Characterization and Classification of Vegetation Canopy Structure and Distribution within the Great Smoky Mountains National Park using LiDAR

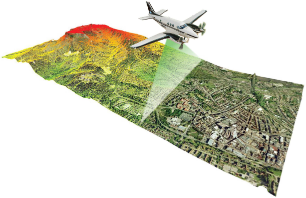
Jitendra Kumar[1], Jon Weiner[2], William W. Hargrove[3], Steven Norman[3], Forrest M. Hoffman[1], Doug Newcomb[4]

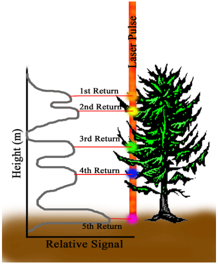
[1]Oak Ridge National Laboratory, [2]University of California Berkeley, [3]USDA Forest Service, [4]U.S. Fish and Wildlife Service

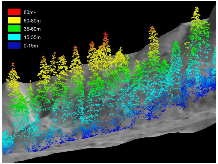
US – International Association of Landscape Ecology, Asheville, NC, USA April 06, 2016

Introduction

- Forests form a complex mosaic of diverse tree and coexisting plant and animal species.
- The structure of vegetation reveals information about stand age and height; forest composition, health, and disturbance; and suitability as habitat for birds and other animal species.
- Airborne Light Detection and Ranging (LiDAR) enables large scale remote sensing of topography, built infrastructure, and vegetation structure.
- Multiple laser “returns” produce “point clouds” used to map the ground surface, buildings, roads, and utility infrastructure, and to reconstruct the structure of vegetation canopies.
- Large data volumes pose significant computational challenges to employing LiDAR to monitor and manage forests and animal habitats.

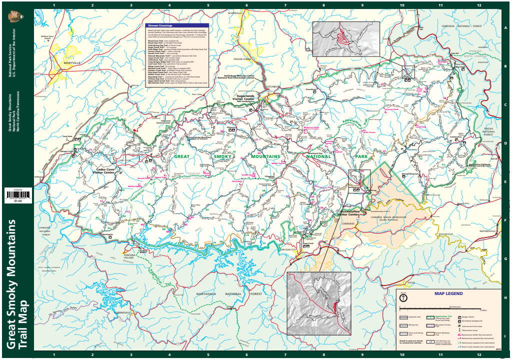


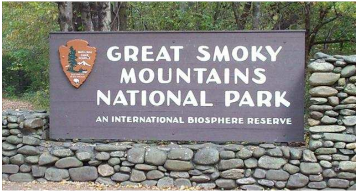




Great Smoky Mountains National Park (GSMNP)

- The GSMNP is the most visited national park in the U.S., and it hosts a rich ecosystem of plants and wildlife.
- The Park encompasses 816 sq. miles in Tennessee and North Carolina and ranges in elevation from 876 to 6,643 feet above mean sea level.





Objectives

- Develop computational tools and worflow for processing and knowledge extraction from massive LiDAR data sets
- Map and characterize the 3-D structure and distribution of the vegetation canopy in GSMNP

LiDAR Tiles for GSMNP

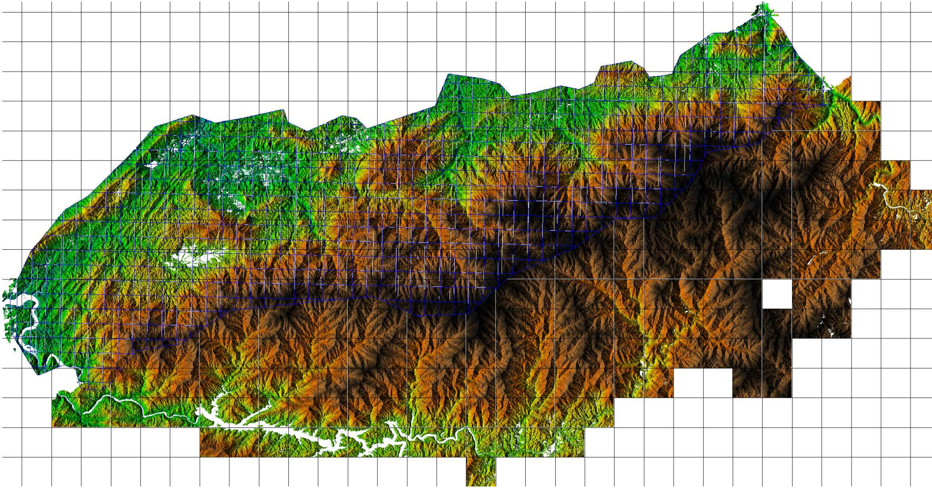
Tennessee

- LiDAR data for 540 sq. miles of the Tennessee portion of the GSMNP and the Foothills Parkway from 1,658 flight miles were collected during February–April 2011 by the U. of Georgia and Photo Science, Inc.
- Four multiple discrete returns per pulse were collected at a rate of 20.2 Hz from a nominal flying height of 1,981 m above ground level.
- Overlapping data were split into 724 non-overlapping 1,500 $\times$ 1,500 m tiles, which we obtained from the National Park Service.
- 724 LiDAR tiles (approx. size 98 GB) projected onto a 3.0 m resolution digital elevation model (DEM) derived from the LiDAR point cloud.
- Projection: UTM Units: meters

North Carolina

- LiDAR data for North Carolina was collected by NC Floodplain Mapping Program in 2005.
- Overlapping data were split into non-overlapping 10,000 $\times$ 10,000 ft tiles, which we obtained from the NC Floodplain Mapping Program.
- 184 LiDAR tiles (approx. size 8.9 GB) projected onto a 3.0 m resolution digital elevation model (DEM) derived from the LiDAR point cloud.
- Projection: NC State Plane Units: ft

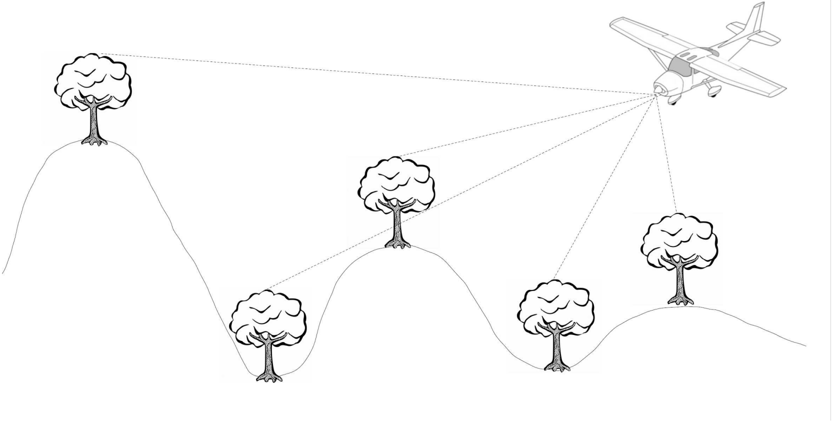
LiDAR Data for GSMNP



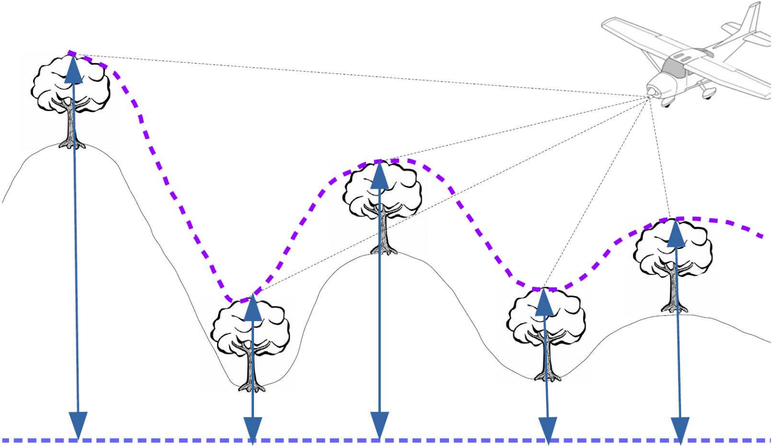
Computational Workflow and Data Processing

