and negative elevations, which we filtered out during processing.
- We used a 1m vertical resolution to bin the LiDAR points in each 30m × 30m cell, then used these bins to create vertical canopy structures.
- Once we created these vertical canopy structures, we used a *k*-means cluster analysis algorithm to classify the landscape according to canopy structure.
- We then prepared maps of the spatial distribution of the canopy structures, and compared them to vegetation maps to determine the correspondence of canopy structures to vegetation type.

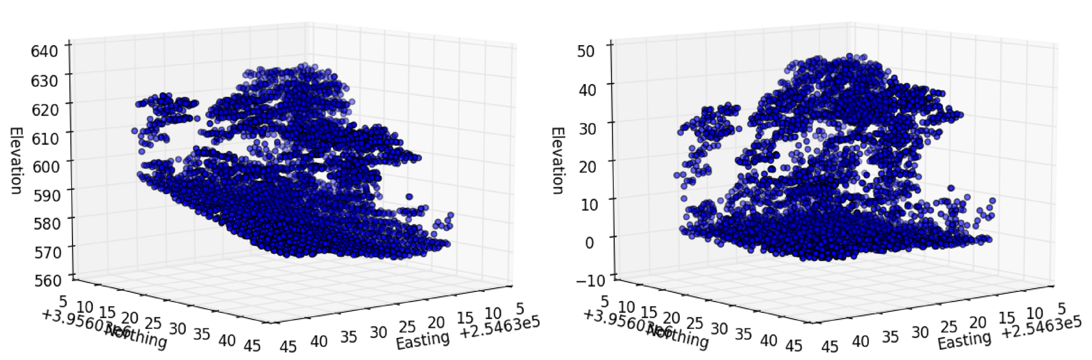

Figure 2: *(a) Raw LiDAR point cloud taken from a typical GSMNP cove forest that shows the influence of the underlying topography (b) Detrended LiDAR point cloud that shows above ground elevations*

Vertical Canopy Structures and *K*-means Clustering

- Vertical canopy structures were constructed using counts of LiDAR returns in 1m vertical bins post filtration and detrending.
- The *k*-means clustering algorithm classified each cell's vertical structure based on its similarity to random seeds.
- When we did the *k*-means cluster analysis, we determined that 30 classes gave a good balance between discriminating unique canopy structures and minimizing outliers.
- The 30 unique canopy structures from the cluster analysis are shown in Figure 4, color coded to the spatial map in Figure 5.
- In Figure 4, note the two most frequent structures are 28 and 18 with over 7% of the total area each, while 3 and 11 are outliers covering less than 0.2% of the total area each.

