Title

A Vegetation Mapping Strategy for Conifer Forests by Combining Airborne LiDAR Data and Aerial Imagery

Journal

Canadian Journal of Remote Sensing, 42(1)

ISSN

0703-8992

Authors

Su, Yanjun
Guo, Qinghua
Fry, Danny L
et al.

Publication Date

2016-01-02

DOI

10.1080/07038992.2016.1131114

Peer reviewed

Canadian Journal of Remote Sensing, 42:1–15, 2016

ISSN: 0703-8992 print / 1712-7971 online
DOI: 10.1080/07038992.2016.1131114

Taylor & Francis
Taylor & Francis Group

A Vegetation Mapping Strategy for Conifer Forests by Combining Airborne LiDAR Data and Aerial Imagery

Yanjun Su[1], Qinghua Guo[1*], Danny L. Fry[2], Brandon M. Collins[3,4], Maggi Kelly[2], Jacob P. Flanagan[1], and John J. Battles[2]

[1]*Sierra Nevada Research Institute, School of Engineering, University of California at Merced, 5200 North Lake Road, Merced, California 95343, USA*
[2]*Department of Environmental Science, Policy, and Management, University of California at Berkeley, Berkeley, California 94720, USA*
[3]*USDA Forest Service, Pacific Southwest Research Station, Davis, California 95418, USA*
[4]*Center for Fire Research and Outreach, University of California at Berkeley, Berkeley, California 94720, USA*

Abstract. Accurate vegetation mapping is critical for natural resources management, ecological analysis, and hydrological modeling, among other tasks. Remotely sensed multispectral and hyperspectral imageries have proved to be valuable inputs to the vegetation mapping process, but they can provide only limited vegetation structure characteristics, which are critical for differentiating vegetation communities in compositionally homogeneous forests. Light detection and ranging (LiDAR) can accurately measure the forest vertical and horizontal structures and provide a great opportunity for solving this problem. This study introduces a strategy using both multispectral aerial imagery and LiDAR data to map vegetation composition and structure over large spatial scales. Our approach included the use of a *Bayesian information criterion algorithm* to determine the optimized number of vegetation groups within mixed conifer forests in two study areas in the Sierra Nevada, California, and an unsupervised classification technique and post hoc analysis to map these vegetation groups across both study areas. The results show that the proposed strategy can recognize four and seven vegetation groups at the two study areas, respectively. Each vegetation group has its unique vegetation structure characteristics or vegetation species composition. The overall accuracy and kappa coefficient of the vegetation mapping results are over 78% and 0.64 for both study sites.

Résumé. La cartographie précise de la végétation est essentielle entre autres pour la gestion des ressources naturelles, l'analyse écologique, et la modélisation hydrologique. Les approches d'imagerie multispectrale et hyperspectrale par télédétection se sont avérées de précieuses contributions au processus de la cartographie de la végétation, mais elles ne peuvent fournir qu'un nombre limité de caractéristiques sur la structure de la végétation, qui sont essentielles pour différencier les communautés végétales dans les forêts de composition homogènes. La télédétection par laser «light detection and ranging» (LiDAR) peut mesurer avec précision les structures verticales et horizontales de la forêt, et fournit une formidable opportunité de résoudre ce problème. Cette étude présente une stratégie qui utilise à la fois l'imagerie multispectrale aérienne et des données LiDAR pour cartographier la composition et la structure de la végétation à grandes échelles spatiales. Notre approche comprenait l'utilisation d'un algorithme du critère d'information Bayésien pour déterminer le nombre optimal de groupes de végétation dans les forêts mixtes de conifères sur deux zones d'étude dans les Sierra Nevada, en Californie, ainsi qu'une technique de classification non supervisée et une analyse post hoc pour cartographier ces groupes de végétation dans les deux zones d'étude. Les résultats montrent que la stratégie proposée peut reconnaitre quatre et sept groupes de végétation dans les deux zones d'étude respectivement. Chaque groupe de végétation a des caractéristiques uniques de structure de la végétation ou de composition des espèces de la végétation. La précision globale et le coefficient kappa des résultats de la cartographie de la végétation sont de plus de 78% et 0,64 pour les deux sites d'étude.

INTRODUCTION

Vegetation mapping is the process of characterizing vegetation units across a landscape from measured environmental parameters (Franklin 1995; Pedrotti 2012). Typically, these units convey information about the dominant plant species present and the morphological structure of the vegetation (e.g., a mesic hardwood or a high-elevation meadow). Accurate and up-to-date vegetation maps are critical for managers and scientists because they serve a range of functions in natural resource management (e.g., forest inventory, timber harvest, wildfire risk control, wildlife protection), ecological and hydrological modeling, and climate change studies (Chuvieco and Congalton 1988; Tal-

Received 7 May 2015. Accepted 30 November 2015.
*Corresponding author e-mail: guo.qinghua@gmail.com.

bot and Markon 1988; Daly et al. 1994; Stephens 1998; Pearce et al. 2001; Mermoz et al. 2005; Alvarez et al. 2013). Traditional methods for vegetation mapping usually rely on field surveys, literature reviews, aerial photography interpretation, and collateral and ancillary data analysis (Pedrotti 2012). However, these methods are expensive and time consuming. Consequently, vegetation maps produced by the traditional approaches reflect past conditions when released and are not updated frequently (Daly et al. 1994).

Remote sensing has proved to be a powerful tool for vegetation mapping by employing image classification techniques. Multispectral remote sensing imagery such as Landsat, SPOT, MODIS, AVHRR, IKONOS, and QuickBird are among of the most commonly used. For example, Franklin (1986) used the Landsat Thematic Mapper (TM) simulator data to discriminate the composition of conifer forests in the Klamath Mountains in northern California. Carpenter et al. (1999) produced a lifeform map for the Sierra Nevada mountain range in California from Landsat TM data by applying the ARTMAP neural network method. Liu et al. (2006) mapped the distribution of forest disease, sudden oak death, in northern California from two-year images obtained by Airborne Data Acquisition and a Registration system. Mallinis et al. (2008) used an object-based classification method to delineate vegetation polygons in a conifer forest from Quickbird imagery. Wang et al. (2004) combined pixel-based and object-based classification methods to map the different mangrove canopy types along the Caribbean coast of Panama. Zhang et al. (2003) and Knight et al. (2006) monitored vegetation to produce phenology-based land cover maps from MODIS data. As well as multispectral data, hyperspectral imagery is another frequently used data type in vegetation mapping (Hirano et al. 2003; Li et al. 2005). The use of hyperspectral data can produce more finely classified vegetation mapping results than multispectral data can (Xu and Gong 2007; Adam et al. 2010), because hyperspectral sensors are designed to collect data from hundreds of continuous spectral channels compared with multispectral sensors with broad wavelength intervals.

All of these studies that use both multispectral and hyperspectral imagery usually focus only on mapping either the land cover type or the vegetation composition. Examining the detailed structure characteristics in forests has rarely been considered because of the limited penetration capability of multispectral and hyperspectral data. However, this information also plays a very important role in many ecological studies. For example, Lindenmayer et al. (2000) advocated that forest-structure-based parameters can impact biodiversity and should be taken into account in forest management. Zielinski et al. (2006) and García-Feced et al. (2011) demonstrated that forest structure information was critical for mapping the habitat of Pacific fisher (*Pekania pennanti*) and California spotted owl (*Strix occidentalis occidentalis*). Graham et al. (2004), Agee and Skinner (2005) and Peterson et al. (2005) all pointed to the important role that forest structure has on wildfire behavior and argued that modifying forest structure through forest treatment might be necessary to reduce fire risk in many dry conifer forest types. Developing methods to integrate structure information into the process of vegetation mapping is an important area of research.

Light detection and ranging (LiDAR), an active remote sensing technique, can accurately measure the three-dimensional distribution of surface objects (Lefsky et al. 2002). The focused and narrow laser beam used by LiDAR sensors has a strong penetration capability in forest areas (Lim et al. 2003; Jensen 2009; Su and Guo 2014). It has been well documented that LiDAR data can be used to derive highly reliable forest structure parameters such as tree height (Nilsson 1996; Andersen et al. 2006; Su et al. 2015), canopy cover (Lim et al. 2003; Korhonen et al. 2011), leaf area index (Riaño et al. 2004; Jensen et al. 2008), stand volume (Nilsson 1996; Naesset 1997), and tree diameter (Popescu 2007; Huang et al. 2011). The capacity to resolve forest structure parameters provides a great opportunity for developing vegetation-mapping strategies (Kramer et al. 2014). Donoghue et al. (2007) and Heinzel and Koch (2011) explored the possibility of identifying tree species mixtures from parameters derived from LiDAR data. Ørka et al. (2009) and Kim et al. (2009) used LiDAR intensity data to differentiate broadleaf and needleleaf trees. Reitberger et al. (2008) used full-waveform LiDAR data to classify deciduous and coniferous trees. Holmgren and Persson (2004) identified individual tree species, including Norway spruce (*Picea abies L. Karst*), Scots pine (*Pinus sylvestris L.*), and deciduous trees, by analyzing individual crown shape and rich tree structure parameters derived from LiDAR data. However, due to the lack of forest canopy spectral information, the accuracy of tree species classification from LiDAR data is limited in complex vegetation conditions.