- We employed a process-parallel approach to extract and analyze LiDAR point cloud data using python.
- To estimate vegetation heights above ground level, elevations from the 3.0 m DEM were subtracted from point cloud data.
- The resulting points were grouped into 1 m vertical bins, up to 75 m, at a horizontal resolution of 30 $\times$ 30 m.
- Anomalous high points (aerosols, birds) and low points (steep slopes, surface litter) were filtered out.
- Corrections were made for low height vegetation (shrubs and grasses) and for many returns at the same elevation.

LiDAR 101

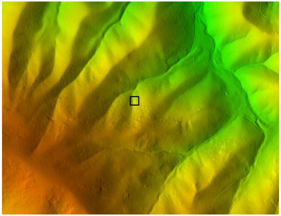


LiDAR 101

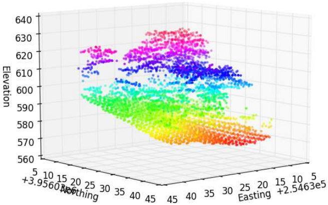


LiDAR 101

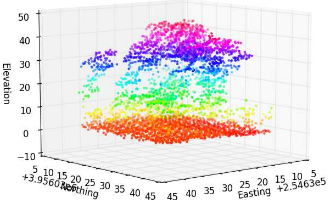
LiDAR Point Cloud Example: 30m pixel



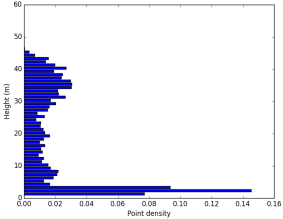
a) 3-D LiDAR point cloud extent at 30×30 m (black square) shown in a typical GSMNP cove forest.



b) Raw LiDAR point cloud (3,985 points), showing imprints of underlying topography.

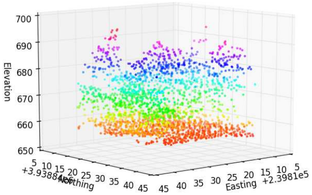


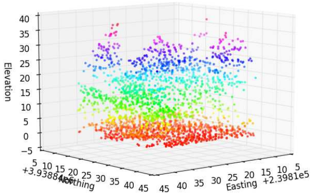
c) LiDAR point cloud after topographic detrending and filtering (3,936 points).

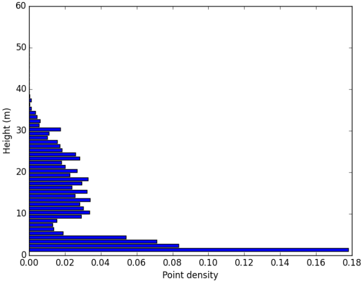


d) Vertical distribution of LiDAR point density in a cove forest dominated by tall trees and a dense understory.

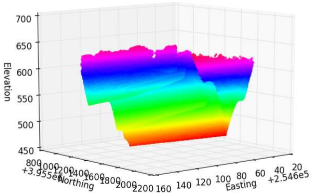
LiDAR Point Cloud Example: 30m pixel

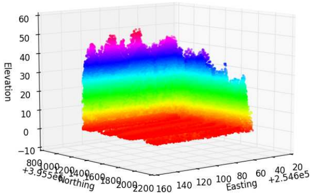


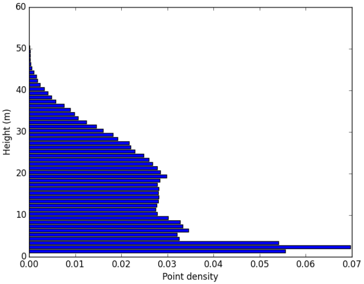




LiDAR Point Cloud Example: 100m pixel





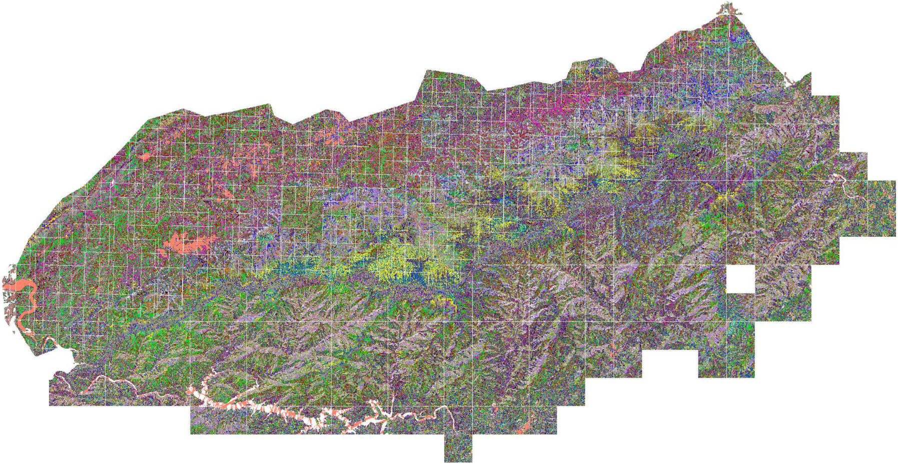


Computational Workflow and Data Processing

- We employed a process-parallel approach to extract and analyze LiDAR point cloud data using python.
- To estimate vegetation heights above ground level, elevations from the 3.0 m DEM were subtracted from point cloud data.
- The resulting points were grouped into 1 m vertical bins, up to 75 m, at a horizontal resolution of 30 × 30 m.
- Anomalous high points (aerosols, birds) and low points (steep slopes, surface litter) were filtered out.
- Corrections were made for low height vegetation (shrubs and grasses) and for many returns at the same elevation.