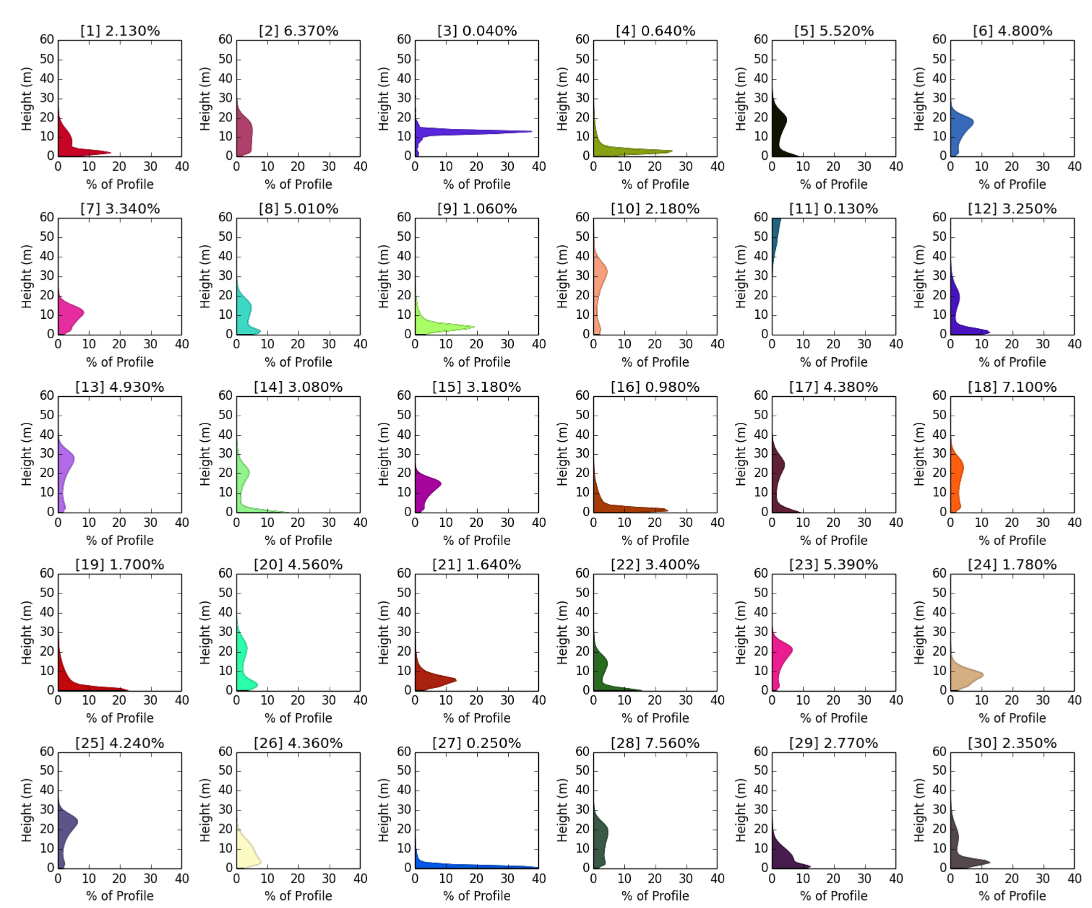

Figure 3: *30 canopy structures with percent map coverage*

Spatial Distribution of Vertical Canopy Structures

- Figure 5 shows the spatial distribution of the 30 canopy structures from Figure 4 across the Tennessee side of GSMNP.
- Low height vegetation regions, which often have high levels of noise, were separately classified as type 0, seen in yellow in Figure 5.
- Figure 6 shows the distribution of maximum canopy heights.
- Gridlines in Figures 5 and 6 were part of the original data (most probably due to processing errors) and hence were unable to be removed.