The integration of LiDAR data and multispectral/hyperspectral imagery has been used to address the limitation of using only LiDAR data in vegetation mapping. For example, Cho et al. (2012), Colgan et al. (2012) and Naidoo et al. (2012) mapped tree species compositions in African savannas through the combination of LiDAR data and hyperspectral data using maximum likelihood, Random Forest, and Support Vector Machine classifiers, respectively; Dalponte et al. (2012) and Hill and Thomson (2005) classified tree species compositions of broadleaf and coniferous mixed forests through the fusion of spectral and LiDAR data; Holmgren et al. (2008) and Koukoulas and Blackburn (2005) used a maximum likelihood classifier to identify individual tree species from LiDAR-derived structure parameters and multispectral information in deciduous and coniferous forests, respectively. It has been reported that the integration of LiDAR data and optical imagery can increase the vegetation composition classification accuracy by 16%–20% in rangelands, compared to using only LiDAR data or optical imagery (Bork and Su 2007). However, most of these studies on mapping vegetation units are still focusing mainly on classifying vegetated from nonvegetated areas or detecting differences in species composition. Forest

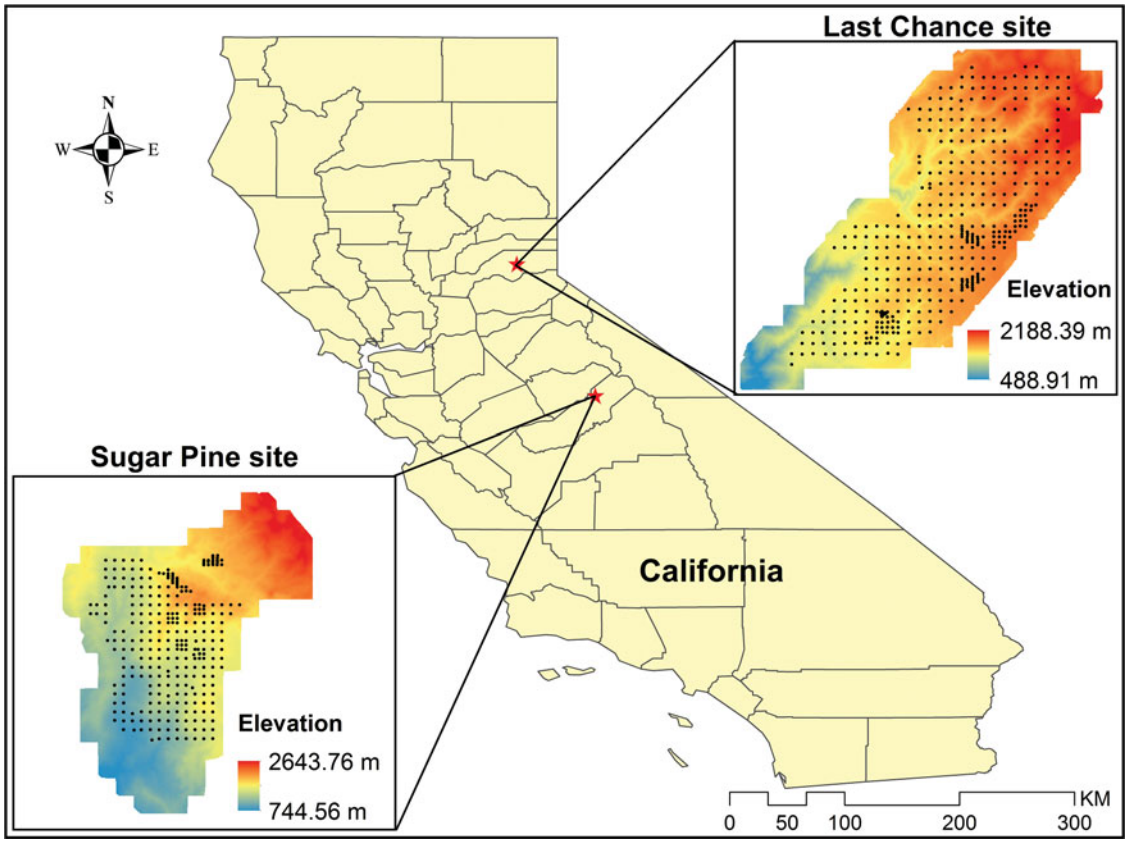

FIG. 1. The geolocations and terrain information of the Last Chance and Sugar Pine study sites with the distribution of field plots.

structure characteristics, which can be estimated by statistical imputation methods that incorporate field measurements with LiDAR data and optical imagery (Falkowski et al. 2010; Hummel at al. 2011; Wallerman and Holmgren 2007), are rarely considered in classification systems.

The objective of this study is to develop and test a new strategy to map vegetation communities in two mixed conifer forests by considering both the dominant tree species composition and vegetation structure characteristics. Multispectral aerial imagery and airborne LiDAR data were integrated, along with a robust network of systematically established field plots in the vegetation mapping process. An unsupervised classification scheme using an automatic cluster determination algorithm based on Bayesian information criterion (BIC) and k-means classification was applied to the fused data to map the vegetation, and a post hoc analysis based on field measurements was used to interpret the ecological properties for each vegetation unit.

MATERIALS AND METHODS

Study Areas

Our two forest study sites are located in the Sierra Nevada mountain range, California, USA (Figure 1). The northern site, Last Chance, covers an area of 92.1 km^2, and the southern site, Sugar Pine, covers an area of 72.8 km^2. The elevation ranges from 280 m to 2190 m for the Last Chance site and from 500 m to 2650 m for the Sugar Pine site, and the average elevation for both study sites is over 1500 m. Trees common to the Sierran mixed conifer and true fir forests dominate the vegetation cover at both sites. The major species present include: ponderosa pine (*Pinus ponderosa*), incense-cedar (*Calocedrus decurrens*), sugar pine (*Pinus lambertiana*), white fir (*Abies concolor*), California red fir (*Abies magnifica*), and Douglas-fir (*Pseudotsuga menziesii*). Within the mixed conifer stands, the major hardwoods are black oak (*Quercus kelloggii*) and canyon live oak (*Quercus chrysolepis*). Forest cover is relatively homogeneous at both the study sites, but the Last Chance site has more heterogeneity than the Sugar Pine site.

Field Measurements

Plot measurements (12.62 m in radius and 500 m^2 in area) were taken in the summer of 2007 and 2008 (Figure 1). The same plot selection procedure was applied to determine the location of 372 and 268 evenly distributed plots at the Last Chance site and Sugar Pine site, respectively. A random point was first chosen to be used as the center of the first plot in each study site. Then, this plot center was taken as a seed point to build

a grid on a 500 m spacing in the four cardinal directions, and the following plot centers were placed on the intersections of the grid. Within watersheds for specific research purposes (e.g., studying hydrological responses to forest fuel treatments), the sampling was intensified to a 250 m by 250 m grid. The position of each plot center in the field was located using a Trimble™ GeoXH GPS. If there were any landing or road surfaces within the plot footprint, the plot center was randomly moved by 25 m in one of the four cardinal directions.

Within each plot, field measurements on individual live trees included: tree species, tree height, diameter at breast height (DBH, breast height = 1.37 m) and height to live crown base. Trees were defined as individuals at least 5 cm in DBH. Moreover, the plot-level canopy cover was measured using a sight tube with 25 sampling points. The plot-level Lorey's height and total basal area were calculated from field measurements and used in the vegetation mapping process in this study, and these can be calculated from the following equations.

$$\mathrm{LH}_z = \frac{\sum_{i=1}^{n_z} \mathrm{BA}_i \times \mathrm{H}_i}{\sum_{i=1}^{n_z} \mathrm{BA}_i}$$

$$\mathrm{TBA}_z = \sum_{i=1}^{n_z} \mathrm{BA}_i,$$

where LH_z and TBA_z represent the Lorey's height and total basal area of the zth plot, and BA_i and H_i are the basal area and tree height of the ith tree in the zth plot.

LiDAR Data

Small footprint airborne LiDAR data covering the Sugar Pine site and Last Chance site were acquired in September 2007 and September 2008 using an Optech GEMINI airborne laser terrain mapper (ALTM) from the National Center of Airborne Laser Mapping at the University of Houston. It was mounted on a twin-engine Cessna Skymaster and was flown at 600 m–700 m above the ground. The ALTM sensor was operated at 100 kHz with a scanning frequency of 40 Hz–60 Hz and a total scan angle of 24°–28°. The average swath width of a single pass was around 510 m, and the overlap between two adjoining swaths was 65% of the swath width. The point density was 6–10 points/m^2, and positioning accuracy was about 10 cm horizontally and 10 cm–15 cm vertically.

Overall, there are 13 layers derived from the raw LiDAR point cloud for both study sites, including the canopy height model (CHM), canopy cover, and 11 canopy quantile metrics. The CHM was calculated by the difference between the LiDAR-derived digital elevation model (DEM) and digital surface model (DSM), which were interpolated from the LiDAR ground returns and LiDAR first returns, respectively. The interpolation algorithm used in this study was ordinary kriging, which has been proved to be more accurate than other schemes (e.g., inverse distance weighted or spline) for interpolating DEM and DSM from LiDAR -derived elevation points (Lloyd and Atkinson 2002; Clark et al. 2004; Guo et al. 2010).

