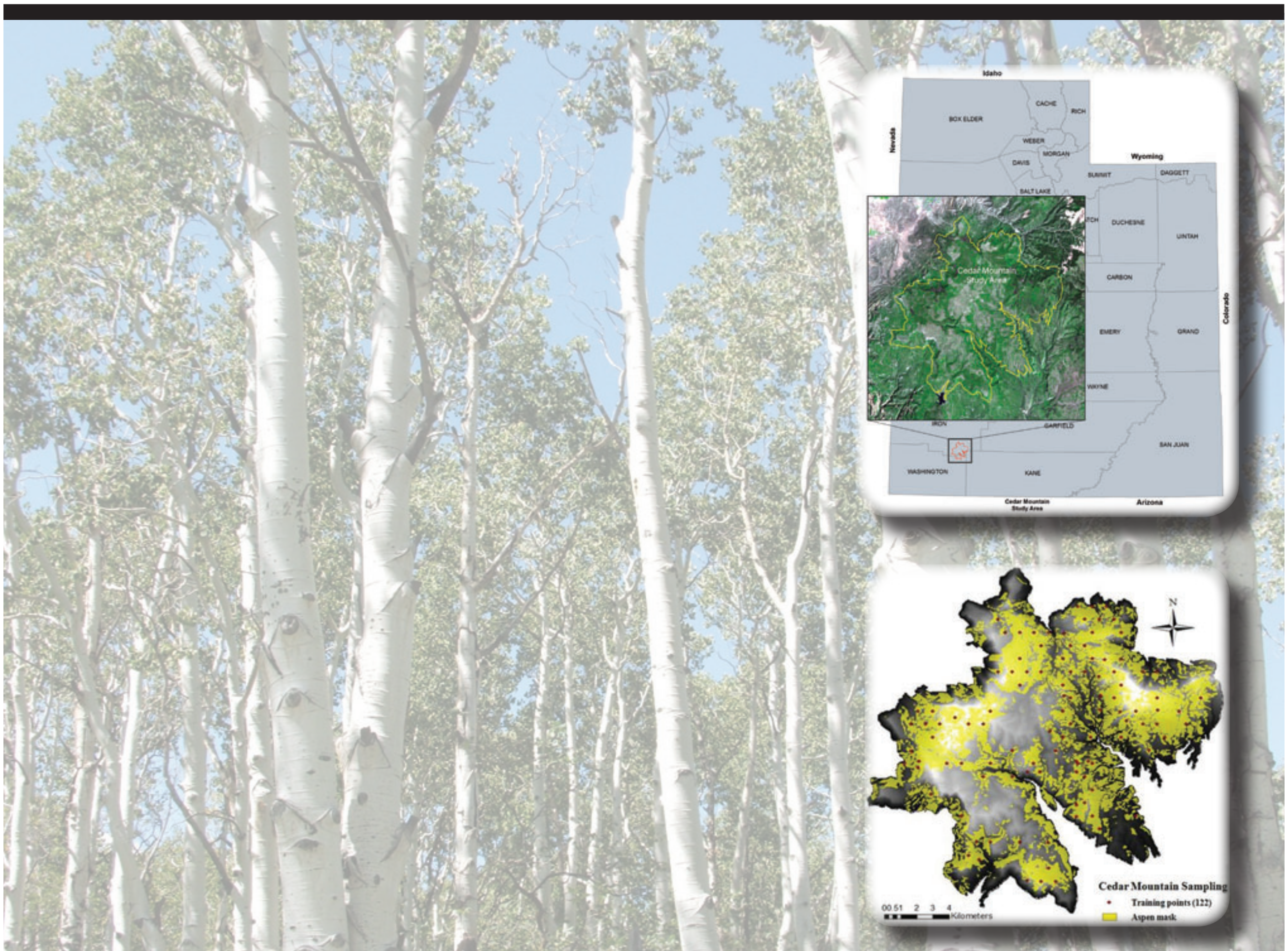
Moderate-Scale Mapping Methods of Aspen Stand Types:

A Case Study for Cedar Mountain in Southern Utah

Chad M. Oukrop, David M. Evans, Dale L. Bartos, R. Douglas Ramsey, Ronald J. Ryel



United States Department of Agriculture / Forest Service
Rocky Mountain Research Station
General Technical Report RMRS-GTR-259
July 2011

Citation

Oukrop, Chad M.; Evans, David M.; Bartos, Dale L.; Ramsey, R. Douglas; Ryel, Ronald J. 2011. **Moderate-scale mapping methods of aspen stand types: a case study for Cedar Mountain in southern Utah.** Gen. Tech. Rep. RMRS-GTR-259. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 18 p.

