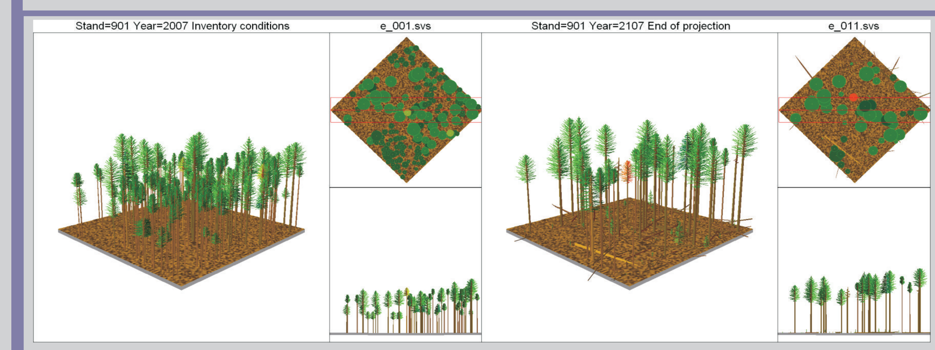
Predicting Tree-Level Forest Structure From LiDAR Data

Michael J. Falkowski[1], Paul E. Gessler[1], Andrew T. Hudak[2], and Nicholas L. Crookston[2]
[1]University of Idaho College of Natural Resources GLED, [2]USDA Forest Service Rocky Mountain Research Station

INTRODUCTION

Abstract

This research evaluates the efficacy of k-nearest neighbor (k-NN) imputation models incorporating LiDAR data to predict and map tree-level forest structure data (individual tree height, diameter at breast height, and species) across a 30,000 ha study area in Northern Idaho, USA. The primary objective is to provide spatially explicit data to parameterize the Forest Vegetation Simulator (FVS), a forest growth model that operates at the individual tree-level. Eventually forest growth will be modeled across the entire study area. In addition to FVS parameterization, the imputed forest structure data could be used for many purposes including forest commodity assessment, carbon accounting, wildlife habitat modeling, etc. The final k-NN imputation model utilizes LiDAR derived height measurements and LiDAR topographic variables to predict tree-level forest structure and species composition data. When compared to an independent forest inventory dataset, the imputed forest structure data had a species composition accuracy of 50%. The accuracy of forest structural attributes calculated from the imputed dataset were quite high when compared to the independent forest inventory data; the root mean square error of imputed basal area and quadratic mean diameter estimates were 5.28 m^2 / ha and 0.81cm, respectively. Furthermore, FVS growth projections based upon the imputed tree-level forest structure data follow similar trends as compared to FVS growth projections based upon the independent forest inventory data. These results indicate that the imputation methods presented herein could eventually be used to parameterize FVS across the entire study area, facilitating the modeling of forest dynamics across the entire region.



StandID	Year	Age	Tpa	BA	SDI	CCF	TopHt	QMD
901	2007	0	200	167	281	154	90	12.36732
901	2017	10	158	157	256	142	92	13.50231
901	2027	20	203	150	259	132	94	11.64478
901	2037	30	146	147	239	126	97	13.60051
901	2047	40	192	144	247	120	99	11.72584
901	2057	50	134	142	229	117	101	13.97679
901	2067	60	148	140	230	113	103	13.17049
901	2077	70	105	140	215	111	106	15.62386
901	2087	80	143	137	225	107	108	13.29856
901	2097	90	97	138	209	106	110	16.16558
901	2107	100	136	137	222	104	111	13.58681

Figure 1. Graphical and Tabular Examples of FVS Forest Growth Projections

The Forest Vegetation Simulator

FVS is a forest growth model used to aid in forest management decision-making (Crookston and Dixon, 2005; Figure 1).

It is an empirically driven model that operates at the individual tree-level.

Requires tree-level forest inventory data for parameterization (i.e., Measurements of diameter at breast height (DBH), species, height, etc. for each tree in a plot or stand).

Obtaining broad-scale, spatially continuous inventory data is difficult.