The canopy cover was calculated by a CHM-based method, a reliable and consistent approach for estimating canopy cover from LiDAR data (Lucas et al. 2006). First, a fine resolution CHM (1 × 1 m^2) was calculated from the LiDAR point cloud using the aforementioned algorithm, and the pixels above a selected height threshold were coded as 1 or 0 otherwise. The height threshold was set as 2 m in this study to match field-based canopy cover measurements. Then, this coded CHM was used to overlap with a 20 × 20-m^2 grid, and the canopy cover was calculated as the percentage of the number of coded CHM pixels with a value of 1 to the total number of coded CHM pixels within each 20 × 20-m^2 grid. The final canopy cover layer was produced in 20-m resolution to roughly match the scale of field plots.

Canopy quantile metrics, representing the height below X% of the LiDAR point cloud, are one of most frequently used LiDAR products for estimating the forest parameters that cannot be obtained directly from a LiDAR point cloud, e.g., DBH and biomass (Lim and Treitz 2004; Thomas et al. 2006). In this study, 11 quantile metrics, including 0%, 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, 99%, and 100%, were calculated in 20-m resolution directly from the LiDAR point cloud.

Aerial Imagery

The 2005 National Agriculture Imagery Program (NAIP) color-infrared (CIR) aerial imagery in 1 × 1 m^2 resolution (composed of green band, red band, and near-infrared (NIR) band) are used in the vegetation mapping procedure of this study. The NAIP program is run by the Farm Service of the US Department of Agriculture (USDA) for the purpose of making high-resolution digital orthographies available to maintain common land units. All NAIP images were taken under permitted weather conditions, and followed the specification of no more than 10%-cloud cover per quarter quad tile. The Aerial Photography Field Office has adjusted and balanced the dynamic range of each image tile to the full range of digital number (DN) value (0–255), and orthorectified each image file using the National Elevation Dataset before releasing the data (Hart and Veblen 2015). To ensure the NAIP imagery coregistered with LiDAR data, we georeferenced the NAIP imagery using over 20 correspondence points for each study site selected from NAIP imagery and LiDAR-derived products (i.e., DEMs and CHMs).

In addition to the three spectral bands, seven texture layers (including mean, variance, homogeneity, contrast, dissimilarity, entropy, and second moment) were extracted from each spectral band using the gray-level co-occurrence matrix (GLCM) filtering method. GLCM is defined over an image to be the distribution of co-occurring values at a given offset (Δx, Δy) (Haralick et al. 1973; Anys et al. 1994; Soh and Tsatsoulis 1999), which

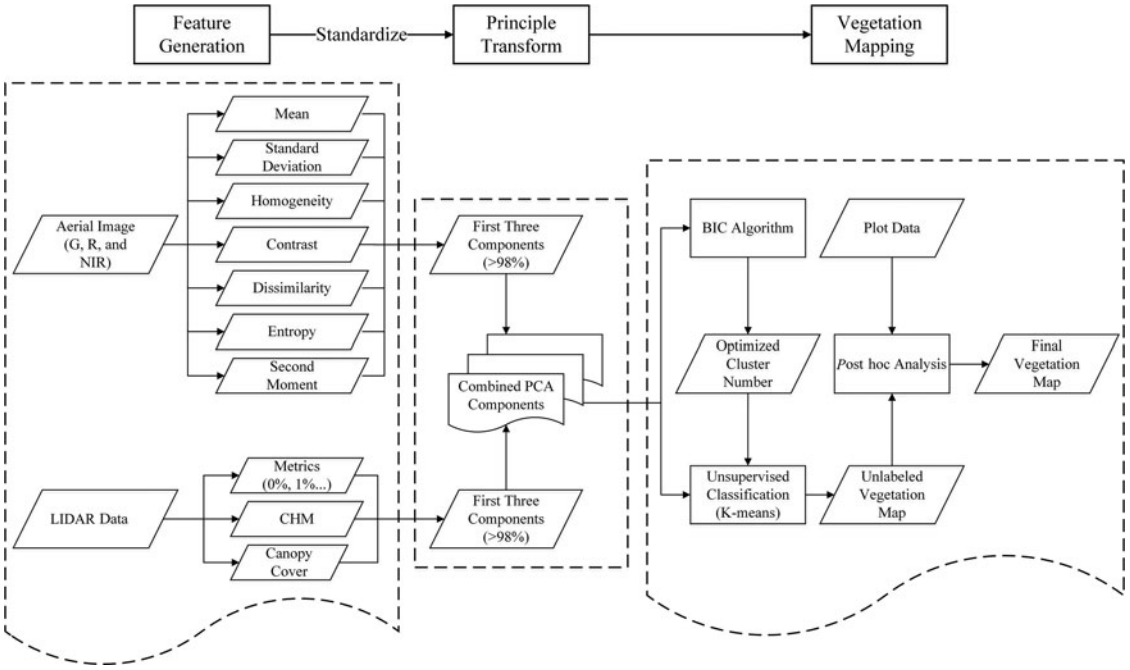

FIG. 2. Procedure for the vegetation-mapping strategy used in this study.

can be mathematically described as

$$\mathrm{GLCM}_{\Delta x,\Delta y}(i,j)=\sum_{p=1}^{m}\sum_{q=1}^{n}\begin{cases}1, \text{if } I(p,q)=i \text{ and } \\ \quad I(p+\Delta x, q+\Delta y)=j \\ 0, \text{otherwise,}\end{cases}$$

where (i, j) is one DN values combination of the image I at the given offset $(\Delta x, \Delta y)$, (p, q) are the spatial position indexes in the image I, and (m, n) are the number of rows and columns of the image I. The offset $(\Delta x, \Delta y)$ is determined by the angular relation between the neighboring pixels and spatial resolution of the image. The texture parameters for the corresponding GLCM can be calculated using equations provided by Haralick et al. (1973), and will not be discussed in detail here. In this study, a 3×3 moving window was used to generate GLCMs and calculate corresponding texture parameters for each cell. To match the spatial scale of the field plots and LiDAR products, the NAIP imagery and obtained texture layers were resampled to the resolution of 20×20 m^2 using the weighted mean value method (Jakubowksi et al. 2013). All of the following vegetation mapping procedures used the resampled NAIP imagery and texture layers.

Vegetation Mapping Strategy

There are, overall, 24 aerial imagery derived features (including the spectral bands and derived texture layers) and 13 LiDAR-derived features initially available for this analysis. This large number of potential input layers for vegetation mapping could negatively influence the results, given the likelihood of redundant information captured by the layers. Many algorithms have been developed to reduce the dimensionality of an input dataset, e.g., principal component analysis (PCA), linear discriminant analysis, correspondence analysis, and detrended correspondence analysis. As one of the most commonly used techniques, the PCA algorithm has been proven to be effective at removing redundant information in remotely sensed data (Mutlu et al. 2008; Pohl and Van Genderen 1998). Therefore, in this study, the standardized PCA method was first applied separately to the aerial-imagery-derived and LiDAR-derived features (Figure 2). The first three PCA components from aerial-imagery-derived features and the first three components from LiDAR-derived features were combined as the input for the vegetation mapping strategy. An unsupervised classification strategy and post hoc analysis integrated with field measurements was then applied on the six PCA components to define vegetation groups and delineate the boundaries of different groups (Figure 2). The detailed descriptions for the unsupervised classification strategy and post hoc analysis are provided following.

Unsupervised Classification Strategy

The specific number and character of vegetation groups within a particular forest are usually unknown prior to the vegetation-mapping process. Thus, one of the main challenges for vegetation mapping is to identify distinct vegetation groups and delineate boundaries among groups. In this study, an automatic cluster-number determination algorithm based on BIC, developed by Chiu et al. (2001), was combined with k-means unsupervised classification to initially map the vegetation. BIC is a robust measure for model selection among a finite set of

models and is defined as:

$$\mathrm{BIC}_k = -2l_k + r_k \log n,$$

where k is the cluster number, l_k is the classification likelihood function, r_k is the number of independent parameters, and n is the number of observations.

To obtain the optimized cluster number, a large maximum cluster number was first defined. In this study we used the hierarchical cluster analysis of species composition (linkage method = Ward's; distance measure = Euclidean) following the method described in McCune et al. (2002) to determine the maximum number of vegetation groups in both study sites. BIC values for all possible cluster numbers (from one to the defined maximum cluster number) were then calculated. With these BIC values, the optimized number of clusters was determined in two steps. First, the initial value of the cluster number was estimated. Let $d\mathrm{BIC}(k)$ be the change of BIC values from two adjacent cluster numbers ($d\mathrm{BIC}(k) = \mathrm{BIC}_k - \mathrm{BIC}_{k-1}$),and $r\mathrm{BIC}(k)$ be the ratio of BIC from k clusters and BIC from only one cluster ($r\mathrm{BIC}(k) = \mathrm{BIC}_k/\mathrm{BIC}_1$). If the $d\mathrm{BIC}(2)$ was larger than 0, the initial cluster number was set as one; otherwise, the initial cluster number was set equal to the number of clusters where $r\mathrm{BIC}(k)$ was smaller than 0.04 for the first time. Second, if the initial cluster number was one, the final cluster number was set as one; otherwise, the ratio change in log-likelihood distance was further used to optimize the cluster number. Let $R(k)$ be the ratio of log-likelihood distances (d_k) from two adjacent cluster numbers ($R(k) = d_k/d_{k-1}$). The ratio of change in log-likelihood was computed as $R(k_1)/R(k_2)$, where k_1 and k_2 were the cluster numbers of the two largest $R(k)$ smaller than the obtained initial cluster number. If the ratio of change was larger than 1.15, the final cluster number was set equal to k_1; otherwise, it was set equal to the maximum value between k_1 and k_2. It should be noted that all the thresholds used in the BIC algorithm were determined by statistical experiments by Chiu et al (2001).