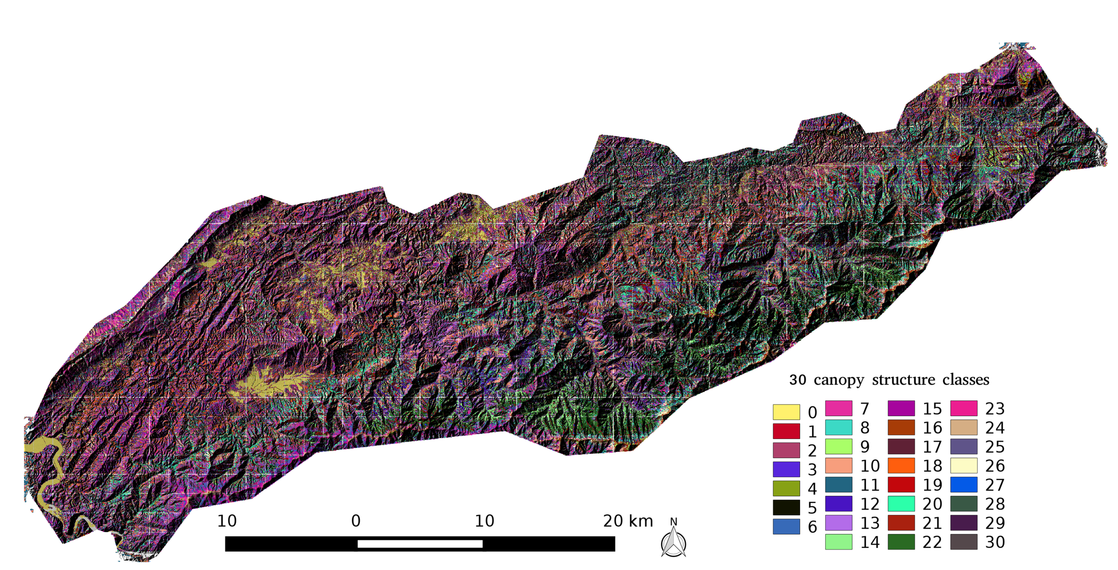

Figure 4: *Spatial distribution of 30 canopy structures*

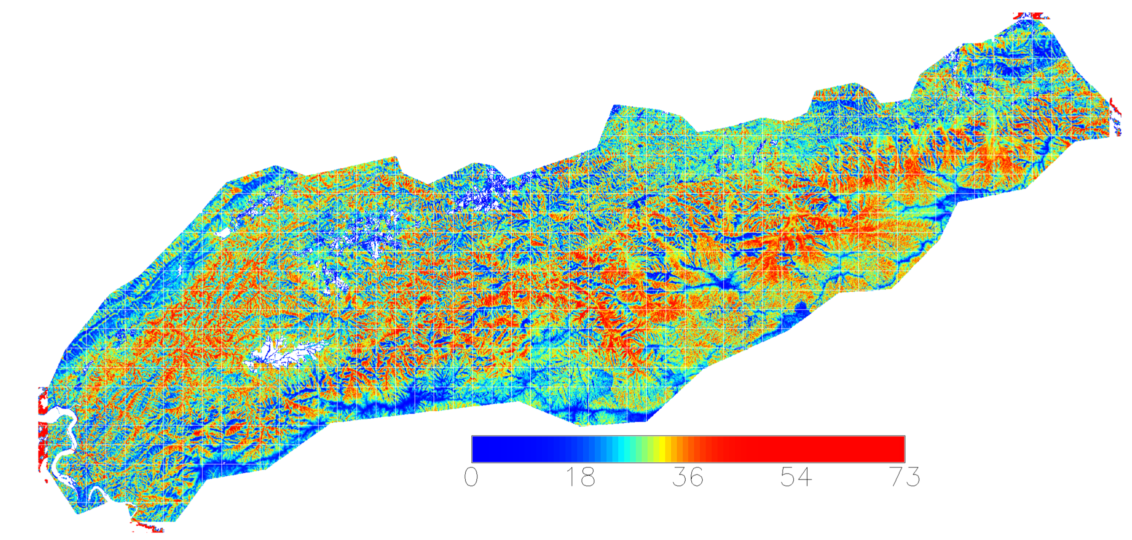

Figure 5: *Spatial distribution of the maximum canopy height*

Validation of Analysis

- We validated our results with the best available vegetation maps from GSMNP, which were available at the same 30m spatial resolution as our LiDAR analysis, and performed two case studies.
- The first case study was at Great Smoky Mountains Institute at Tremont (GSMIT) sites that are primarily surrounded by montane cove and hemlock forests that are known to have tall, dense canopies.
- Comparisons between Figures 6(a) and 6(b) show high correlation between the LiDAR-derived classes 10 and 13, some of the tallest and densest structures, and the montane cove vegetation surrounding the GSMIT sites, which are marked with blue circles.
- Our second case study was at Cades Cove, a popular tourist destination in the park dominated by short (<1m) successional grasses.
- Figure 7(a) shows Cades Cove with the LiDAR-derived class 0, indicating vegetation shorter than 1m, and this correspondence well with the vegetation map of Cades Cove shown in 7(b).