Map of 30 Vegetation Canopy Structure Classes

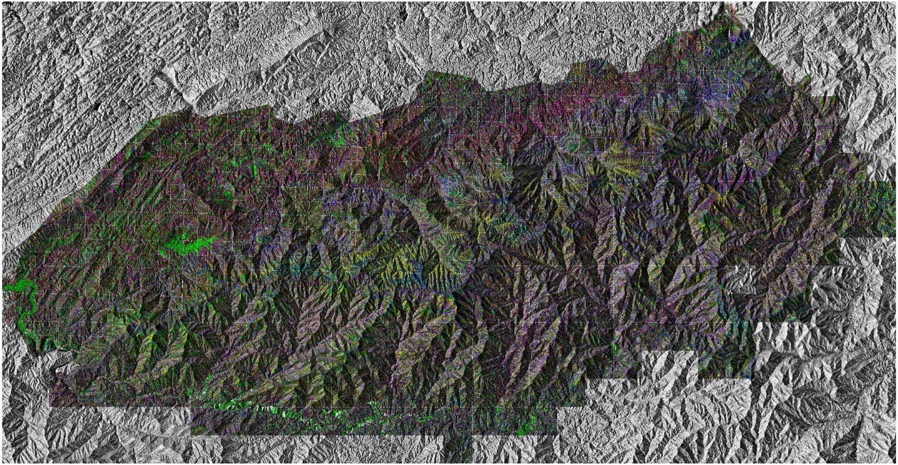
Vertical profile distributions were input to the cluster analysis, considering all tiles simultaneously.



This map shows the 30 most-different classes of vegetation canopy structure, randomly colored, as identified by *k*-means clustering for the Great Smoky Mountains National Park.

Map of 30 Vegetation Canopy Structure Classes

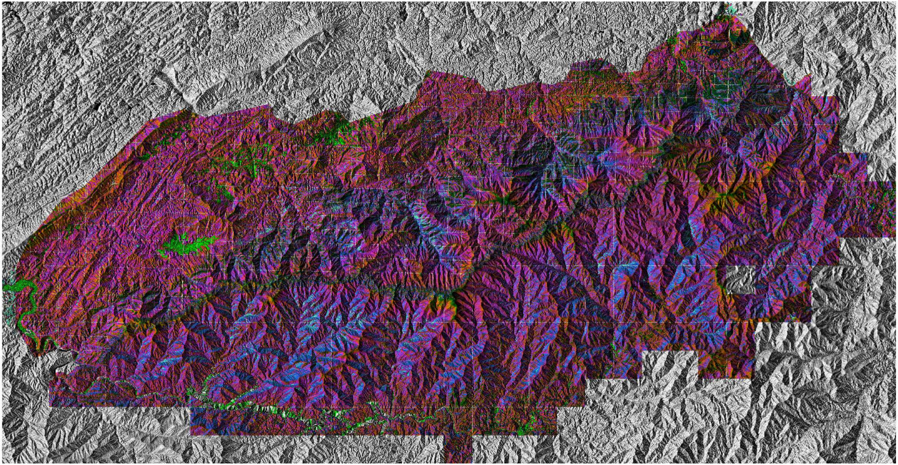
Vertical profile distributions were input to the cluster analysis, considering all tiles simultaneously.



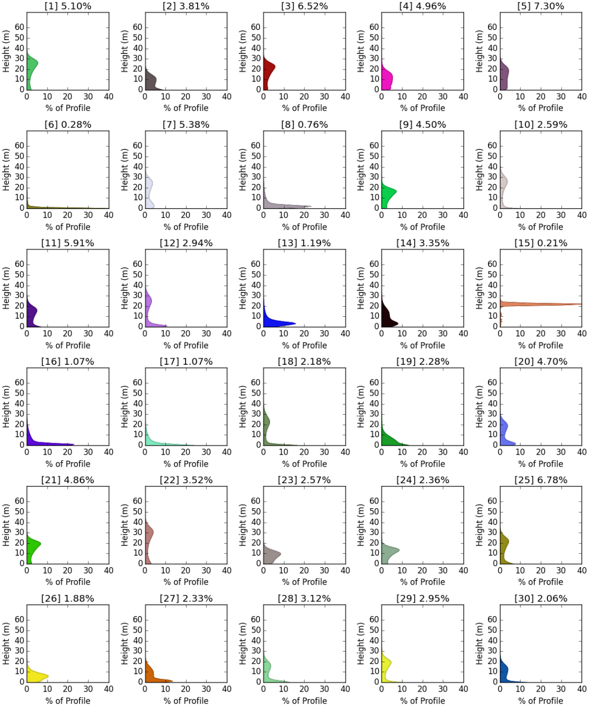
This map shows the 30 most-different classes of vegetation canopy structure, randomly colored, as identified by k-means clustering for the Great Smoky Mountains National Park.

Map of 30 Vegetation Canopy Structure Classes

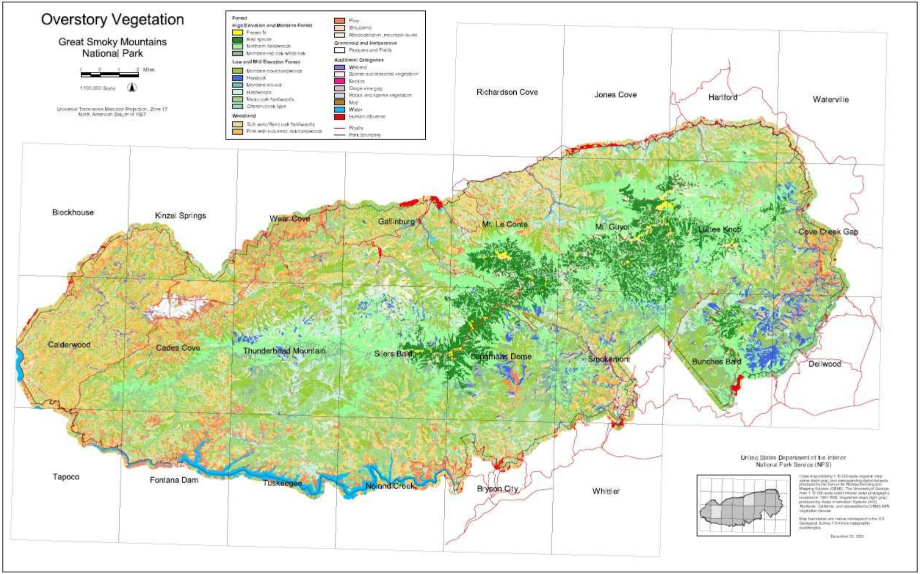
Vertical profile distributions were input to the cluster analysis, considering all tiles simultaneously.



This map shows the 30 most-different classes of vegetation canopy structure, randomly colored, as identified by *k*-means clustering for the Great Smoky Mountains National Park.