With the optimized cluster number, we used a k-means clustering algorithm to delineate the boundary of different vegetation types. K-means divides observations into a predefined number of clusters, and each observation belongs to the cluster with the nearest mean (Hartigan 1975), which can be mathematically described as:

$$\arg\min \sum_{i=1}^{k_{BIC}} \sum_{x_j \in S_i} \left\| x_j - u_i \right\|^2,$$

where k_{BIC} is the predefined number of clusters, x_j is the jth observation vector, S_i is the ith set of observation vectors, and μ_i is the mean point of the ith set. In this study, the maximum iterations for k-means unsupervised classification was set to 10, and the change threshold of the mean points was set to 5%.

Post hoc Analysis

Field measurements were used to describe the dominant tree species composition and forest structure characteristics. The unsupervised vegetation group for each plot was extracted by overlapping the plot location with an unsupervised classification result. Then, for all plots belonging to the same unsupervised classification group, we analyzed their dominant tree species and forest structure characteristics measured from the field. The dominant tree species were defined by the proportions of different tree species weighted by basal area, and the forest structure characteristics were defined by the plot-level basal area, Lorey's height, and canopy cover. Finally, these plot-derived dominant tree species information and forest structure characteristics were used to determine the property of each unsupervised classification group. It should be noted that approximately two-thirds of the plots (273 in Last Chance and 177 in Sugar Pine) were randomly selected and used to define vegetation group properties. The other plots were reserved to validate the vegetation mapping result.

Accuracy Assessment

PCA ordination analysis, one type of multivariate analysis that can depict species relationships in low-dimensional space (Gauch 1982), was used to evaluate the capability of proposed vegetation mapping strategy on differentiating tree species. It has been widely used as a complement to other data-clustering techniques that help identify repeatable vegetation patterns and discontinuities in species composition (Lepš and Šmilauer 2003). In this study, relative species abundance for ordination analysis was represented by basal area (i.e., the ratio of basal area for each tree species to the total basal area of all trees at a plot). Moreover, the permutation test, a type of robust nonparametric statistical significance test (Nichols and Holmes 2002), was used to evaluate the capability of the proposed vegetation-mapping strategy on recognizing different structure characteristics, because the field-measured forest structure parameters are not normally distributed based on the Shapiro–Wilk test ($\alpha = 0.05$) (Table 1).

In addition, the total accuracy (TA) and kappa coefficient (κ) were also calculated for the purpose of evaluating vegetation-mapping results, which can be denoted as

$$\mathrm{TA} = \frac{a}{N}$$
$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)},$$

where a is the number of plots whose vegetation group agree with the vegetation-mapping result, N is the total number of plots used for accuracy assessment; *Pr(a)* is the relative observed agreement, and *Pr(e)* is the hypothetical probability of chance agreement. The 95% confidence interval for the TA was calculated using the method provided by Foody (2009). About one-third of the plot measurements at each study site were used

TABLE 1
Tests of normality for the forest structure parameters using Shapiro–Wilk test

	Last Chance Site			Sugar Pine Site		
	Statistic	df	Sig.	Statistic	df	Sig.
Lorey's Height	0.630	370	0.000	0.988	268	0.030
Basal Area	0.489	370	0.000	0.941	268	0.000
Canopy Cover	0.988	370	0.003	0.048	268	0.000

to calculate TA and κ. The vegetation group assignments for these test plots were determined by the minimum Mahalanobis distance between these plots and the center of each vegetation group. The parameters used for calculating the Mahalanobis distance include the three forest structure parameters and the coordinates on the primary and secondary axes from the ordination analysis. The center for each vegetation group was calculated by the means of plots used to name vegetation groups. To minimize the influence of the different scales of parameters, all parameters were normalized before calculating the Mahalanobis distance.

RESULTS

Optimized Cluster Number Determination

In this study, the hierarchical cluster analysis result showed that there was never any support for more than eight vegetation classes at either study site. Thus, as a conservative starting point, we approximately doubled the estimate from preliminary results (i.e., 15 vegetation classes) and set it as the upper limit of the BIC cluster number determination algorithm. As shown in Table 2, all *d*BIC values for the Last Chance site were smaller than zero, and the cluster number was 14 when the *r*BIC was smaller than 0.04 for the first time. The initial cluster number was set as 14 for the Last Chance site. When the cluster number was smaller than 14, the two largest *R(k)* values were from results having two clusters and seven clusters. Due to the fact that the ratio between these two *R(k)* was smaller than 1.15, the final optimized cluster number for the Last Chance site was set to seven. Similarly, the final optimized cluster number for the Sugar Pine site was set to four. It should be noted that the initial cluster number for the Sugar Pine site was set to 15 (i.e., the predefined maximum cluster number) because all the *r*BIC values were larger than 0.04.

TABLE 2
The optimized cluster number determination results using Bayesian information criterion (BIC) algorithm for the Last Chance and Sugar Pine study sites

Last Chance Site					Sugar Pine Site				
k	BIC	dBIC[a]	rBIC[b]	R(k)[c]	k	BIC	dBIC[a]	rBIC[b]	R(k)[c]
1	1390210.558				1	1478604.199			
2	1132790.048	−257420.510	1.000	1.942	2	1216658.927	−261945.272	1.000	2.154
3	1000327.759	−132462.289	.515	1.386	3	1095132.435	−121526.492	.464	1.177
4	904805.634	−95522.126	.371	1.236	4	991913.123	−103219.312	.394	1.920
5	827541.566	−77264.068	.300	1.139	5	938229.015	−53684.108	.205	1.055
6	759733.224	−67808.342	.263	1.502	6	887345.487	−50883.528	.194	1.140
7	714639.554	−45093.670	.175	1.774	7	842710.399	−44635.088	.170	1.237
8	689292.817	−25346.737	.098	1.139	8	806647.821	−36062.578	.138	1.461
9	667049.072	−22243.744	.086	1.119	9	782010.413	−24637.409	.094	1.047
10	647181.368	−19867.704	.077	1.318	10	758479.133	−23531.280	.090	1.029
11	632149.235	−15032.134	.058	1.165	11	735610.756	−22868.377	.087	1.009
12	619263.081	−12886.154	.050	1.098	12	712946.401	−22664.355	.087	1.008
13	607542.700	−11720.381	.046	1.407	13	690455.682	−22490.719	.086	1.321
14	599259.290	−8283.410	.032	1.162	14	673467.151	−16988.531	.065	1.058
15	592152.613	−7106.676	.028	1.014	15	657416.013	−16051.138	.061	1.005

[a]The changes (*d*BIC) are from the previous number of clusters in the table.
[b]The ratios of changes (*r*BIC) are relative to the change for the two-cluster solution.
[c]The ratios of distance measures (*R(k)*) are based on the current number of clusters against the previous number of clusters.

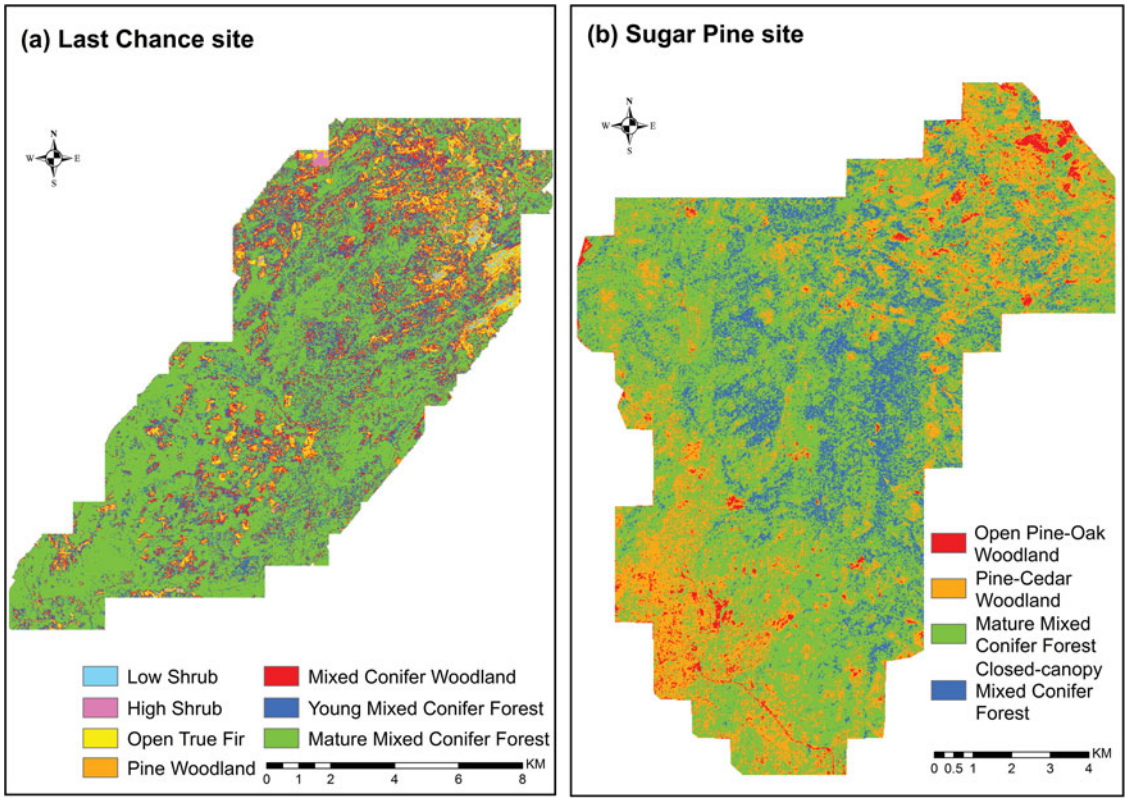

FIG. 3. Labeled vegetation-mapping results for the Last Chance and Sugar Pine sites.

