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]

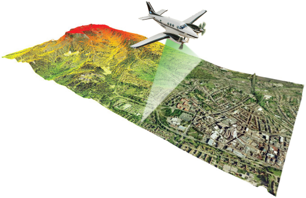
[1]Oak Ridge National Laboratory, [2]University of California Berkeley, [3]USDA Forest Service, [4]U.S. Fish and Wildlife Service

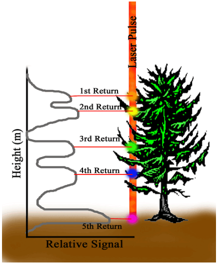
10th International Workshop on Spatial and Spatiotemporal Data Mining
Atlantic City, New Jersey, USA
November 14, 2015

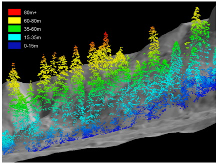
CLIMATE CHANGE SCIENCE INSTITUTE
OAK RIDGE NATIONAL LABORATORY

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.

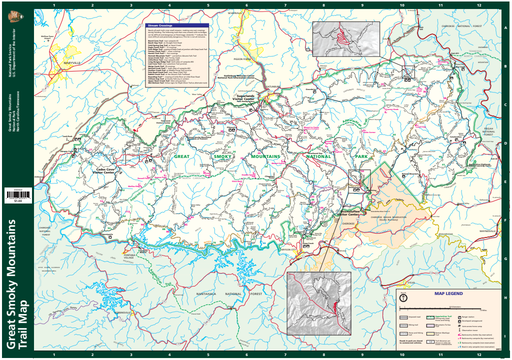


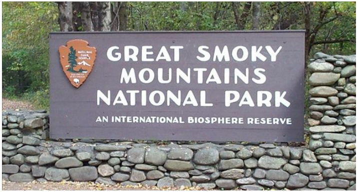




Great Smoky Mountains National Park (GSMNP)

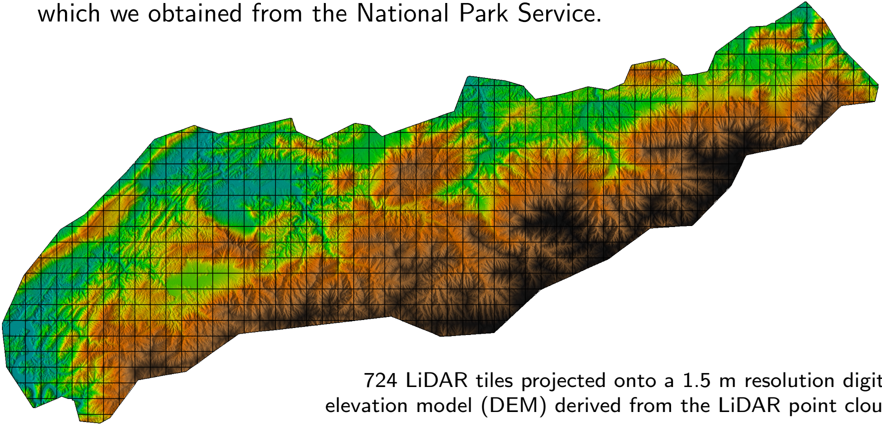
- The GSMNP is the most visited national park in the U.S., and it hosts a rich ecosystem of plants and wildlife.
- The Park encompasses 816 sq. miles in Tennessee and North Carolina and ranges in elevation from 876 to 6,643 feet above mean sea level.





LiDAR Tiles for Tennessee side of GSMNP

- 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.

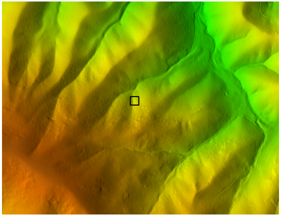


724 LiDAR tiles projected onto a 1.5 m resolution digital elevation model (DEM) derived from the LiDAR point cloud.

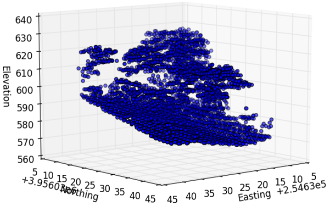
Computational Workflow and Data Processing

- 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 1.5 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.