[1] 5.10%
[2] 3.81%
[3] 6.52%
[4] 4.96%
[5] 7.30%
[6] 0.28%
[7] 5.38%
[8] 0.76%
[9] 4.50%
[10] 2.59%
[11] 5.91%
[12] 2.94%
[13] 1.19%
[14] 3.35%
[15] 0.21%
[16] 1.07%
[17] 1.07%
[18] 2.18%
[19] 2.28%
[20] 4.70%
[21] 4.86%
[22] 3.52%
[23] 2.57%
[24] 2.36%
[25] 6.78%
[26] 1.88%
[27] 2.33%
[28] 3.12%
[29] 2.95%
[30] 2.06%
Height (m)
% of Profile

Overstory Vegetation Cover Map for GSMNP

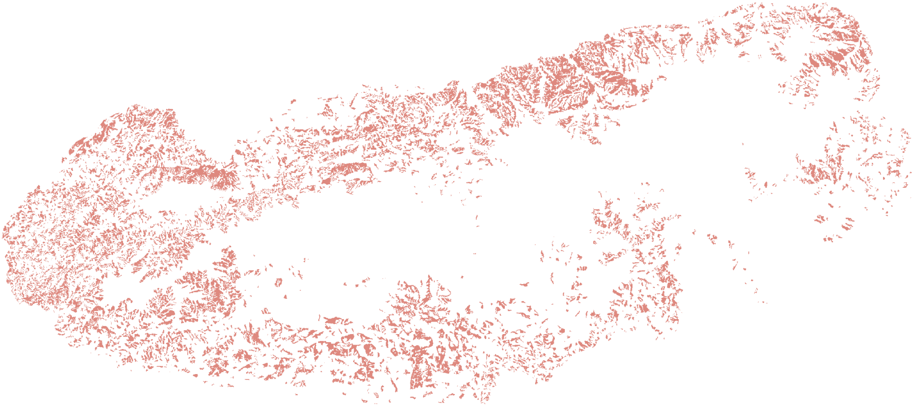


Overstory Vegetation Cover Map for GSMNP

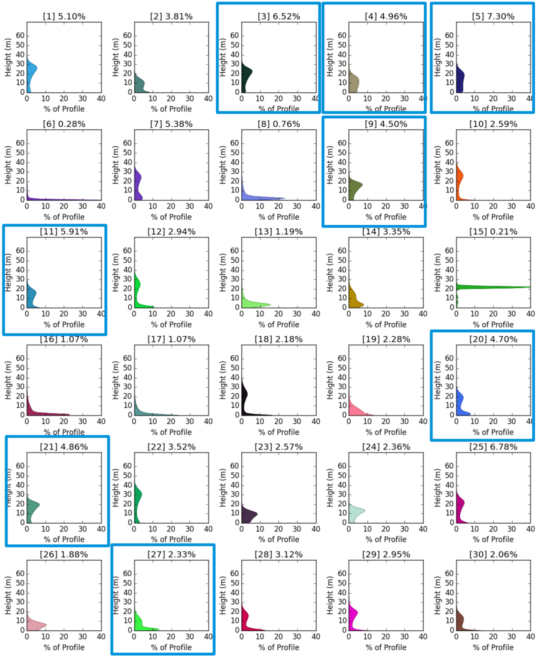
ID	Forest type	Percent Area
1001	Yellow pine forests	7.03
1003	Floodplain forests	1.21
1004	Ericaceous shrubs (non-heath bald type)	0.47
1005	Ericaceous shrubs (heath bald type)	1.00
1006	Successional hardwood forests	3.65
1007	Chestnut oak forests	14.08
1008	High elevation beech/red oak forests	1.53
1009	High elevation red oak/white oak forests	2.52
1010	Northern hardwood/acid hardwood forests	16.13
1011	Northern hardwood/boulderfield forests	4.09
1012	White pine forests	1.23
1013	Spruce-fir forests	2.57
1014	Roads	0.23
1015	Successional or modified vegetation	0.04
1016	Sparse vegetation	0.04
1017	Rock	0.10
1018	Mud/gravel	0.22
1019	Water	1.33
1020	Dead vegetation	0.06
1021	Exotic vegetation	0.00

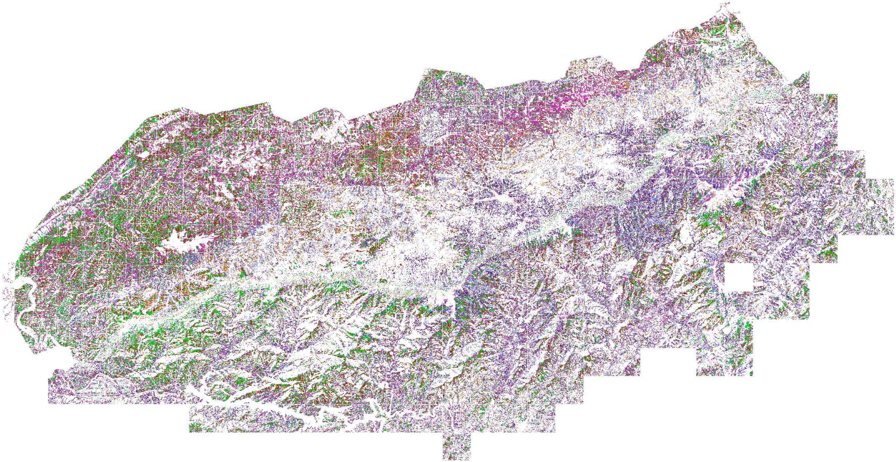
ID	Forest Type	Percent Area
1022	Human influence	0.68
1023	Shrubs	0.35
1024	Alluvial vegetation	0.00
3890	Successional or modified vegetation	0.21
3893	Successional or modified vegetation	0.03
4048	Successional or modified vegetation	0.51
4242	Grassy balds	0.01
6192	Montane oak-hickory forests	11.17
6272	Spruce-fir forests	0.68
6286	Montane oak-hickory forests	0.99
7102	Hemlock forests	0.36
7136	Hemlock forests	1.50
7230	Montane oak-hickory forests	7.93
7517	White pine forests	0.10
7519	White pine forests	0.76
7543	Montane cove forests	3.87
7692	Montane oak-hickory forests	1.16
7695	Montane cove forests	2.47
7710	Montane cove forests	9.04
7878	Montane cove forests	0.64