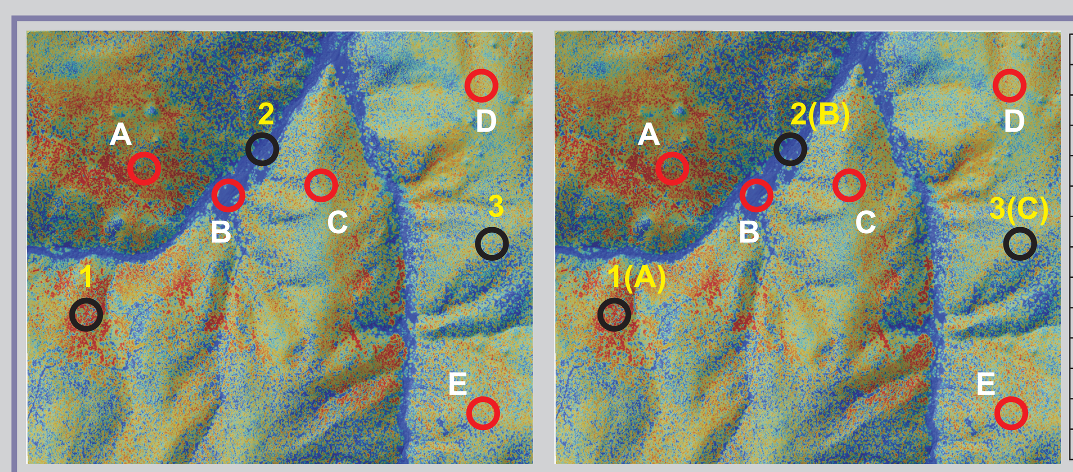
k-NN imputation models incorporating LiDAR data may provide a means to obtain spatially continuous predictions of tree level information at the landscape scale.

K-Nearest Neighbors Imputation

Multivariate statistical technique that uses nearest neighbors (in multivariate space) to predict missing data values (Figure 2).

Has been use extensively to predict stand-level forest inventory data (e.g., basal area, mean diameter, stand height, etc.) from remotely sensed data (Mäkelä and Pekkarinen, 2004; LeMay and Temesgen, 2005).

To date k-NN imputation has not been evaluated in the prediction of tree-level forest structure data.



Sample Plots	Avg LiDAR Ht	Aspect	Basal Area
A	35 m	110	350 Sq ft/ ac
B	1 m	-1	20 Sq ft/ ac
C	25 m	350	250 Sq ft/ ac
D	35 m	292	300 Sq ft/ ac
E	30 m	290	275 Sq ft/ ac
New Plots			
1	34 m	115	?
2	1.5 m	-1	?
3	22 m	353	?
Imputed Values			
1	34 m	115	350 Sq ft / ac
2	1.5 m	-1	20 Sq ft / ac
3	22 m	353	250 Sq ft / ac

Figure 2. Graphical Example of Imputation - Missing data in new plots are imputed from nearest sample plots (in multivariate space)

Objectives

Evaluate the efficacy of k-NN imputation for predicting tree-level forest structure data in un-inventoried areas.

The accuracy of k-NN predictions will be evaluated by comparing them to independent forest inventory data.

Parameterize FVS with the imputed tree-level forest structure data as well as with the independent forest inventory data.

Compare FVS forest growth projections derived from each data set.