The vegetation mapping results for the Last Chance and Sugar Pine sites are shown in Figure 3. Both sites are dominated by Sierran mixed conifer trees. Specifically, 56% of the Last Chance site was classified as the mature mixed conifer forest, 19% as young mixed conifer forest, and 12.6% as mixed conifer woodland. The young mixed conifer forest was mainly scattered within the mature mixed conifer forest (Figure 3(a)). Pine- and open true fir-dominated forest types were less abundant, covering 7.3% and 3.7% of the study area, respectively (Table 3). These forest types were found mainly at the north end of the study site, and their coverage increased with elevation (Figure 1 and Figure 3(a)). The proportion of the low- and high-shrub types were very small, both around 0.6%. At Sugar Pine, the mature mixed conifer forest again was the most common type, occupying 57.1% of the landscape (Figure 3(b)). Closed-canopy mixed conifer forest was the next most common type, at 25.9% of area, with the greatest concentration in the middle of the study site. The pine-cedar woodland and open pine-oak woodland were distributed at the southeast and northwest of the study site, occupying 13.8% and 3.2%, respectively.

The forest vertical structure information and dominant tree species composition for each vegetation group are shown in Table 3. Naming conventions for the unsupervised groups were based on the dominant tree species (Table 3). If the tree species composition for two vegetation groups is similar, the name recognizes the differences in the forest structures. For example, at the Last Chance site, composition of the dominant tree species for young mixed conifer forest and mature mixed conifer forest are similar, but the mature mixed conifer forest has larger, taller trees and greater canopy cover (Table 3). Note there is no tree information for groups identified as low shrub and high shrub, because no trees were measured with a DBH of 5 cm or greater in these groups.

The capability of the proposed vegetation-mapping strategy to differentiate among dominant species was evaluated by ordination analysis. In Figure 4, the first two axes for both study sites represent over 50% information of all data. The tree species composition among vegetation groups differ greatly with each other at the Last Chance site (Figure 4 (a)). Although the tree species composition of young mixed conifer forest and mature mixed conifer forest are similar (Table 3), the proportion of white fir for the mature mixed conifer forest is larger than that of the young mixed conifer forest, and that for ponderosa pine is smaller (Table 3). At the Sugar Pine site, the proportion of black oak trees for open pine-oak woodland is higher than the other three vegetation groups, which makes it unique among all four vegetation groups (Figure 4 (b)). The tree species compositions for the other three vegetation groups are similar, especially the mature mixed conifer forest and closed-canopy mixed conifer forest. The proportion of white fir and California red fir for the pine-cedar woodland is relatively smaller, compared to the

TABLE 3
Forest structure parameters and dominant tree species for each vegetation group obtained from the k-means unsupervised classification procedure; the dominant tree species are evaluated by the relative basal area of each tree species(Note that certain tree species with too-small relative basal areas for all groups (<1%) were not included in the table)

		Basal Area	Lorey's Height	Canopy Cover	Dominant Tree Species[b] Relative Basal Area (%)								
Group ID	Vegetation Type	(m^2/ha)	(m)	(%)	ABCO	ABMA	CADE	PILA	PIMO	PIPO	PSME	QUKE	LO
					Last Chance Site								
G1	Low Shrub	N/A[a]	N/A[a]	N/A[a]	Manzanita *(Arctostaphylos spp.)*								
G2	High Shrub	N/A[a]	N/A[a]	N/A[a]	Manzanita *(Arctostaphylos spp.)*								
G3	Open True Fir	4.0	10.1	9.2	69	19	0	0	1	11	0	0	0
G4	Pine Woodland	11.2	13.2	21.8	15	5	0	22	0	41	17	0	0
G5	Mixed Conifer Woodland	20.3	15.6	36.4	44	2	8	5	0	21	18	3	0
G6	Young Mixed Conifer Forest	24.7	18.5	46.1	24	1	8	18	0	26	21	1	1
G7	Mature Mixed Conifer Forest	48.3	26.3	61.5	34	4	6	18	0	12	22	3	0
					Sugar Pine Site								
G1	Open Pine-Oak Woodland	11.4	12.2	14.7	0	0	0	3	0	72	0	24	0
G2	Pine-Cedar Woodland	19.8	17.6	38.1	11	1	20	11	0	30	0	10	17
G3	Mature Mixed Conifer Forest	47.3	25.3	66.8	26	1	28	8	0	19	0	8	10
G4	Closed-canopy Mixed Conifer	68.0	32.4	74.6	40	1	29	13	0	9	0	5	2

[a]N/A means the value is not available for corresponding blank.
[b]Species code: ABCO, white fir (*Abies concolor*); ABMA, California red fir (*Abies magnifica*); CADE, incense-cedar (*Calocedrus decurrens*); PILA, sugar pine (*Pinus lambertiana*); PIMO, western white pine (*Pinus monticola*); PIPO, ponderosa pine (*Pinus ponderosa*); PSME, Douglas-fir (*Pseudotsuga menziesii*); QUKE, black oak (*Quercus kelloggii*); LO, canyon live oak (*Quercus chrysolepis*).

mature mixed conifer forest and closed-canopy mixed conifer forest.

The capability of the proposed vegetation-mapping strategy to differentiate the forest vertical structure characteristics was examined by permutation testing under the null hypothesis that the means of vegetation vertical structure parameters among vegetation groups have no difference. Because there were no forest structure parameters for the plots within the low-shrub and high-shrub groups at Last Chance, these two groups were excluded from the permutation test. At the Last Chance site, this null hypothesis is rejected for differences in parameters among all vegetation groups ($\alpha < 0.05$), except the difference of Lorey's height between open true fir and pine woodland and that between pine woodland and mixed conifer woodland (Table 4). For differences in Lorey's height between these two group combinations, the null hypothesis can still be rejected at the significant level of $\alpha = 0.10$. At the Sugar Pine site, the variation in vegetation structure parameters among groups is not as pronounced as at the Last Chance site. The vegetation parameters for the closed-canopy mixed conifer forest are the most distinct. The *p*-values for the differences in all three parameters among the closed-canopy mixed conifer forest and the other three vegetation groups are all smaller than 0.05 except for the difference in canopy cover with mature mixed conifer forest. The basal area and Lorey's height of the mature mixed conifer forest are significantly different from all other groups ($\alpha < 0.05$). However, its canopy cover has no significant difference from all other vegetation groups. The differences in all three parameters between open pine-oak woodland and pine-cedar woodland are not significant.

The accuracy of the vegetation-mapping results was evaluated by the independent plot measurements (Table 5). As can been seen, the overall accuracies of the vegetation-mapping results are around 80% with a 95% confidence interval of $\sim 8\%$ for both study sites, and kappa coefficients are higher than 0.65. At the Last Chance site, the commission errors and omission errors for most vegetation groups are lower than 20%, except the commission errors for the mixed conifer woodland and young mixed conifer forest and the omission error for the mixed conifer woodland. At the Sugar Pine site, all commission and omission errors are lower than 30%, except the omission error for the pine-cedar woodland. The omission rate of the pine-cedar woodland is as high as 41%, and six out of seven omitted pine-cedar woodland plots were misclassified as mature mixed conifer forest.

DISCUSSION

Remote sensing technology has been shown to be extremely helpful for mapping and monitoring vegetation over large spatial