GSMNP: Chestnut Oak Forest

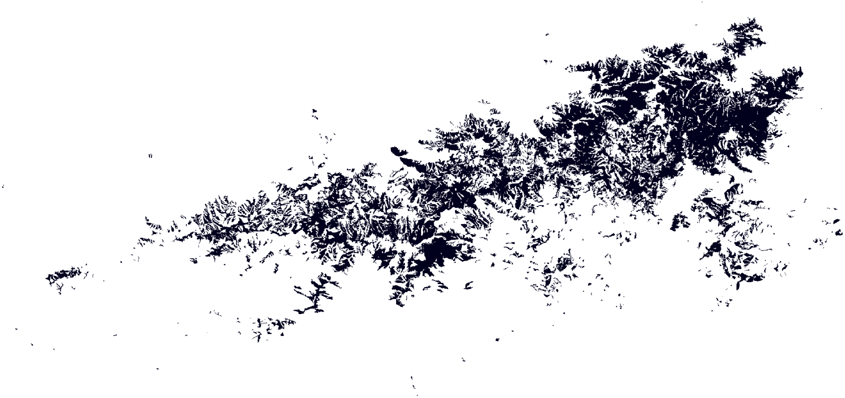


Chestnut Oak Forests

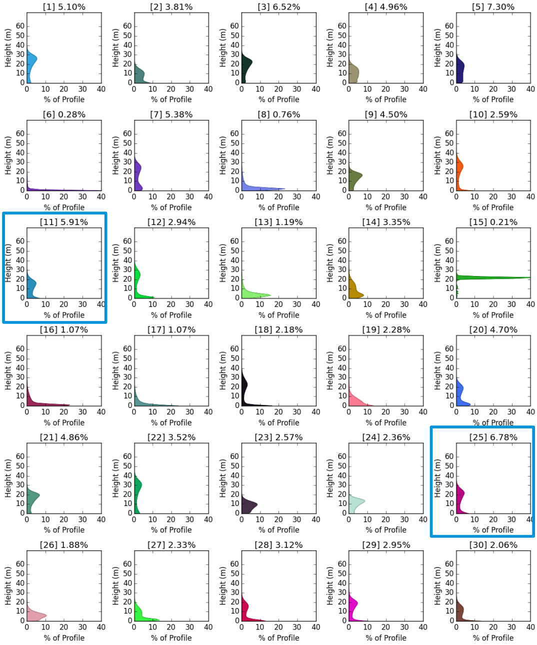


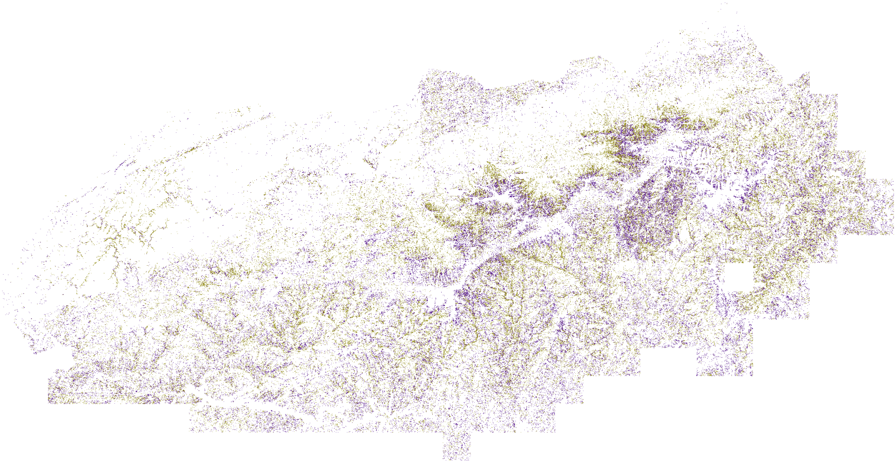


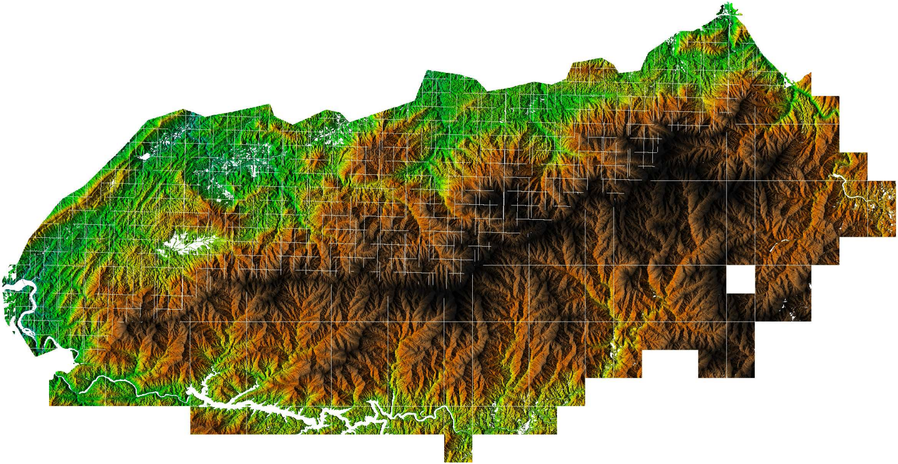
GSMNP: Northern Hardwood Forest



Northern Hardwood Forests

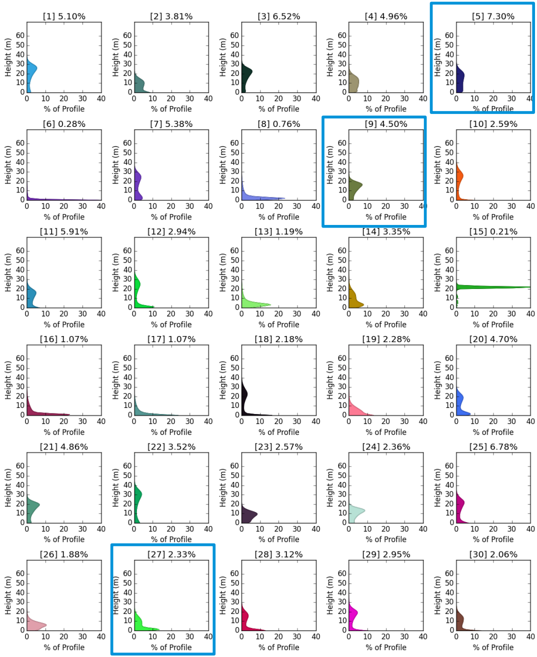


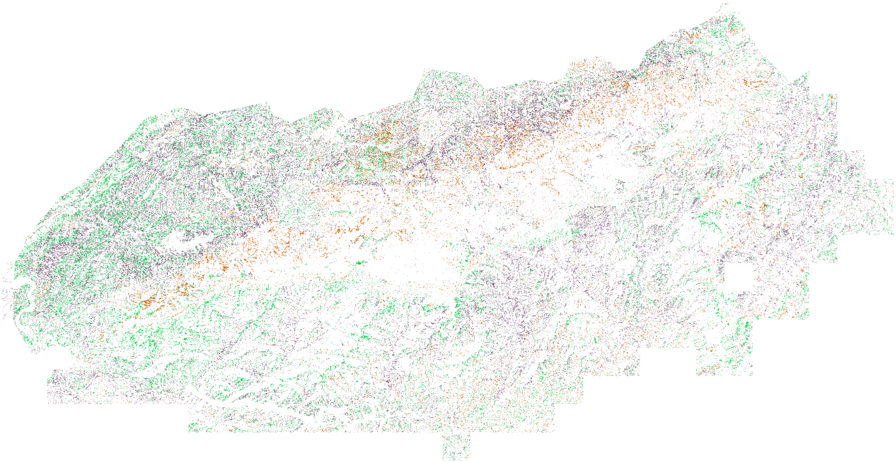




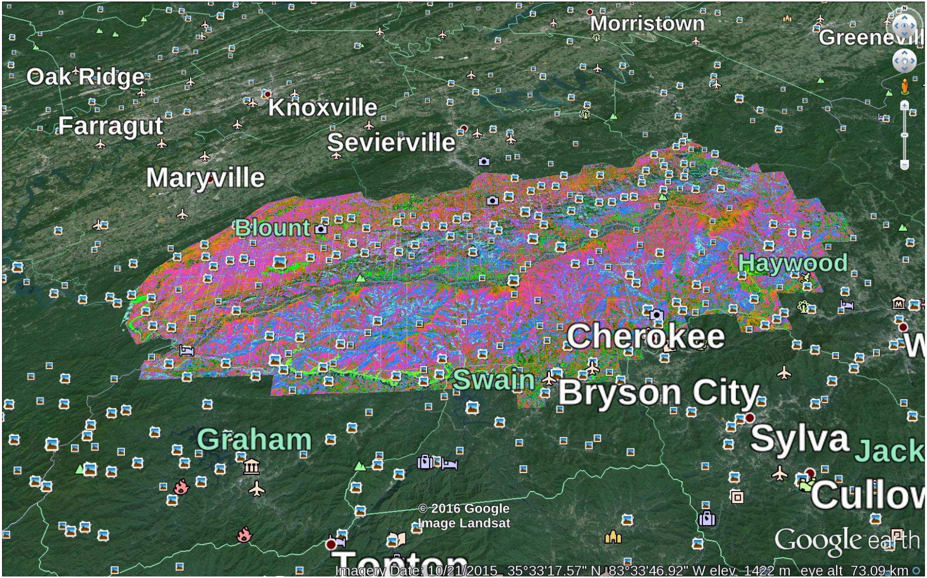
GSMNP: Hemlock Forest