METHODS

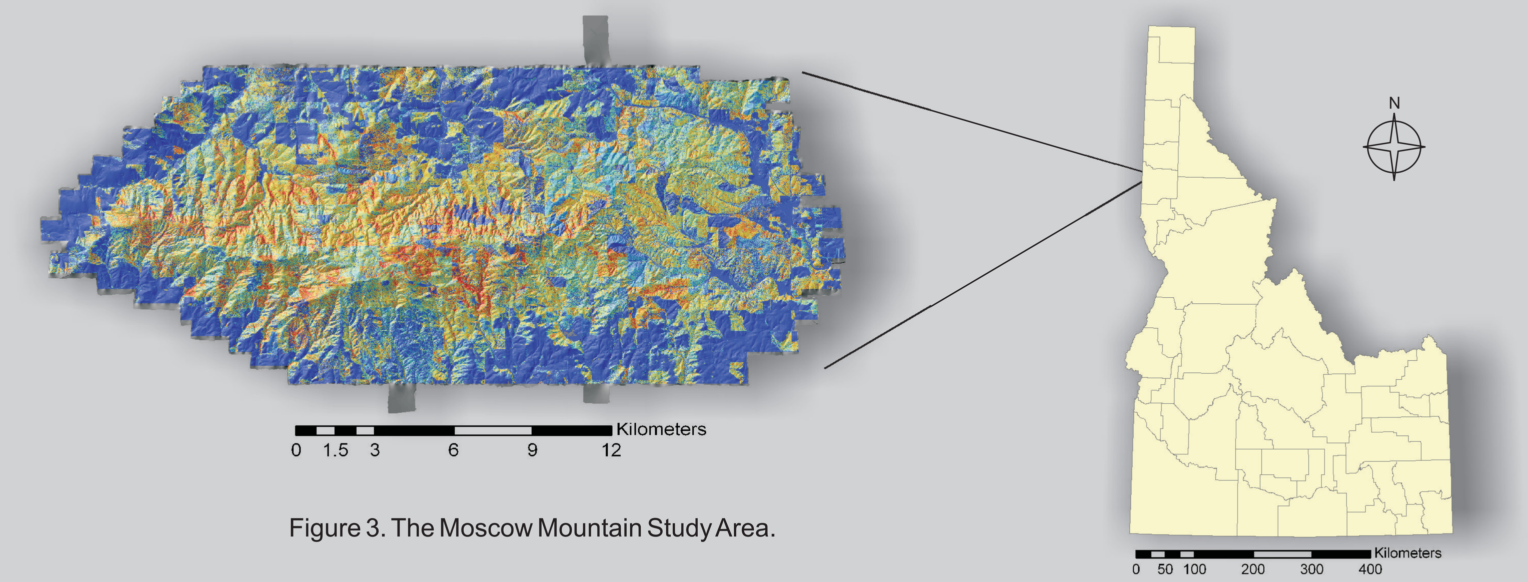


Figure 3. The Moscow Mountain Study Area.

Study area and Forest Inventory Data Collection

This study was conducted on Moscow Mountain, which lies at the extreme western extent of the Clearwater Mountains in Northern Idaho, USA (Figure 3).

Moscow Mountain is topographically complex and primarily comprised of temperate mixed-conifer forest.

Eighty- three 405 m^2 fixed-radius forest inventory plot were surveyed during the summer of 2003.

The University of Idaho Experimental forest collected independent forest inventory data during the summer of 2006, which will be used as validation data in the current study.

LiDAR Acquisition and Processing

Discrete return LiDAR data (1.95 m nominal post spacing) were acquired with an ALS40 system.

Once acquired, the LiDAR data were separated into ground and non-ground returns using the Multiscale Curvature Classification algorithm (Evans and Hudak, 2007).

Following classification, a digital elevation model was created, and the height above ground surface was calculated for all non-ground returns.

Numerous LiDAR derived height metrics and topographic variables were calculated across the Moscow Mountain study area (Table 1, Figure 3).

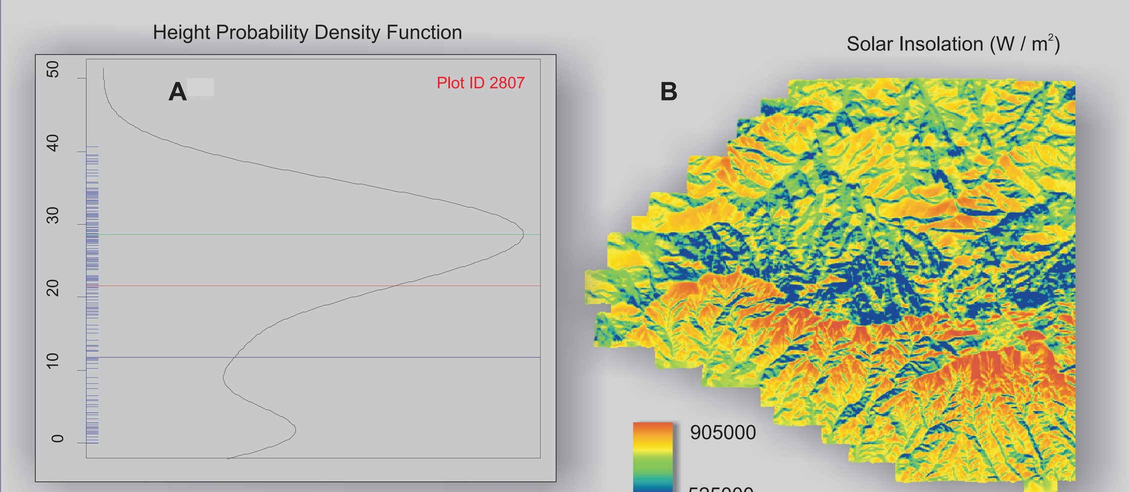


Table of LiDAR Metrics

C

Topographic Metrics	Height Metrics
Elevation	Mean Height
Slope Aspect	Minimum Height
Topographic Position	Maximum Height
Wetness Index	Standard Deviation of Heights
Distance to Streams	Modal Height
Solar insolation	Median Height
Heatload	Percentile Heights (10%-90%)

Figure 4. LiDAR height distribution and example height metrics (A), and example topographic variables (B and C). In figure 4A the black line is the probability density function of LiDAR heights at an inventory plot. The green, red, and purple lines represent the modal height, mean height, and the standard deviation of heights, respectively. The blue tic marks on the Y axis represent individual LiDAR returns. Figure 4 B is solar insolation (W / m^2) calculated from the LiDAR DEM. Figure 4 C is a table listing topographic and height metrics utilized in this study.