Abstract

Quaking aspen (*Populus tremuloides* Michx.) are the most widely distributed tree species across North America, but its dominance is declining in many areas of the western United States, with certain areas experiencing rapid mortality events over the past decade. The loss of aspen from western landscapes will continue to profoundly impact biological, commercial, and aesthetic resources associated with aspen. However, many options are available for its restoration. Advances in remote sensing technologies offer cost-effective means to produce spatial and quantitative information on the distribution and severity of declining aspen at many scales. This report describes the development and application of transferable remote sensing and geographic information system methodologies to accurately classify aspen condition within areas of delineated aspen woodland cover. These methodologies were applied on Cedar Mountain in southern Utah within the Colorado Plateau to map three aspen stand conditions (healthy, damaged, and seral) successfully. Using moderate-scale imagery (2008 Landsat TM data), digital elevation model derivatives, high-resolution National Agriculture Imagery Program imagery, and a decision tree modeling approach, a spatially explicit 2008 landscape assessment of Cedar Mountain aspen was produced with an overall accuracy of 81.3% (Kappa [κ] or KHAT accuracy measure = 0.69, n = 445). Of the total area mapped as aspen within the 12,139-ha study area, healthy aspen was the most abundant with 49% (5960 ha), followed by damaged with 35% (4210 ha), and seral with an estimated 16% (1968 ha) coverage. Aspen classification maps, derived from remotely sensed digital imagery and ancillary datasets, can offer objective management information to land managers to utilize when planning, implementing, and evaluating aspen restoration activities.

Keywords: quaking aspen, sudden aspen decline, Landsat TM, NAIP imagery, succession, GIS

Authors

Chad M. Oukrop is an Ecologist with the Forest Inventory and Analysis located at the Forestry Sciences Laboratory, Anchorage, Alaska.

David M. Evans is an Ecological Site Inventory Specialist with the Natural Resource Conservation Service, Sonora Soil Survey Office, Sonora, California.

Dale L. Bartos is an Aspen Ecologist with the USDA Forest Service, Rocky Mountain Research Station and an Adjunct Assistant Professor in the Department of Wildland Resources at Utah State University, Logan.

R. Douglas Ramsey is a Professor in the Department of Wildland Resources at Utah State University, Logan.

Ronald J. Ryel is a Professor in the Department of Wildland Resources at Utah State University, Logan.

You may order additional copies of this publication by sending your mailing information in label form through one of the following media. Please specify the publication title and series number.

Publishing Services

Telephone	(970) 498-1392
FAX	(970) 498-1122
E-mail	rschneider@fs.fed.us
Website	http://www.fs.fed.us/rm/publications
Mailing address	Publications Distribution Rocky Mountain Research Station 240 West Prospect Road Fort Collins, CO 80526

Contents

Moderate-Scale Mapping Methods of Aspen Stand Types:

A Case Study for Cedar Mountain in Southern Utah

Chad M. Oukrop, David M. Evans, Dale L. Bartos, R. Douglas Ramsey, Ronald J. Ryel

Introduction

Quaking aspen (*Populus tremuloides*, Michx.) are the most widely distributed native tree species in North America, occurring broadly from the northeastern coast across the North American boreal forest into Alaska and southward through the Rocky Mountains into Mexico (Baker 1925). Aspen are the predominant deciduous tree of the Rocky Mountain Region, with the highest abundances in Colorado and Utah (Preston 1976; Bartos 2001). Communities dominated by aspen are noted for forage production, understory diversity, wildlife habitat, watershed protection, water yield, timber products, and aesthetic appeal (Preston 1976; Bartos and Campbell 1998; Bartos 2001; LaMalfa and Ryel 2008).