Hemlock Forests

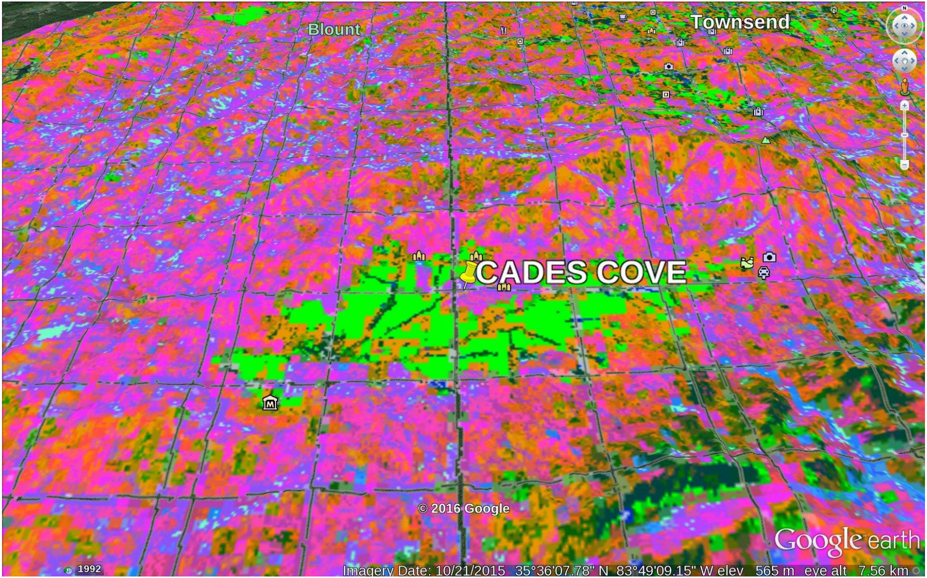


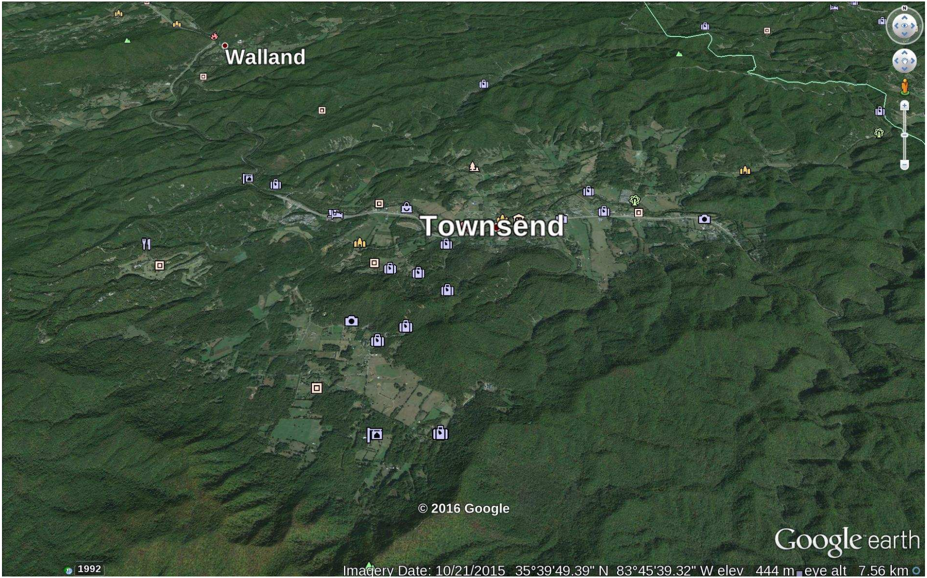


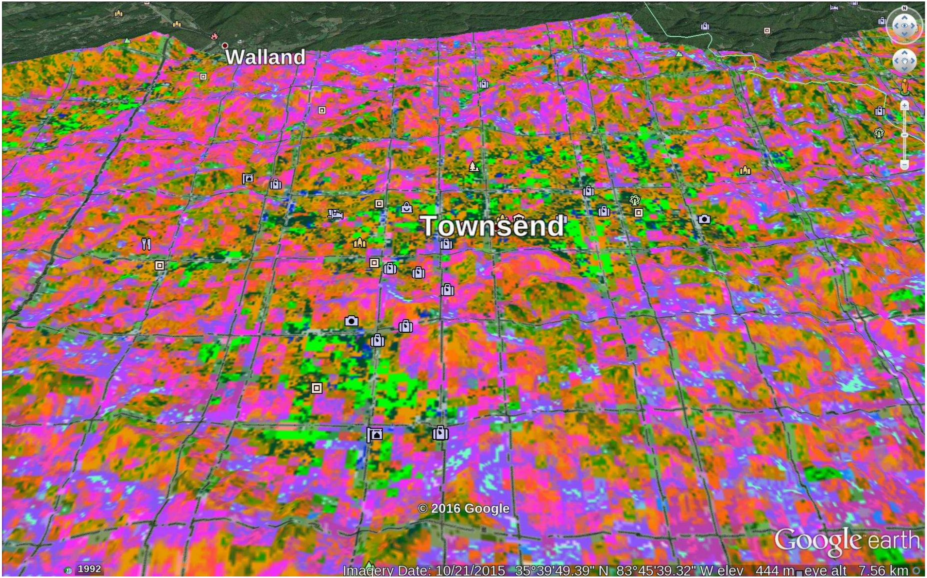
Exploring in Google Earth




Blount
Townsend
CADES COVE
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°36'07.78" N 83°49'09.15" W elev 565 m eye alt 7.56 km


Blount
Townsend
CADES COVE
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°36'07.78" N 83°49'09.15" W elev 565 m eye alt 7.56 km