Imputation Model Development

A two-step process was employed to predict tree structural information at each of the validation inventory plots.

First, imputation models were developed to predict plot-level forest structure and species information (e.g., forest type and basal area) from plot-level LiDAR height metrics and LiDAR derived DEM variables.

Each Imputation model was evaluated based upon prediction accuracy and parsimony.

For the second step, we calculated multivariate distance between each of the original inventory plots and each of the validation plots.

Tree-level data from the original forest inventory were then assigned to the closest (in multivariate space) validation inventory plot (Figure 1).

Imputation Model Evaluation

The accuracy of the imputed tree-level forest structure data was determined by comparing it to tree-level forest structure data measured during the validation inventory.

FVS was parameterized with both the imputed and validation inventory data. Forest growth was then projected in 10-year increments for 100 years via FVS.

To further evaluate the accuracy of the imputed tree-level forest structure data, the growth projections for six randomly selected stands were compared.

Any multivariate distance could be used. However, for this study multivariate distance was based upon the randomForest proximity matrix (see Breiman, 2001 for a description of the randomForest method, and Hudak et al., Accepted for a description of the randomForest based multivariate distance).

RESULTS

Imputation Model Results

The final imputation model predicted forest type and basal area from three LiDAR derived topographic variables (heat load, elevation, and slope) and four LiDAR derived height metrics (mean height, median height, height of the 10th percentile, and height of the 75th percentile).

The modeled estimate of species composition prediction accuracy was only 37%, while the modeled estimate of basal area error (root mean square difference) was 13.04 m^2 / ha (56.83 ft^2 / ac).

When compared to the independent forest inventory data, the error rates for species composition and forest structure were much lower (Table 1).

Inventory Metric	RMSE	Correlation
Basal Area (Ft^2 / ac)	29.83	0.92
Trees Per Acre	18.95	0.86
Top Height (Ft)	8.65	0.81
Quadric Mean Diameter (In)	0.32	0.99
Total Cubic Feet (Ft^3)	779.1	0.93
Species Composition Error	50%	

Table 1. Error Statistics for the Imputed Forest Inventory Data.

Model Evaluation and FVS Parameterization

FVS growth projections based upon the imputed data followed similar trends as compared to FVS growth projections based upon the validation inventory data.

This is regardless of species composition prediction accuracy. In fact, the species Composition of stands 901 and 915 were correctly predicted by the imputation model.