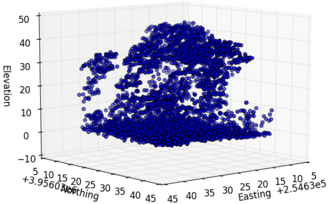
LiDAR Point Cloud Example



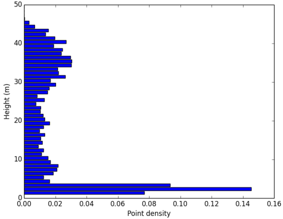
a) 3-D LiDAR point cloud extent at 30 × 30 m (black square) shown in a typical GSMNP cove forest.



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

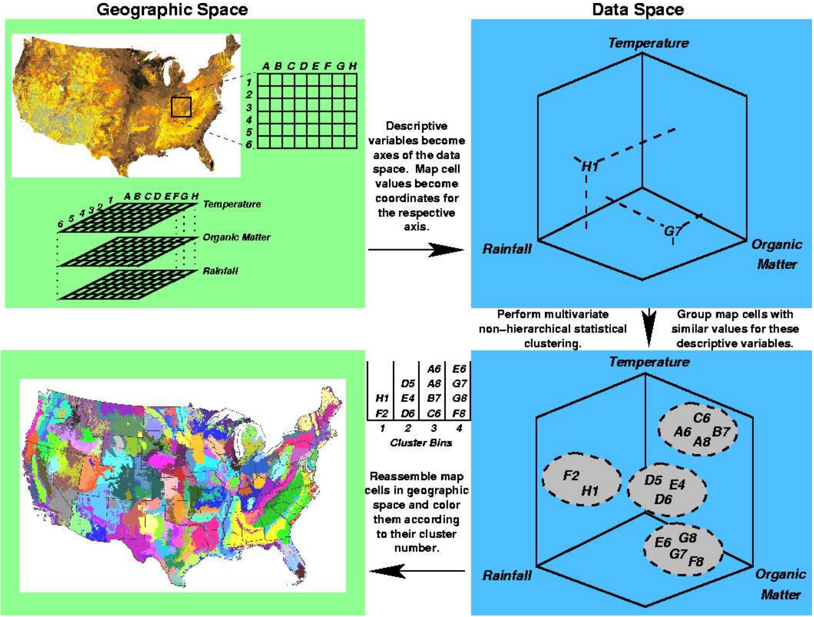


c) LiDAR point cloud after topographic detrending and filtering (3,936 points).



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

Multivariate Spatiotemporal Clustering (MSTC)



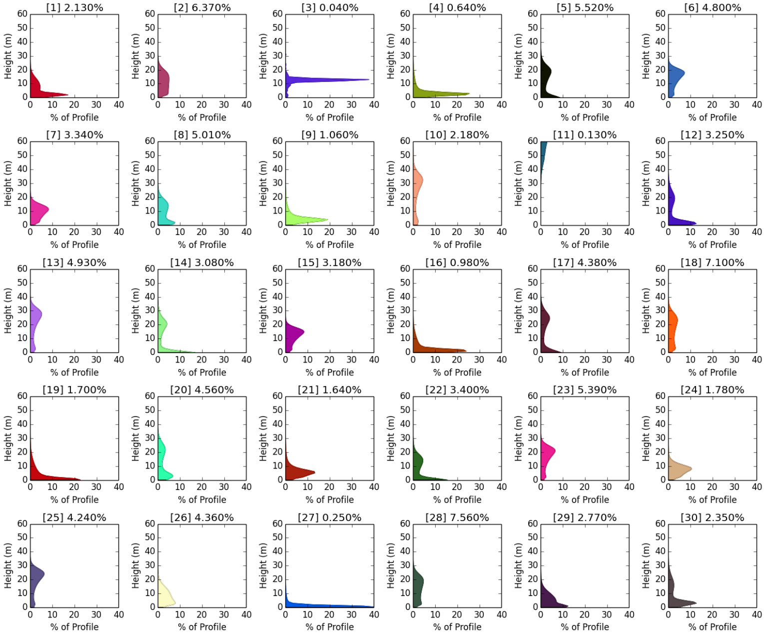
(Hargrove and Hoffman, 2004)

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 Tennessee portion of the Great Smoky Mountains National Park.