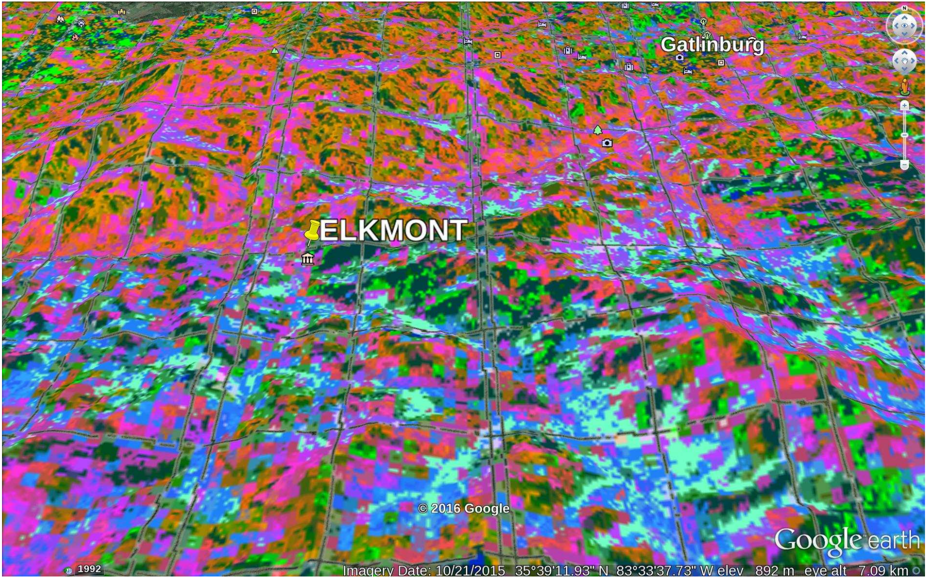

Walland
Townsend
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°39'49.39" N 83°45'39.32" W elev 444 m eye alt 7.56 km

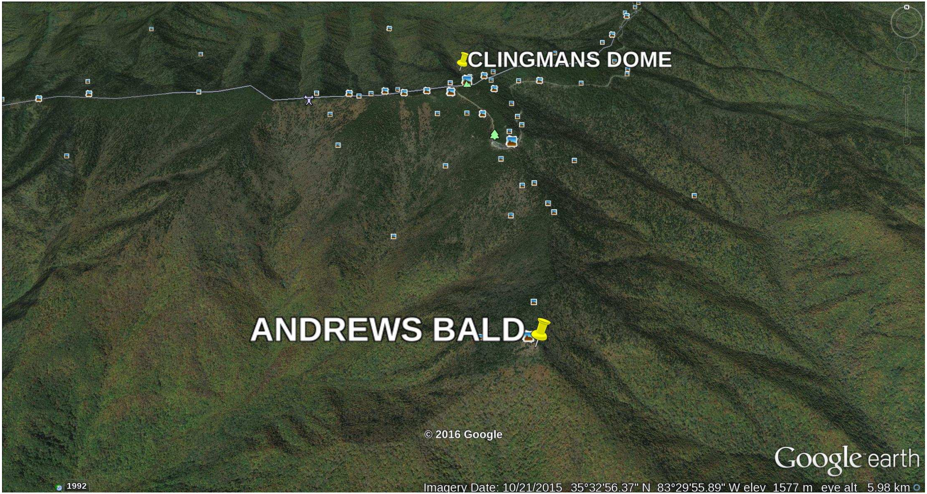

Walland
Townsend
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°39'49.39" N 83°45'39.32" W elev 444 m eye alt 7.56 km

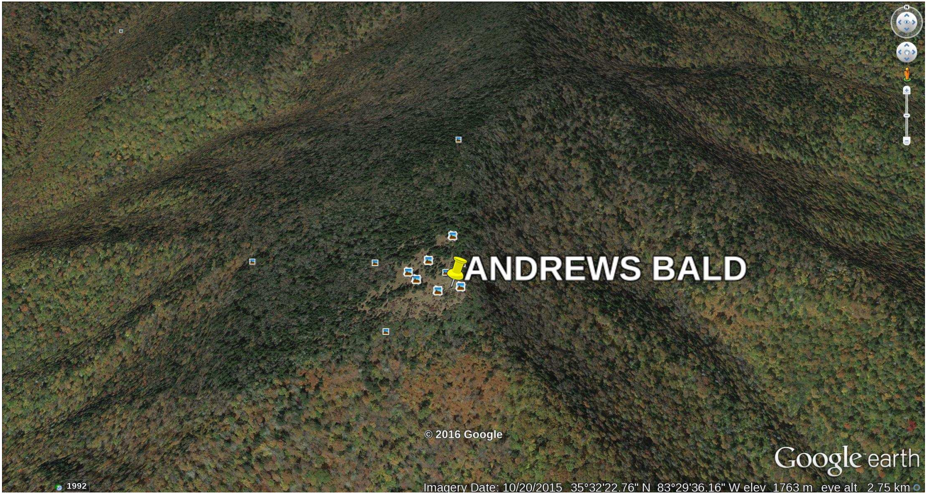

CLINGMANS DOME
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°33'38.97" N 83°30'25.53" W elev 1962 m eye alt 13.89 km

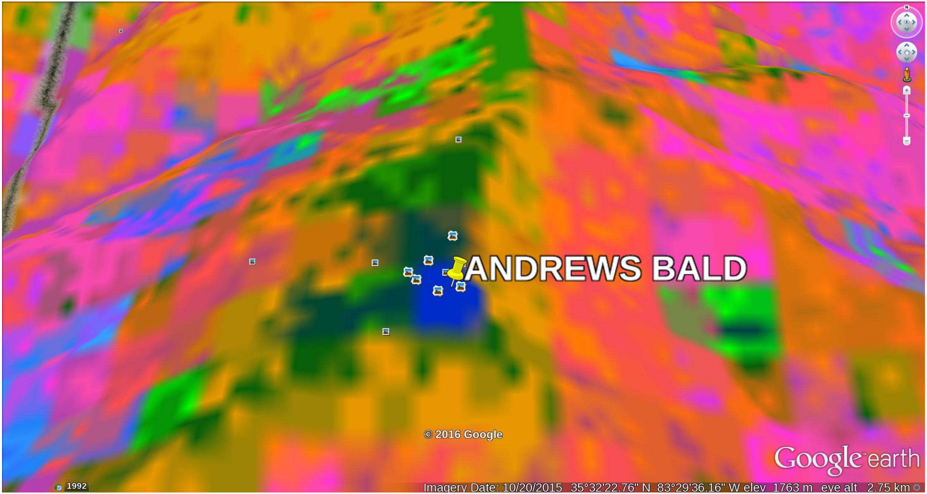

CLINGMANS DOME
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°33'38.97" N 83°30'25.53" W elev 1962 m eye alt 13.89 km


Gatlinburg
ELKMONT
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°39'11.93" N 83°33'37.73" W elev 892 m eye alt 7.09 km


Gatlinburg
ELKMONT
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°39'11.93" N 83°33'37.73" W elev 892 m eye alt 7.09 km

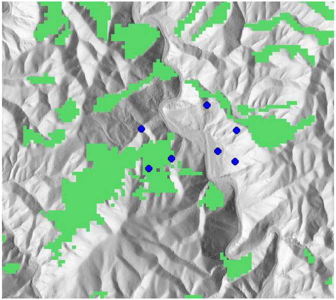

CLINGMANS DOME
ANDREWS BALD
© 2016 Google
Google earth
1992
Imagery Date: 10/21/2015 35°32'56.37" N 83°29'55.89" W elev 1577 m eye alt 5.98 km


ANDREWS BALD
© 2016 Google
Google earth
1992
Imagery Date: 10/20/2015 35°32'22.76" N 83°29'36.16" W elev 1763 m eye alt 2.75 km

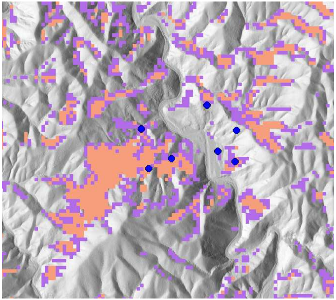

ANDREWS BALD
© 2016 Google
Google earth
1992
Imagery Date: 10/20/2015 35°32'22.76" N 83°29'36.16" W elev 1763 m eye alt 2.75 km

Great Smoky Mountain Institute at Tremont

- The Great Smoky Mountain Institute at Tremont (GSMIT) is surrounded by "Mountain Cove" and "Hemlock" forests with tall, dense canopies and low understory vegetation.
- We found strong spatial correspondence between the "Mountain Cove" forest and the vegetation structure classes representing the taller vegetation in the Park.
- Individual tree records from "Citizen Science" phenology plots, located at blue circles, were studied for ground truthing.