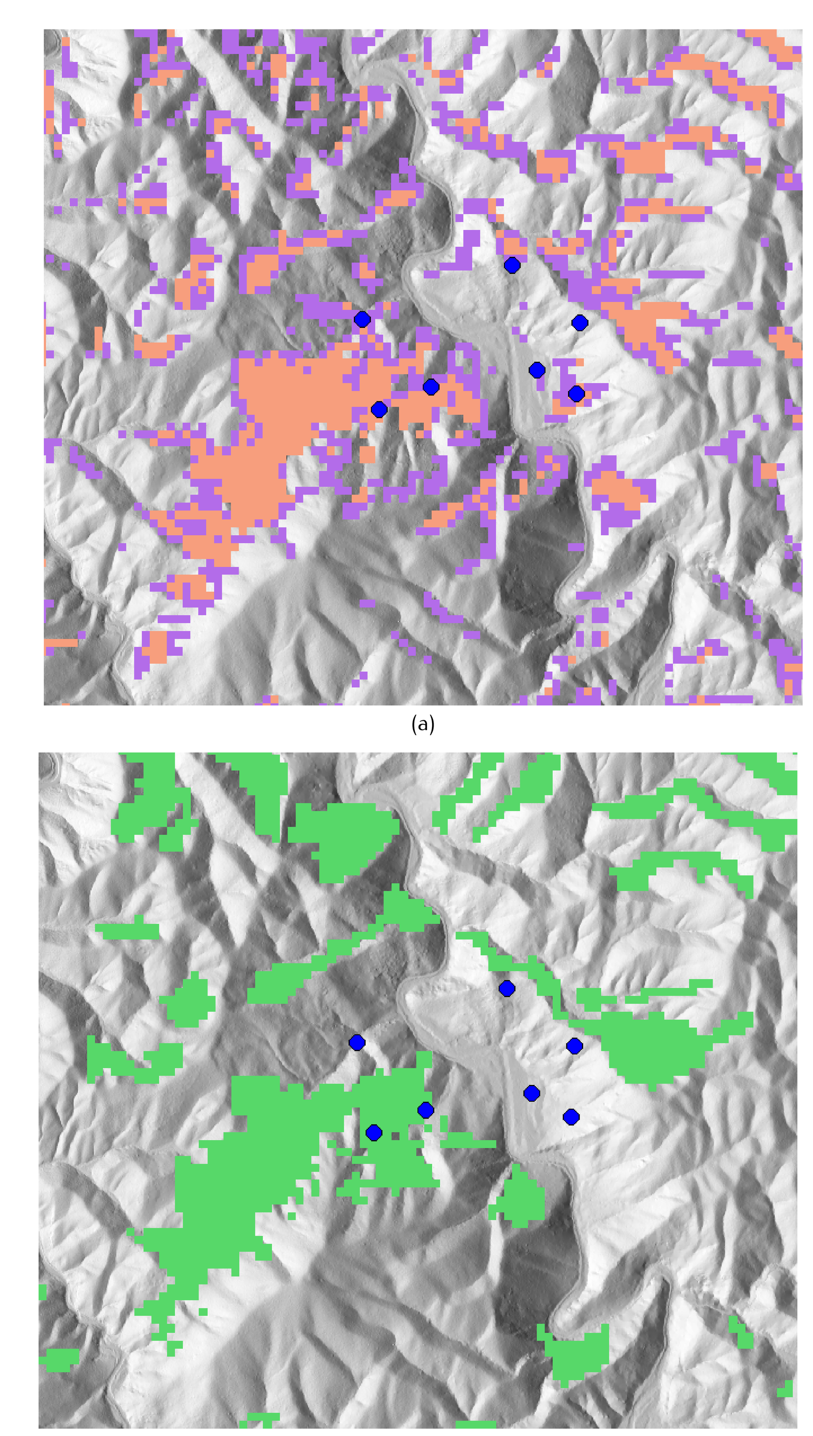

Figure 6: *(a) LiDAR classes 10 and 13 around the GSMIT sites indicated with blue dots (b) Montane cove forest as mapped by [1]*