30 Vegetation Canopy Structure Prototypes

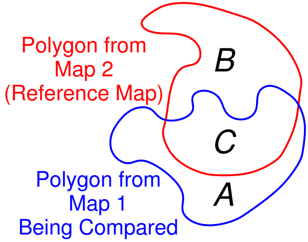


Automated Supervision for Unsupervised Classification

- Clustering is an unsupervised classification technique, so ecoregions have no descriptive labels (e.g., **chestnut oak forest** or **spruce-fir forest**).
- **Label stealing** allows us to perform automated "supervision" by "stealing" the best corresponding human-created descriptive labels to assign to ecoregions or vegetation structure classes.
- We employed a tool called **Mapcurves** to select the best vegetation class labels from vegetation type maps delineated by human experts for the GSMNP (Madden, 2014).
- We considered a library of vegetation type land cover maps at 30 m resolution, and chose the label with the highest **goodness-of-fit (GOF)** score for every ecoregion polygon.

Mapcurves: A Method for Comparing Categorical Maps

- Hargrove et al. (2006) developed a method for quantitatively comparing categorical maps that is
  - independent of differences in resolution,
  - independent of the number of categories in maps, and
  - independent of the directionality of comparison.

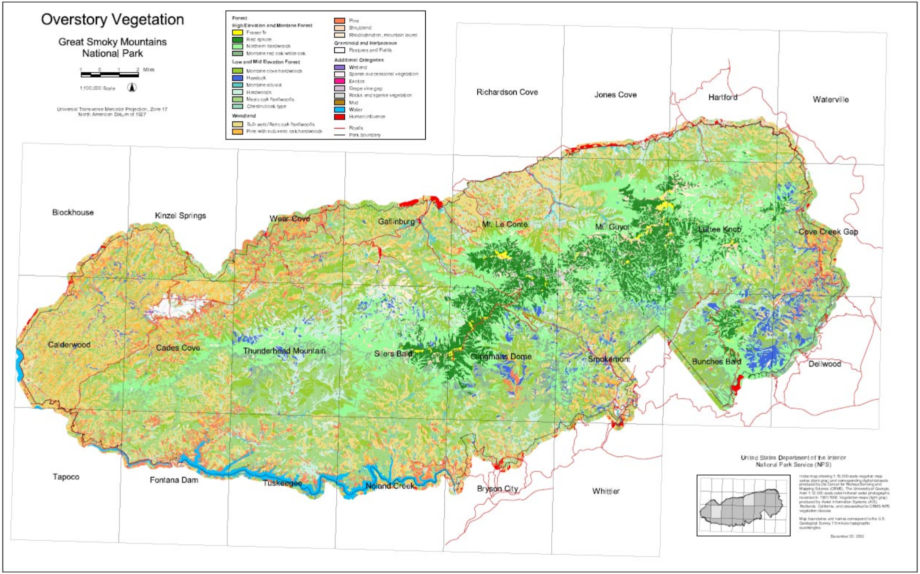


Goodness of Fit (GOF) is a unitless measure of spatial overlap between map categories:

$$\text{GOF} = \sum_{\text{polygons}} \frac{C}{B + C} \times \frac{C}{A + C}$$

- GOF provides "credit" for the area of overlap, but also "debit" for the area of non-overlap.
- Mapcurves comparisons allow us to reclassify any map in terms of any other map (*i.e.*, color Map 2 like Map 1).
- A greyscale GOF map shows the degree of correspondence between two maps based on the highest GOF score.

Overstory Vegetation Cover Map for GSMNP



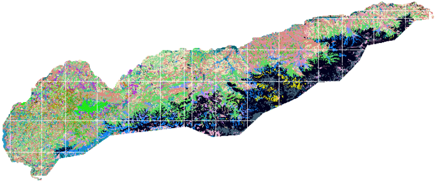
Translating Cluster Numbers to Vegetation Types

	Dominant Vegetation Type
0	Successional or modified vegetation
1	Chestnut Oak Forest
2	Chestnut Oak Forest
~~3~~	~~Successional or Modified Vegetation~~
4	Chestnut Oak Forest
5	Northern Hardwood/acid Hardwood Forest
6	Chestnut Oak Forest
7	Yellow Pine Forest
8	Northern Hardwood/acid Hardwood Forest
9	Chestnut Oak Forest
10	Montane Cove Forest
~~11~~	~~Chestnut Oak Forest~~
12	Northern Hardwood/acid Hardwood Forest
13	Montane Oak-hickory Forest
14	Northern Hardwood/acid Hardwood Forest
15	Yellow Pine Forest