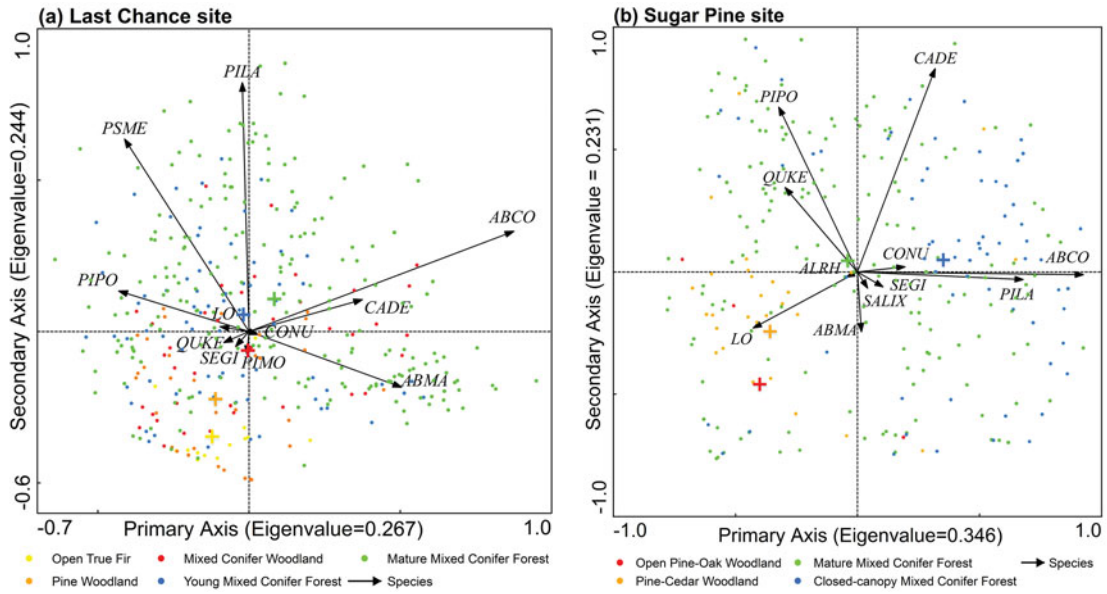

FIG. 4. Ordination analysis results for the Last Chance and Sugar Pine sites. The "+" symbol in each color represents the centroid of the vegetation group represented by the corresponding color in each figure. Species code: ABCO, white fir (*Abies concolor*); ABMA, California red fir (*Abies magnifica*); ALRH, white alder (*Alnus rhombifolia*); CADE, incense-cedar (*Calocedrus decurrens*); CONU, mountain dogwood; LO, canyon live oak (*Quercus chrysolepis*); PILA, sugar pine (*Pinus lambertiana*); PIMO, western white pine (*Pinus monticola*); PIPO, ponderosa pine (*Pinus ponderosa*); PSME, Douglas-fir (*Pseudotsuga menziesii*); QUKE, black oak (*Quercus kelloggii*); SALIX, peachleaf willow (*Salix amygdaloides*); SEGI, giant sequoia (*Sequoiadendron giganteum*).

scales (Xie et al. 2008). However, choosing a classification system that comprehensively captures vegetation community composition and structure is still a major challenge for vegetation mapping from remotely sensed data (Rapp et al. 2005). Traditionally, the number of vegetation units and/or the properties of vegetation units within a forest have been predefined by the prior knowledge of experts from previous experience or field sampling data (Bork and Su 2007; Carpenter et al. 1999; Naidoo et al. 2012). However, this could lead to biased or inconsistent classification systems across regions and might not result in optimal breaks among different vegetation communities. Heinzel and Koch (2011) found that the accuracy of vegetation mapping can increase from 57% to 91% with corresponding decreases in the number of vegetation classes from six to two. It is critical to determine the optimal number of groups that balances the value of recognizing differences in vegetation structure and composition with the reliability of identifying these differences.

By combining the LiDAR data and high-resolution aerial image, this study used a novel automatic cluster number determination algorithm and k-means unsupervised classification to define an optimized classification system. The classification of each vegetation group was determined by fully considering both the vegetation structure characteristics and dominant tree species composition. The results at both study sites show that the proposed vegetation mapping strategy can differentiate vegetation groups by vegetation structure parameters or dominant species composition or both (Figure 3). At the Last Chance site, the small differences in the relative abundance of the common tree species were captured along with steep gradients in structure (Figure 4a, Table 3, and Table 4). Although the tree species composition for the young mixed conifer forest and mature mixed conifer forest were very similar, trees in mature mixed conifer forest were considerably larger than in young mixed conifer forest (Table 3). Similarly, for the low-shrub and high-shrub groups, which were both dominated by manzanita (*Arctostaphylos* spp.), the latter was about 30 cm higher on average than the former, based on the LiDAR-derived CHM. At Sugar Pine, the unsupervised classification clearly detected the pine-oak vegetation type from the matrix of mixed conifer forests (Figure 4b) as well as the structural gradient present (Table 3, Table 4).

Forest structure information, which has been difficult to incorporate in previous vegetation-mapping strategies, is an important factor that has influence on various ecological applications (Peterson et al. 2005; Zielinski et al. 2006) and should be used in the procedure of developing vegetation maps for forest management (Lindenmayer et al. 2000). This is particularly true in more compositionally homogeneous forests. In these forests, traditional vegetation-mapping methods, which rely on passive remote sensing data, might miss the underlying structural differences within the forest. By including LiDAR data, the proposed vegetation-mapping strategy can detect differences in vegetation vertical structure characteristics that, in turn, inform the assessment of wildlife habitat suitability, wildfire hazard, and water

TABLE 4
The p-values of permutation test for the differences in forest structure parameters among different vegetation groups in the Last Chance and Sugar Pine study sites

	Basal Area					Lorey's Height					Canopy Cover				
	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]
	Last Chance Site														
G3[a]	1.000	0.004	0.001	0.000	0.000	1.000	0.066	0.009	0.001	0.000	1.000	0.000	0.000	0.000	0.000
G4[a]	0.004	1.000	0.002	0.000	0.000	0.066	1.000	0.078	0.001	0.000	0.000	1.000	0.000	0.000	0.000
G5[a]	0.001	0.002	1.000	0.013	0.000	0.009	0.078	1.000	0.037	0.000	0.000	0.000	1.000	0.001	0.000
G6[a]	0.000	0.000	0.013	1.000	0.000	0.001	0.001	0.037	1.000	0.000	0.000	0.000	0.001	1.000	0.000
G7[a]	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000
	G1[a]	G2[a]	G3[a]	G4[a]	N/A[b]	G1[a]	G2[a]	G3[a]	G4[a]	N/A[b]	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]
	Sugar Pine Site														
G1[a]	1.000	0.920	0.109	0.013	N/A[b]	1.000	0.806	0.245	0.014	N/A[b]	1.000	0.403	0.684	0.000	N/A[b]
G2[a]	0.920	1.000	0.000	0.000	N/A[b]	0.806	1.000	0.019	0.000	N/A[b]	0.403	1.000	0.613	0.000	N/A[b]
G3[a]	0.109	0.000	1.000	0.000	N/A[b]	0.245	0.019	1.000	0.000	N/A[b]	0.684	0.613	1.000	0.334	N/A[b]
G4[a]	0.013	0.000	0.000	1.000	N/A[b]	0.014	0.000	0.000	1.000	N/A[b]	0.000	0.000	0.334	1.000	N/A[b]

[a]G3 to G7 and G1 to G4 for the Last Chance site and Sugar Pine site represent the corresponding vegetation group listed in Table 3.
[b]N/A means value is not available for corresponding blank.

yield. For example, the Sugar Pine site is dominated by three vegetation types, pine-cedar woodland, mature mixed conifer forest, and closed-canopy mixed canopy forest (Figure 3(b)), which have similar tree species composition (Table 3). Without considering forest vertical structure characteristics from LiDAR data, these three vegetation groups might be classified only as one larger group.

The field measurements of species composition and plot-level forest structure support the results obtained by the unsupervised classification strategy. The proposed vegetation-mapping strategy can produce sufficiently high overall accuracies (nearly 80% in both cases) and kappa coefficients (over 0.64 at both sites) for most applications in which the vegetation map provides the essential classification and scaling information.

TABLE 5
The confusion matrices and accuracy assessments for the vegetation mapping results of Last Chance site and Sugar Pine site

	Last Chance Site									Sugar Pine Site					
	Reference							Commission	Kappa	Reference				Commission	Kappa
Predicted	G1[a]	G2[a]	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]	Error (%)	Coefficient	G1[a]	G2[a]	G3[a]	G4[a]	Error (%)	Coefficient
G1[a]	1	0	0	0	0	0	0	0.0	0.70	1	0	0	0	0	0.64
G2[a]	0	1	0	0	0	0	0	0.0		0	10	1	0	9.1	
G3[a]	0	0	4	0	0	0	0	0.0		0	6	42	5	20.7	
G4[a]	0	0	0	11	2	0	0	15.4		0	1	6	18	28.0	
G5[a]	0	0	0	2	5	0	3	50		N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A[b]	
G6[a]	0	0	0	0	3	12	6	42.9		N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A[b]	
G7[a]	0	0	0	0	1	3	46	8.0		N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A[b]	
Omission Error (%)	0.0	0.0	0.0	15.4	54.5	20.0	16.3	N/A[b]		0.0	41.2	14.3	21.7	N/A[b]	
Overall Accuracy (%)	80.0±7.9 (95% confidence interval)									78.9±8.3 (95% confidence interval)					

[a]G1 to G7 and G1 to G4 for the Last Chance site and Sugar Pine site represent the corresponding vegetation group listed in Table 3.
[b]N/A means value is not available for corresponding blank.

Moreover, the overall accuracy and kappa coefficient obtained from the proposed vegetation-mapping strategy are comparable to most previous supervised vegetation-mapping strategies integrating LiDAR data and multispectral imagery (Bork and Su 2007; Dalponte et al. 2012; Cho et al. 2012).

Although the commission and omission errors for certain vegetation groups were high, they might be caused by misregistration between plot measurements and remotely sensed data (LiDAR data and aerial imagery). The plot locations were measured using a GPS in the field. Although it can produce centimeter-level positioning accuracy in most cases, the blocking effect of forest canopy can reduce the GPS positioning accuracy significantly (Sigrist et al. 1999). The possible positioning error may lead to poor coregistration with remotely sensed data. Particularly, this misregistration could have a pronounced effect on the commission and omission errors of vegetation groups that do not cluster together. For example, the young mixed conifer forest in the Last Chance site had both a relatively high commission error and omission error. Instead of aggregating together, young mixed conifer forest was mainly scattered within mature mixed conifer forest (Figure 3(a)). A commission error of 66.7% for young mixed conifer forest was due to the misclassification as mature mixed conifer forest.