Despite their apparent merit, aspen communities in portions of the Intermountain West are in decline. Possible factors contributing to aspen decline are climatic change, Twentieth Century wildfire suppression that has led to conifer succession (Bartos 2001), domestic and native ungulate browsing of regenerating aspen (Bartos 2001; Sexton and others 2006), and insect and disease outbreaks that affect stressed aspen stands (Hogg and Schwarz 1999). More recently, rapid rates of aspen mortality were reported in southwestern Colorado (Worrall and others 2008), northern Arizona (M.L. Fairweather personal communication), southern Utah (J. Bowns personal communication), and Montana (W.D. Sheppard, personal communication). This recent phenomenon is referred to as Sudden Aspen Decline (SAD) and is characterized by rapid overstory mortality with little to no understory regeneration (Bartos and Shepperd 2010). Complete mortality of the stand generally occurs within a two to three year period, and there are striking similarities in the suddenness and synchronicity of the decline in stands showing SAD. The severity of SAD differs considerably from the typical aspen decline, which can be defined as either advancing conifer succession or the gradual deterioration of vigor and health (10 to 20 years or more) (Sinclair and Lyon 2005).

There is an increased focus on the threats to natural resources (hydrological, biological, and aesthetic values) concomitant with the loss of aspen from the Intermountain West (Bartos and Shepperd 2010). Resource managers, however, often lack cost-effective resources needed to properly assess and restore aspen over large areas. Advances in remote sensing technology (Prince and others 1995; Moskal and Franklin 2004) offer viable options to acquire extensive spatial and quantitative information on the location, extent, and severity of aspen decline at most ownership and management levels. Aspen distribution maps, derived from remotely sensed imagery and ancillary datasets, can offer land managers objective, large-scale management information for planning, implementing, and evaluating aspen restoration activities.

In the past decade, numerous efforts have been made to map aspen ecosystems. With an increased availability of space-borne sensors that collect imagery at multiple spatial and spectral resolutions (e.g., MODIS, Landsat, SPOT, and IKONOS), coupled with improved computing and processing power; scientist, analysts, and land managers alike have developed better techniques to map aspen systems at local and regional scales (Heide 2002; Strand and others 2009; Lowry and others 2007). However, these efforts have often had very low accuracy measures and are unreliable for management purposes. Furthermore, no studies have addressed SAD specifically, nor have they discriminated pure (i.e., persistent) aspen stand types into independent classes. Perhaps the most relevant study regarding SAD was the Worrall and others (2008) study, which utilized an aerial sketch-mapping technique to classify aspen into healthy and damaged classes. In this study they found extensive aspen mortality in the San Juan range of southern Colorado that was strikingly similar to that found on Cedar Mountain in southern Utah. However, aerial sketch-mapping techniques are often very expensive and are clouded by spatial error and surveyor subjectivity. Consequently, estimation and classification of stand data can often be misleading and erroneous. Thus, effective remote sensing geographic information