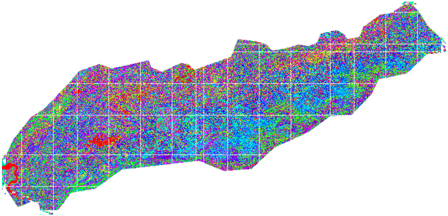
	Dominant Vegetation Type
16	Chestnut Oak Forest
17	Montane Cove Forest
18	Montane Oak-Hickory Forest
19	Chestnut Oak Forest
20	Montane Oak-Hickory Forest
21	Spruce-Fir Forest
22	Northern Hardwood/Acid Hardwood Forest
23	Chestnut Oak Forest
24	Yellow Pine Forest
25	Montane Oak-Hickory Forest
26	Chestnut Oak Forest
27	Ericaceous Shrubs (Heath Bald Type)
28	Chestnut Oak Forest
29	Yellow Pine Forest
30	Chestnut Oak Forest

- Clusters 3 and 11 are anomalous and represent very small areas.
- Labels are used to subsequently combine clusters or reclass regions.
- Labels and vertical distribution plots indicate vegetation cover, tree height and stand age, and suggest animal habitat occurrence.

Vegetation Type Label Stealing



Masked 30 m vegetation type map for Tennessee GSMNP



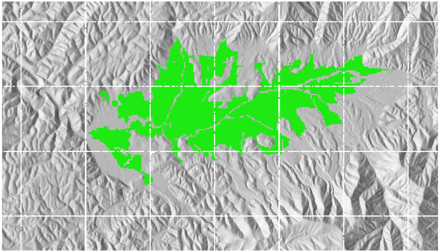
Masked $k = 30$ LiDAR classification

Reclassed $k = 30$ LiDAR classification

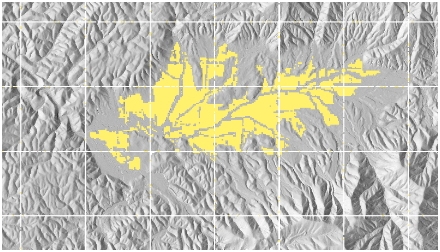
Goodness of Fit (GOF)

Validation: Cades Cove Valley

- Cades Cove, now a popular tourist destination, was a 19th century agrarian settlement.
- The area consists of woodlots interspersed within old fields that are mowed and burned to mimic the original settlement (Thiemann et al., 2009).
- Identified as "Successional or modified vegetation" in the vegetation type map, the area corresponds well with the low height (<1 m tall) vegetation class derived from LiDAR.



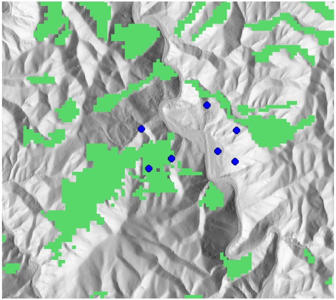
"Successional or modified vegetation" from (Madden, 2014).



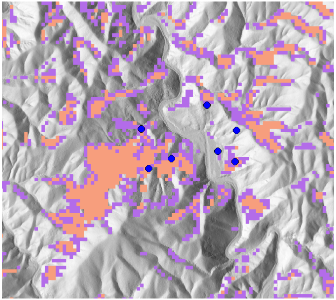
Low height (<1 m tall) vegetation class derived from LiDAR.

Validation: 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 10 and 13, which represent the tallest 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 extract descriptors for the unsupervised clustering results and to validate results in Cades Cove and GSMIT.
- **Label Stealing** offers a useful method for interpreting and understanding vegetation canopy structure delineations.
- 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.

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

W. W. Hargrove and F. M. Hoffman. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environ. Manage.*, 34(Supplement 1):S39–S60, Apr. 2004. doi:10.1007/s00267-003-1084-0.

W. W. Hargrove, F. M. Hoffman, and P. F. Hessburg. Mapcurves: A quantitative method for comparing categorical maps. *J. Geograph. Syst.*, 8(2):187–208, July 2006. doi:10.1007/s10109-006-0025-x.

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`.

J. A. Thiemann, C. R. Webster, M. A. Jenkins, P. M. Hurley, J. H. Rock, and P. S. White. Herbaceous-layer impoverishment in a post-agricultural southern Appalachian landscape. *The American Midland Naturalist*, 162:148–168, 2009. doi:10.1674/0003-0031-162.1.148.