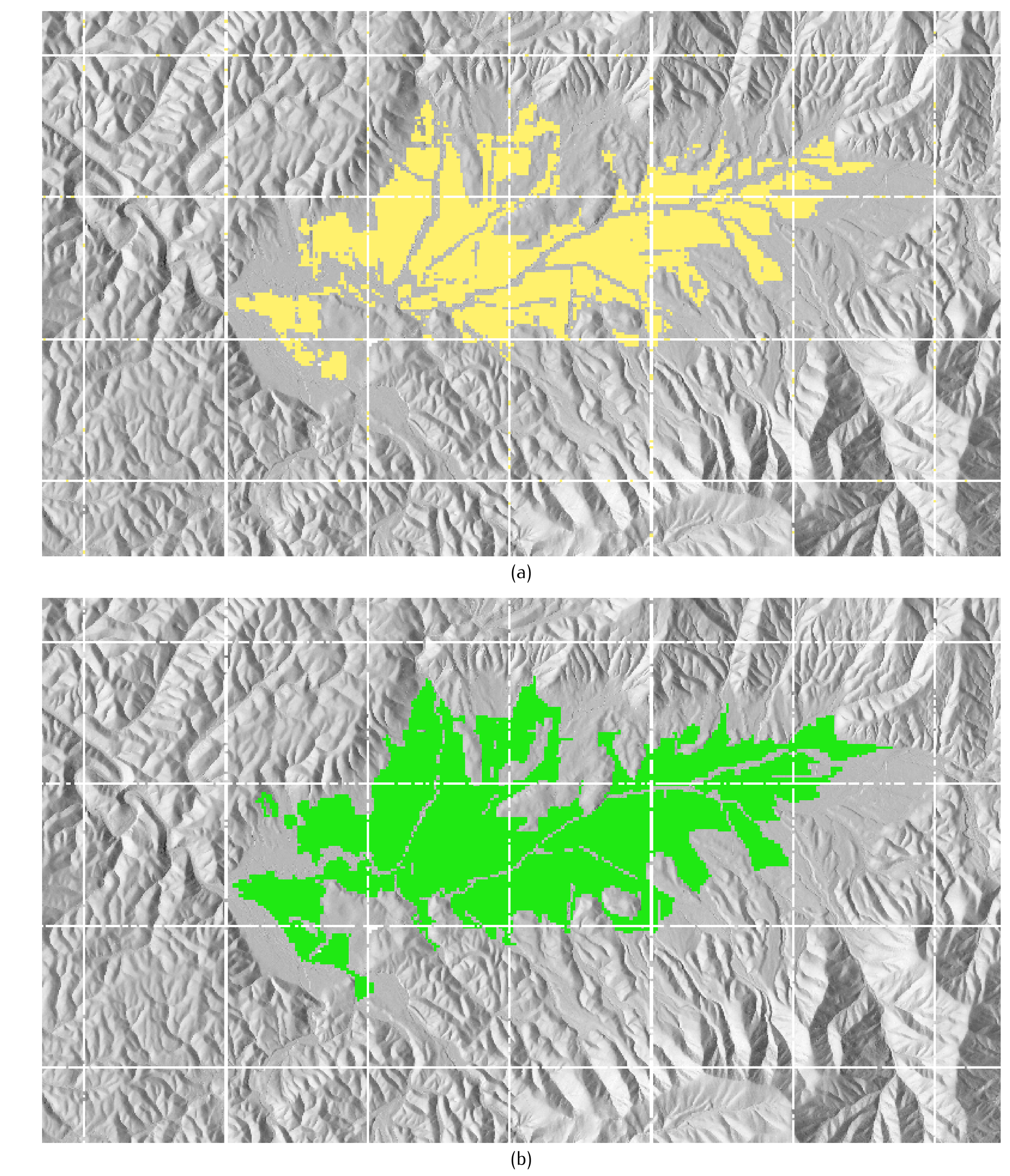

Figure 7: *(a) LiDAR class 0 around Cades Cove, indicating low-height (<1m) vegetation (b) Successional or modified vegetation in the Cades Cove area mapped by [1]*

Summary

- Using LiDAR data sets provided by the National Park Service, we created vertical canopy structures and then classified them using a *k*-means clustering algorithm.
- We took the spatial distribution of the canopy structures and compared them to vegetation maps to validate our results.
- Two initial validation tests show strong correlation between canopy structure and vegetation, but further case studies are needed to establish confidence.
- This method offers the ability to discern remote vegetation using LiDAR and could guide future high resolution vegetation mapping efforts by the NPS.

References and Data Products

1. M. Madden, "Overstory Vegetation at Great Smoky Mountains National Park, Tennessee and North Carolina, Reference Code: 1047498," 2014.[Online]. Available: https://irma.nps.gov/App/Reference/Profile/1047498
2. Kumar, Jitendra, Jon Weiner, William W. Hargrove, Steven P. Norman, Forrest M. Hoffman, and Doug Newcomb. November 17, 2015. "*Characterization and classification of vegetation canopy structure and distribution within the Great Smoky Mountains National Park using LiDAR.*" In Proceedings of the 15th IEEE International Conference on Data Mining Workshops (ICDMW 2015), pages 1478–1485. Institute of Electrical and Electronics Engineers (IEEE). *doi:10.1109/ICDMW.2015.178*

Data available at:
Kumar, J., J. Weiner, W.W. Hargrove, S.P. Norman, F.M. Hoffman, and D. Newcomb. 2015. LiDAR-derived Vegetation Canopy Structure, Great Smoky Mountains National Park, 2011. ORNL DAAC, Oak Ridge, Tennessee, USA. *doi:http://dx.doi.org/10.3334/ORNLDAAC/1286*

Contact: Jon Weiner
Email: jonweiner@berkeley.edu