The quality of NAIP aerial imagery could be another factor that influences the vegetation-mapping accuracy. As known, there is nonlinear color balancing effect existing in the NAIP imagery due to the dynamic range of different image tiles and different data-acquiring time (Hart and Veblen 2015). Moreover, the absolute horizontal accuracy for the NAIP imagery is around 6 m at a 95% confidence level (USDA Farm Service Agency 2015). Although this study has tried to reduce the influence of misregistration between NAIP imagery and LiDAR products by matching correspondence points, it still cannot be totally eliminated. Further study is still needed to address how the nonlinear color balancing effect and horizontal accuracy influence the vegetation-mapping accuracy. Moreover, it has been frequently reported that hyperspectral data outperformed multispectral data in recognizing plant species (Adam et al. 2010; Xu and Gong 2007), and there have been studies showing that the integration of hyperspectral data and LiDAR data can produce more accurate vegetation maps than the integration of multispectral data and LiDAR data (Dalponte et al. 2012).

CONCLUSIONS

This study proposed a vegetation-mapping strategy through the combination of multispectral aerial imagery and LiDAR data. Both the vegetation structure and composition information were taken into consideration of the determination of classification system. The BIC algorithm was used to automatically optimize the number of vegetation units within two mixed conifer forests, and the property of each vegetation group was identified by post hoc analysis based on field measurements. The results show that the proposed vegetation-mapping strategy is a robust method to map vegetation in mixed conifer forests with a sufficient high accuracy. The overall accuracy and kappa coefficient are over 78% and 0.64 for both study sites. Each identified vegetation group can be differentiated from others by vegetation structure parameters or dominant species composition or both. The obtained vegetation maps have the potential to considerably improve the identification of critical habitat for species of concern (e.g., Pacific fisher and California spotted owl), as well as identifying wildfire risk through characterizing ladder fuels (Kramer et al. 2014).

FUNDING

This is SNAMP Publication Number #41. The Sierra Nevada Adaptive Management Project is funded by USDA Forest Service Region 5, USDA Forest Service Pacific Southwest Research Station, US Fish and Wildlife Service, California Department of Water Resources, California Department of Fish and Game, California Department of Forestry and Fire Protection, and the Sierra Nevada Conservancy.

REFERENCES

Adam, E., Mutanga, O., and Rugege, D. 2010. "Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review." *Wetlands Ecology and Management*, Vol. 18(No. 3): pp. 281–296.

Agee, J.K., and Skinner, C.N. 2005. "Basic principles of forest fuel reduction treatments." *Forest Ecology and Management*, Vol. 211(No. 1): pp. 83–96.

Alvarez, O., Guo, Q., Klinger, R.C., Li, W., and Doherty, P. 2013. "Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation" *International Journal of Climatology*, Vol. 34(No. 7): pp. 2258–2268.

Andersen, H.-E., Reutebuch, S.E., and McGaughey, R.J. 2006. "A rigorous assessment of tree height measurements obtained using airborne LiDAR and conventional field methods." *Canadian Journal of Remote Sensing*, Vol. 32(No. 5): pp. 355–366.

Anys, H., Bannari, A., He, D., and Morin, D. 1994. "Texture analysis for the mapping of urban areas using airborne MEIS-II images." In *Proceedings of the First International Airborne Remote Sensing Conference and Exhibition*, 11–15 September 1994, Strasbourg, France, (ERIM), Vol. III, pp. 231–245.

Bork, E.W., and Su, J.G. 2007. "Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis." *Remote Sensing of Environment*, Vol. 111(No. 1): pp. 11–24.

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., Woodcock, C.E., and Franklin, J. 1999. "A neural network method for efficient vegetation mapping." *Remote Sensing of Environment*, Vol. 70(No. 3): pp. 326–338.

Chiu, T., Fang, D., Chen, J., Wang, Y., and Jeris, C. 2001. "A robust and scalable clustering algorithm for mixed type attributes in large database environment." In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, 263–268. New York: ACM.

Cho, M.A., Mathieu, R., Asner, G.P., Naidoo, L., van Aardt, J., Ramoelo, A., Debba, P., Wessels, K., Main, R., and Smit, I.P. 2012. "Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system." *Remote Sensing of Environment*, Vol. 125: pp. 214–226.

Chuvieco, E., and Congalton, R.G. 1988. "Mapping and inventory of forest fires from digital processing of TM data." *Geocarto International*, Vol. 3(No. 4): pp. 41–53.

Clark, M.L., Clark, D.B., and Roberts, D.A. 2004. "Small-footprint LiDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape." *Remote Sensing of Environment*, Vol. 91(No. 1): pp. 68–89.

Colgan, M.S., Baldeck, C.A., Féret, J.-B., and Asner, G.P. 2012. "Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data." *Remote Sensing*, Vol. 4(No. 11): pp. 3462–3480.

Daly, C., Neilson, R.P., and Phillips, D.L. 1994. "A statistical-topographic model for mapping climatological precipitation over mountainous terrain." *Journal of Applied Meteorology*, Vol. 33(No. 2): pp. 140–158.

Dalponte, M., Bruzzone, L., and Gianelle, D. 2012. "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data." *Remote Sensing of Environment*, Vol. 123: pp. 258–270.

Donoghue, D.N., Watt, P.J., Cox, N.J., & Wilson, J. (2007). Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment*, vol. 110(no. 4): pp. 509–522.

Falkowski, M.J., Hudak, A.T., Crookston, N.L., Gessler, P.E., Uebler, E.H., and Smith, A.M. 2010. "Landscape-scale parameterization of a tree-level forest growth model: a k-nearest neighbor imputation approach incorporating LiDAR data." *Canadian Journal of Forest Research*, Vol. 40(No. 4): pp. 184–199.

Foody, G. M., 2009." Sample size determination for image classification accuracy assessment and comparison." *International Journal of Remote Sensing*, Vol. 30(No. 20): pp. 5273–5291.

Franklin, J. 1986. "Thematic Mapper analysis of coniferous forest structure and composition." *International Journal of Remote Sensing*, Vol. 7(No. 10): pp. 1287–1301.

Franklin, J. 1995. "Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients."*Progress in Physical Geography*, Vol. 19(No. 4): pp. 474–499.

García-Feced, C., Tempel, D.J., and Kelly, M. 2011. "LiDAR as a tool to characterize wildlife habitat: California spotted owl nesting habitat as an example." *Journal of Forestry*, Vol. 109(No. 8): pp. 436–443.

Gauch, H.G. 1982. *Multivariate analysis in community ecology*. New York, NY: Cambridge University Press.

Graham, R.T., McCaffrey, S., and Jain, T.B. 2004. *Science Basis for Changing Forest Structure to Modify Wildfire Behavior and Severity*, General Technical Report, RMRS-120. Fort Collins, Colorado, USA: USDA Forest Service, Rocky Mountain Research Station.

Guo, Q., Li, W., Yu, H., and Alvarez, O. 2010. "Effects of topographic variability and LiDAR sampling density on several DEM interpolation methods." *Photogrammetric Engineering and Remote Sensing*, Vol. 76(No. 63): pp. 701–712.

Haralick, R.M., Shanmugam, K., and Dinstein, I.H. 1973. "Textural features for image classification." *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-3(No. 6): pp. 610–621.

Hart, S.J., and Veblen, T.T. 2015. "Detection of spruce beetle-induced tree mortality using high-and medium-resolution remotely sensed imagery." *Remote Sensing of Environment*, Vol. 168(No. 10): pp.134–145.

Hartigan, J.A. 1975. *Clustering algorithms*. John Wiley & Sons.

Heinzel, J., and Koch, B. 2011. "Exploring full-waveform LiDAR parameters for tree species classification." *International Journal of Applied Earth Observation and Geoinformation*, Vol. 13(No. 1): pp. 152–160.

Hill, R., and Thomson, A. 2005. "Mapping woodland species composition and structure using airborne spectral and LiDAR data." *International Journal of Remote Sensing*, Vol. 26(No. 17): pp. 3763–3779.

Hirano, A., Madden, M., and Welch, R. 2003. "Hyperspectral image data for mapping wetland vegetation." *Wetlands*, Vol. 23(No. 2): pp. 436–448.

Holmgren, J., and Persson, Å. 2004. "Identifying species of individual trees using airborne laser scanner." *Remote Sensing of Environment*, Vol. 90(No. 4): pp. 415–423.

Holmgren, J., Persson, Å., and Söderman, U. 2008. "Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images." *International Journal of Remote Sensing*, Vol. 29(No. 5): pp. 1537–1552.

Huang, H., Li, Z., Gong, P., Cheng, X., Clinton, N., Cao, C., Ni, W., and Wang, L. 2011. "Automated methods for measuring DBH and tree heights with a commercial scanning LiDAR." *Photogrammetric Engineering and Remote Sensing*, Vol. 77(No. 3): pp. 219–227.

Hummel, S., Hudak, A., Uebler, E., Falkowski, M., and Megown, K. 2011. "A comparison of accuracy and cost of LiDAR versus stand exam data for landscape management on the Malheur National Forest." *Journal of Forestry*, Vol. 109(No. 5): pp. 267–273.

Jakubowksi, M.K., Guo, Q., Collins, B., Stephens, S., and Kelly, M. 2013. "Predicting surface fuel models and fuel metrics using LiDAR and CIR imagery in a dense, mountainous forest." *Photogrammetric Engineering and Remote Sensing*, Vol. 79(No. 1): pp. 37–49.