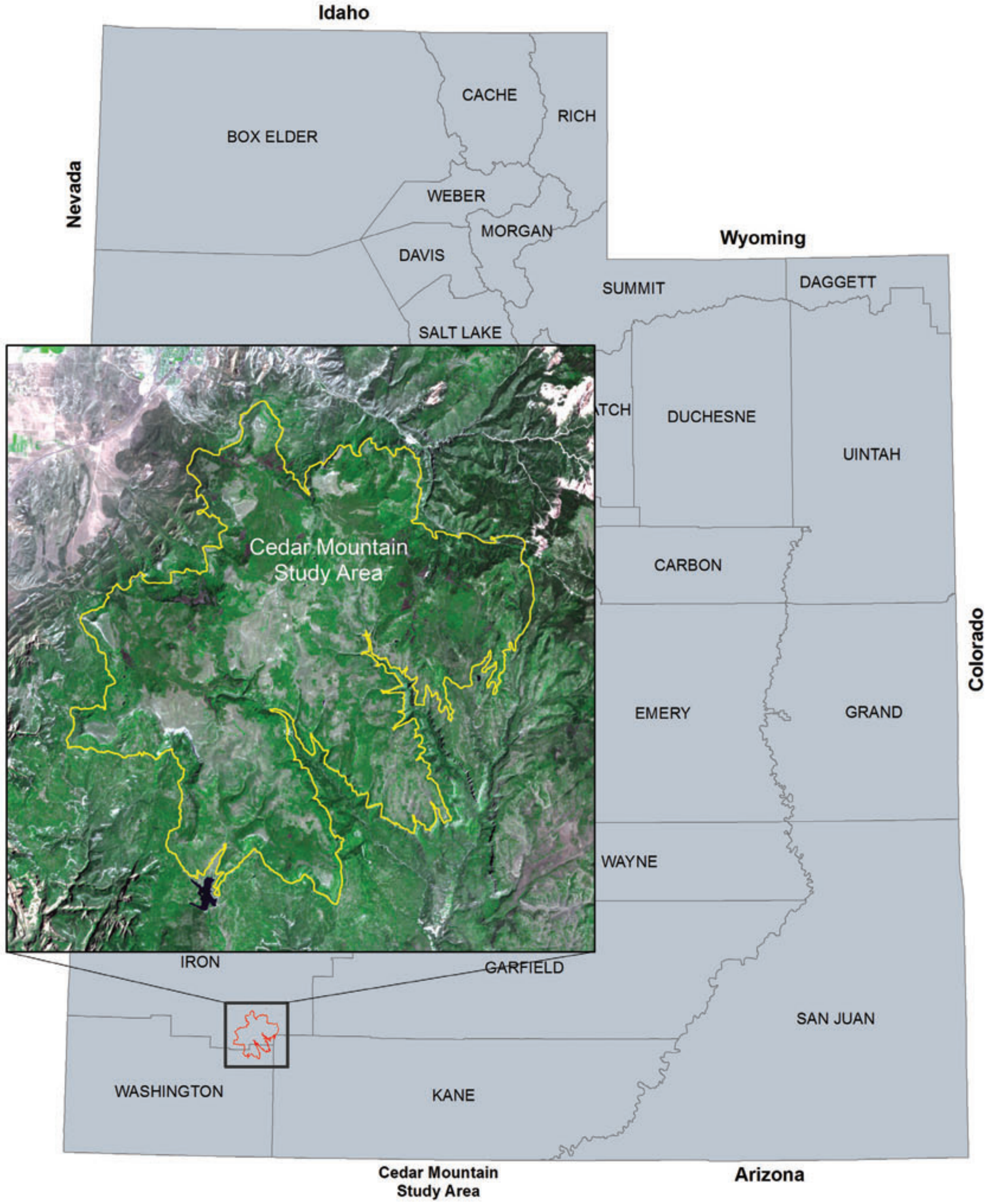


Fig. 1. The Cedar Mountain study area located near Cedar City, Utah.

systems (GIS) methods that are specifically designed to map SAD areas in the Intermountain West for management decisions have yet to be established.

To this end, we utilized remote sensing and GIS technologies to develop transferable methodologies for mapping areas experiencing aspen decline within the Great Basin and Colorado Plateau. In our analysis, we applied a classification and regression tree (CART) approach utilizing Landsat Thematic Mapper (TM) reflectance data, topographic data from digital elevation models (DEMs), and landform data to successfully map aspen stand classes on Cedar Mountain in southern Utah (fig. 1). This analysis provides land managers a spatial resource on the location and coverage of aspen stand classes for Cedar Mountain, as well as a resource that contains methods applicable to other areas experiencing aspen decline throughout the Colorado Plateau and the Great Basin.

Methods

The study area for this project is located on Cedar Mountain in southern Utah near Cedar City, Utah (fig. 1). Cedar Mountain encompasses approximately 27,216 ha of mostly privately owned land and is situated within the Kolob Terrace (2400 to 3162 m elevation)—a broad, relatively flat, lower southwestern tier of the Markagunt Plateau. Cedar Mountain contains extensive aspen mortality and is among the quintessential SAD examples within the Intermountain West.

Landsat TM imagery was selected for this project primarily due to its 25-year history of imaging the Earth. Landsat TM offers the longest running time series of systematic, remotely sensed digital imagery available. While there was no temporal component in this project, the history and systematic coverage of TM data provides the ability to easily apply techniques learned here to other areas. Secondly, the spatial resolution (grain size) of these data tends to fit the requirements for land managers. Thirdly, the spectral resolution of the Landsat TM sensor encompasses important portions of the electromagnetic spectrum (visible, near-infrared, and shortwave-infrared) that are used for vegetation mapping. Lastly, Landsat TM data are free and can be readily downloaded through the U.S. Geological Survey Global Visualizer Viewer (2008).