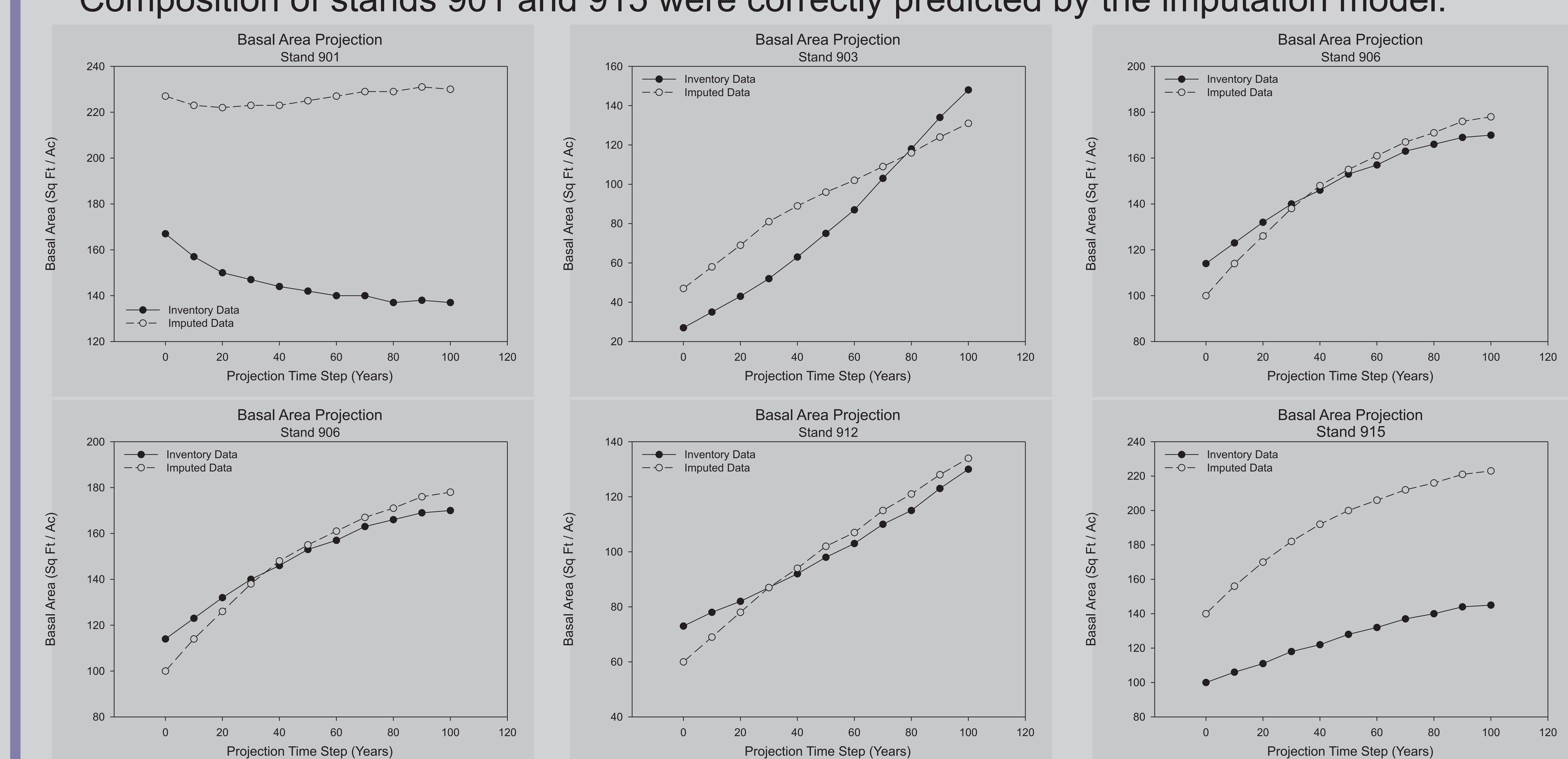


Figure 5. FVS growth projectiosn for the imputed and validation inventory data.

CONCLUSIONS

This study evaluated the efficacy of predicting tree-level forest structure data via k-NN imputation models incorporating LiDAR data. The imputed data were ultimately used to parameterize a tree-level forest growth model. When compared to independent forest inventory data, the imputation had an accuracy of 50% for predicting forest species composition. The accuracy of forest structure metrics (e.g, basal area and quadratic mean diameter were quite high). Furthermore, FVS growth projections based upon the imputed tree-level inventory data followed similar trends as compared to FVS growth projection based upon independent forest inventory data. This finding indicates that the methods presented herein could eventually be use to predict tree-level forest structure data across the entire study area, which could ultimately be use to project forest growth across the region. In addition to projecting forest growth, the imputed tree-level forest structure data could be used for a variety of applications including forest commodity assessment, carbon accounting, wildlife habitat assessment, etc. However, before this is accomplished we plan to further assess the accuracy of the imputation model presented herein.

Literature Cited

Breiman, L., 2001. Random Forests, Machine Learning, 45, 532.
Crookston, N.L., and Dixon G.E., 2005. The forest vegetation simulator: A review of its structure, content and applications. Computer and Electronics in Agriculture. 49, 60-80.
Evans, J.S., and Hudak, A.T., 2007. A multiscale curvature algorithm for classifying discrete return lidar in forested environments. IEEE Transactions on Geoscience and Remote Sensing 45(4): 1029-1038.
Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., and Falkowski M.J., Accepted. Nearest-neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data., Remote Sensing Environment.
LeMay, V., and Temesgen, H., 2005. Comparision of Nearest Neighbor Methods for Estimating Basal Area and Stems per Hectare Using Auxiliary Variables. Forest Science, 51, 109-119.
Makela, H., and Pekkarinen, A., 2004. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. Forest Ecology and Management. 196, 245-255.