Jensen, J.L., Humes, K.S., Vierling, L.A., and Hudak, A.T. 2008. "Discrete return LiDAR-based prediction of leaf area index in two conifer forests." *Remote Sensing of Environment*, Vol. 112(No. 10): pp. 3947–3957.

Jensen, J.R., 2009. *Remote sensing of the environment: An Earth resource perspective (2nd Edition)*. Upper Saddle River, NJ: Pearson Education India.

Kim, S., McGaughey, R.J., Andersen, H.-E., and Schreuder, G. 2009. "Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data." *Remote Sensing of Environment*, Vol. 113(No. 8): pp. 1575–1586.

Knight, J.F., Lunetta, R.S., Ediriwickrema, J., and Khorram, S. 2006. "Regional scale land cover characterization using MODIS-NDVI 250 m multi-temporal imagery: A phenology-based approach." *GIScience & Remote Sensing*, Vol. 43(No. 1): pp. 1–23.

Korhonen, L., Korpela, I., Heiskanen, J., and Maltamo, M. 2011. "Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index." *Remote Sensing of Environment*, Vol. 115(No. 4): pp. 1065–1080.

Koukoulas, S., and Blackburn, G.A. 2005. "Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data." *International Journal of Remote Sensing*, Vol. 26(No. 3): pp. 431–455.

Kramer, H. A., B. M. Collins, S. L. Stephens, and M. Kelly. 2014. "Quantifying ladder fuels: a new approach using LiDAR." *Forests*, Vol. 5(No. 6): pp. 1432–1453.

Lefsky, M.A., Cohen, W.B., Parker, G.G., and Harding, D.J. 2002. "LiDAR remote sensing for ecosystem studies LiDAR, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists." *BioScience*, Vol. 52(No. 1): pp. 19–30.

Lepš, J., and Šmilauer, P. 2003. *Multivariate analysis of ecological data using CANOCO*. Cambridge, UK: Cambridge University Press.

Li, L., Ustin, S., and Lay, M. 2005. "Application of multiple end-member spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China Camp, CA, USA." *International Journal of Remote Sensing*, Vol. 26(No. 23): pp. 5193–5207.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M. 2003. "LiDAR remote sensing of forest structure." *Progress in Physical Geography*, Vol. 27(No. 1): pp. 88–106.

Lim, K.S., and Treitz, P.M. 2004. "Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators." *Scandinavian Journal of Forest Research*, Vol. 19(No. 36): pp. 558–570.

Lindenmayer, D.B., Margules, C.R., and Botkin, D.B. 2000. "Indicators of biodiversity for ecologically sustainable forest management." *Conservation Biology*, Vol. 14(No. 4): pp. 941–950.

Liu, D., Kelly, M., and Gong, P. 2006. "A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery." *Remote Sensing of Environment*, Vol. 101(No. 2): pp. 167–180.

Lloyd, C.D., and Atkinson, P.M. 2002. "Deriving DSMs from LiDAR data with kriging." *International Journal of Remote Sensing*, Vol. 23(No. 12): pp. 2519–2524.

Lucas, R.M., Cronin, N., Lee, A., Moghaddam, M., Witte, C., and Tickle, P. 2006. "Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia." *Remote Sensing of Environment*, Vol. 100(No. 3): pp. 407–25.

Mallinis, G., Koutsias, N., Tsakiri-Strati, M., and Karteris, M. 2008. "Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site." *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 63(No. 2): pp. 237–250.

Mermoz, M., Kitzberger, T., and Veblen, T.T. 2005. "Landscape influences on occurrence and spread of wildfires in Patagonian forests and shrublands." *Ecology*, Vol. 86(No. 10): pp. 2705–2715.

McCune, B., Grace, J. B., and Urban, D. L. 2002. *Analysis of ecological communities (Vol. 28)*. Gleneden Beach, OR: MjM Software Design.

Mutlu, M., Popescu, S.C., Stripling, C., and Spencer, T. 2008. "Mapping surface fuel models using LiDAR and multispectral data fusion for fire behavior." *Remote Sensing of Environment*, Vol. 112(No. 1): pp. 274–285.

Naesset, E., 1997. "Estimating timber volume of forest stands using airborne laser scanner data." *Remote Sensing of Environment*, Vol. 61(No. 2): pp. 246–253.

Naidoo, L., Cho, M., Mathieu, R., and Asner, G. 2012. "Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment." *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 69: pp. 167–179.

Nichols, T.E., and Holmes, A.P. 2002. "Nonparametric permutation tests for functional neuroimaging: a primer with examples." *Human Brain Mapping*, Vol. 15(No. 1): pp. 1–25.

Nilsson, M. 1996. "Estimation of tree heights and stand volume using an airborne LiDAR system." *Remote Sensing of Environment*, Vol. 56(No. 1): pp. 1–7.

Ørka, H.O., Næsset, E., and Bollandsås, O.M. 2009. "Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data." *Remote Sensing of Environment*, Vol. 113(No. 6): pp. 1163–1174.

Pearce, J., Cherry, K., and Whish, G. 2001. "Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution." *Journal of Applied Ecology*, Vol. 38(No. 2): pp. 412–424.

Pedrotti, F. 2012. *Plant and vegetation mapping*. New York: Springer.

Peterson, D.L., Johnson, M.C., Agee, J.K., Jain, T.B., McKenzie, D., and Reinhardt, E.D. 2005. *Forest Structure and Fire Hazard in Dry Forests of the Western United States*, General Technical Report, PNWGTR-268. Portland, Oregon: USDA Forest Service.

Pohl, C., and Van Genderen, J.L. 1998. "Review article multisensor image fusion in remote sensing: concepts, methods and applications." *International Journal of Remote Sensing*, Vol. 19(No. 5): pp. 823–854.

Popescu, S.C. 2007. "Estimating biomass of individual pine trees using airborne LiDAR.' *Biomass and Bioenergy*, Vol. 31(No. 9): pp. 646–655.

Rapp, J., Wang, D., Capen, D., Thompson, E., and Lautzenheiser, T. 2005. 'Evaluating error in using the National Vegetation Classification System for ecological community mapping in northern New England, USA." *Natural Areas Journal*, Vol. 25: pp. 46–54.

Reitberger, J., Krzystek, P., and Stilla, U. 2008. "Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees." *International Journal of Remote Sensing*, Vol. 29(No. 5): pp. 1407–1431.

Riaño, D., Valladares, F., Condés, S., and Chuvieco, E. 2004. "Estimation of leaf area index and covered ground from airborne laser scanner (LiDAR) in two contrasting forests." *Agricultural and Forest Meteorology*, Vol. 124(No. 3): pp. 269–275.

Sigrist, P., Coppin, P., and Hermy, M. 1999. "Impact of forest canopy on quality and accuracy of GPS measurements." *International Journal of Remote Sensing*, Vol. 20(No. 18): pp. 3595–3610.

Soh, L.-K., and Tsatsoulis, C. 1999. "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37(No. 2): pp. 780–795.

Stephens, S.L. 1998. "Evaluation of the effects of silvicultural and fuels treatments on potential fire behaviour in Sierra Nevada mixed-conifer forests." *Forest Ecology and Management*, Vol. 105(No. 1): pp. 21–35.

Su, Y., and Guo, Q. 2014. "A practical method for SRTM DEM correction over vegetated mountain areas." *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 87(No. 1): pp. 216–228.

Su, Y., Guo, Q., Ma, Q., and Li, W. 2015. "SRTM DEM Correction in vegetated mountain areas through the integration of spaceborne LiDAR, airborne LiDAR, and optical imagery." *Remote Sensing*, Vol. 7(No. 9): pp. 11202–11225.

Talbot, S., and Markon, C. 1988. "Intermediate-scale vegetation mapping of Innoko National Wildlife Refuge, Alaska using Landsat MSS

digital data." *Photogrammetric Engineering and Remote Sensing*, Vol. 54(No. 3): pp. 377–383.

Thomas, V., Treitz, P., McCaughey, J., and Morrison, I. 2006. "Mapping stand-level forest biophysical variables for a mixedwood boreal forest using LiDAR: an examination of scanning density." *Canadian Journal of Forest Research*, Vol. 36(No. 1): pp. 34–47.

USDA Farm Service Agency. 2015. NAIP Imagery (online). Accessed 3 May 2015. Available from http://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/index.

Wallerman, J., and Holmgren, J. 2007. "Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data." *Remote Sensing of Environment*, Vol. 110(No. 4): pp. 501–508.

Wang, L., Sousa, W., and Gong, P. 2004. "Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery." *International Journal of Remote Sensing*, Vol. 25(No. 24): pp. 5655–5668.

Xie, Y., Sha, Z., and Yu, M. 2008. "Remote sensing imagery in vegetation mapping: a review." *Journal of Plant Ecology*, Vol. 1(No. 1): pp. 9–23.

Xu, B., and Gong, P. 2007. "Land-use/land-cover classification with multispectral and hyperspectral EO-1 data." *Photogrammetric Engineering and Remote Sensing*, Vol. 73(No. 8): pp. 955–965.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., Reed, B.C., and Huete, A. 2003. "Monitoring vegetation phenology using MODIS." *Remote Sensing of Environment*, Vol. 84(No. 4): pp. 471–475.

Zielinski, W.J., Truex, R.L., Dunk, J.R., and Gaman, T. 2006. "Using forest inventory data to assess fisher resting habitat suitability in California." *Ecological Applications*, Vol. 16(No. 3): pp. 1010–1025.