"Montain Cove" forest from (Madden, 2014).



Tall canopy vegetation classes 10 and 13 derived from LiDAR.

Summary and Conclusions

- We developed an approach, parallel software tools, and workflow for analyzing large volumes of LiDAR point cloud data in a scalable fashion.
- **Multivariate Spatiotemporal Clustering (MSTC)** provides a valuable quantitative framework for stratifying vegetation canopy structure data derived from LiDAR point clouds.
- We applied these tools to LiDAR data from the GSMNP to identify vegetation classes based on overstory/understory distributions.
- We used a spatial overlay method to compare the unsupervised clustering results to existing vegetation maps and to validate results
- These tools and the resulting maps will inform resource management and conservation planning by forest and wildlife managers, who were not previously able to use large, complex LiDAR data sets.

2015 IEEE 15th International Conference on Data Mining Workshops

Characterization and classification of vegetation canopy structure and distribution within the Great Smoky Mountains National Park using LiDAR

Jitendra Kumar*, Jon Weiner†, William W. Hargrove‡, Steven P. Norman‡,
Forrest M. Hoffman* and Doug Newcomb§
*Oak Ridge National Laboratory, Oak Ridge, TN, USA Email: jkumar@climatemodeling.org
†University of California Berkeley, Berkeley, CA, USA
‡USDA Forest Service, Southern Research Station, Asheville, NC, USA
§U.S. Fish and Wildlife Service, Raleigh, NC, USA

Abstract—**Vegetation canopy structure is a critically important habitat characteristic for many threatened and endangered birds and other animal species, and it is key information needed by forest and wildlife managers for monitoring and managing forest resources, conservation planning and fostering biodiversity. Advances in Light Detection and Ranging (LiDAR) technologies have enabled remote sensing-based studies of vegetation canopies by capturing three-dimensional structures, yielding information not available in two-dimensional images of the landscape provided by traditional multi-spectral remote sensing platforms. However, the large volume data sets produced by airborne LiDAR instruments pose a significant computational challenge, requiring algorithms to identify and analyze patterns of interest buried within LiDAR point clouds in a computationally efficient manner, utilizing state-of-art computing infrastructure. We developed and applied a computationally efficient approach to analyze a large volume of LiDAR data and characterized the vegetation canopy structures for 139,859 hectares (540 sq. miles) in the Great Smoky Mountains National Park. This study helps improve our understanding of the distribution of vegetation and animal habitats in this extremely diverse ecosystem.**

I. Introduction

Forest ecosystems are a complex mosaic of diverse plant and tree species, the location and distribution of which are driven by a number of gradients like climate (ex. temperature, precipitation regimes), topography (ex. elevation, slope, aspect), geology (ex. soil types, textures, depth), hydrology (ex. drainage, moisture availability) etc. Diverse combinations of these gradients support diverse composition and distribution of vegetation which in turn supports an array of wildlife. Understanding the vegetation canopy structure is critical to understand, monitor and manage the complex forest ecosystems like those in the Great Smoky Mountain National Park (GSMNP). Vegetation canopies not only help understand the vegetation, but are also a critically important habitat characteristics of many threatened and endangered animal and bird species for which the GSMNP is home.

Remote sensing has been widely used to monitor regional to global forest ecosystems and for mapping of vegetation types. However, traditional remote sensing methods for vegetation classification often use light reflectance from the top layer of vegetation. Advances in Light Detection and Ranging (LiDAR) technologies have enabled remote sensing-based studies of vegetation canopies by providing a three-dimensional representation of vegetation structure throughout the canopy. While the application of LiDAR for study of forest ecosystems is becoming more common, the richness of these data sets are generally under-utilized due to the large volumes of the data produced by these instruments and lack of computational resources and analysis algorithms. Most of the LiDAR studies focus on the development of high resolution Digital Elevation Models, canopy heights and occasionally understory density [1], [2]. While LiDAR dervied metrics have proven to be useful for an array of applications [1]–[5], three-dimensional information provided by the LiDAR are left unutilized.

The objective of this study is to develop methods to realize the potentials of rich LiDAR data set to map and characterize the three-dimensional structure and distribution of vegetation canopies. We develop and apply data analytic techniques to identify the ecologically important and understandable structural types by mining the large and complex volumes of LiDAR data.

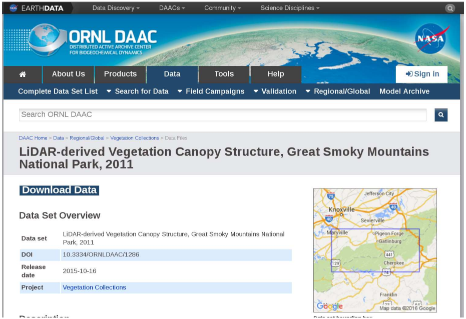
II. Materials

A. Study area

The geographic area for this study was the Great Smoky Mountains National Park (GSMNP), which in part covers the Great Smoky Mountains and the Blue Ridge Mountains, encompassing 816 sq. miles across Tennessee and North Carolina in the United States. Results presented here focus primarily on the Tennessee side of the GSMNP (approximately 540 sq. miles). The GSMNP covers complex topography with elevations ranging from 876–6,643 feet above mean sea level. The GSMNP is ecologically rich and diverse, consisting of about 1,600 species of flowering plants, including 100 native tree species and over 100 native shrub species [6]. The distribution of vegetation in the park is strongly influenced by topography, moisture and other environmental gradients [7].

978-1-4673-8493-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDMW.2015.178

IEEE computer society



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. http://dx.doi.org/10.3334/ORNLDAAC/1286

Acknowledgments

This research was sponsored by the U. S. Department of Agriculture Forest Service Eastern Forest Environmental Threat Assessment Center (EFETAC) and by the Climate and Environmental Sciences Division (CESD) of the Biological and Environmental Research (BER) Program in the U. S. Department of Energy Office of Science. This research used resources of the Oak Ridge Leadership Computing Facility (OLCF) at Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle, LLC, for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725.

References

M. Madden. Overstory vegetation at Great Smoky Mountains National Park, Tennessee and North Carolina (reference code: 1047498), 2014. URL https://irma.nps.gov/App/Reference/Profile/1047498.