Landsat TM images can offer repeat coverage every 16 days. However, cloud cover and data quality tend to limit the selection of imagery. Further, phenological variation in the land cover of interest also limits imagery selection. If multiple scenes are needed for a given study area, mosaicking of adjacent Landsat scenes is required. Image standardization for solar angle illumination, instrument calibration, and atmospheric haze (i.e., path radiance) may be necessary for improved image matching.

Predictor layers used to map Cedar Mountain aspen consisted of core image-derived and ancillary datasets (Appendix A). Core image-derived datasets included individual Landsat TM spectral bands (path 34, row 38) from June 26, 2008, and the brightness, greenness, and wetness (BGW) tasseled cap transformation derived from the Landsat TM bands (Crist and Cicone 1984). Topographic ancillary datasets were extracted from 30-m DEMs obtained from the Utah Automated Geographic Reference Center (2008) and consisted of slope (in degrees), aspect (moisture index transformation), elevation (m), and a 10-class landform dataset (Manis and others 2001). The final model integrated a total of 13 predictor layers (table 1).

A key factor of this analysis was selecting aspen stand types, conditions, or classes that were practical from both a remote sensing and management perspective. From a land management standpoint, the aspen stand classification must have ecological relevance in terms of tangible management implications. Understanding this relationship is important in order to create a product that better informs decisions. From a remote sensing perspective, aspen stand types must be spectrally distinct enough to separate different stands based on their reflectance characteristics (table 2).

Table 1. The 13 predictor layers used in aspen classification model.

Model input	Band #	Description
Landsat TM 5 reflectance	1	Blue
Landsat TM 5 reflectance	2	Green
Landsat TM 5 reflectance	3	Red
Landsat TM 5 reflectance	4	NIR
Landsat TM 5 reflectance	5	MIR
Landsat TM 5 reflectance	6	MIR
Core-image derivative	7	Brightness
Core-image derivative	8	Greenness
Core-image derivative	9	Wetness
DEM	10	Elevation
DEM	11	Slope
DEM	12	Aspect
Landform	13	Landform

Table 2. Descriptions of aspen stand types used for the Cedar Mountain classification.

Healthy	Full aspen crowns with little to no die-off (<25% overstory mortality, <25% conifer cover)
Damaged	Dead or dying aspen stands with considerable to full overstory die-off and/or foliage loss (25-100% overstory mortality, <25% conifer cover)
Seral	Presence of aspen and at least 25% conifer cover within the plot

Note: Condition based on 90 x 90 m plot observation.

For Cedar Mountain, there are two dominant aspen types—persistent and seral. The persistent type can be divided into healthy and damaged (declining) stand conditions. Healthy stand types are self regenerating stands, usually producing pulses of regeneration that maintain grove size over long periods of time (Bartos 2001). Healthy aspen stands often contain a pure overstory, good stand structure (i.e., numerous age cohorts), adequate regeneration, and a diverse understory of grasses, forbs, and shrubs (Mueggler 1988; Kurzel and others 2007). These stands tend to be more resilient to disturbance (i.e., insect infestations and disease) and invasion by introduced species, and they maintain water balance between and within vegetation communities more effectively (Ryel 2004). Declining aspen stands are characterized by overstory mortality, poor stand structure, weak regeneration, and altered understory communities that weaken stand functionality. The seral stand type is characterized by the presence of aspen and conifers inhabiting the landscape simultaneously. Aspen in these systems are regarded as the early successional, disturbance, or pioneer species since they are generally the first to establish following fire, disease, or other disturbances. Although aspen may continue to persist on conifer-dominated sites (late seral) for many years, potentially centuries, eventually the more shade-tolerant conifers reestablish and begin to break up aspen canopies (Loope 1971; Schier 1975). These three aspen types are ecologically and spectrally distinct. Initially, there was interest in separating the "damaged" aspen cover class into multiple cover classes, such as "dead" or "dying." However, the performance of each of these finer classes proved too difficult to separate with acceptable accuracy. Consequently, they were combined into the "damaged" class to reduce overall error. However, creating new classes or splitting classes in other areas may be possible depending on stand characteristics.

Our objective was to classify only aspen into independent stand classes, thus, non-aspen cover was excluded. To identify only areas of aspen, high-resolution (1 m), color infrared digital orthophoto quarter quads (DOQQs) acquired from the National Agriculture Imagery Program (NAIP) (Utah Automated Geographic Reference Center 2008) were used

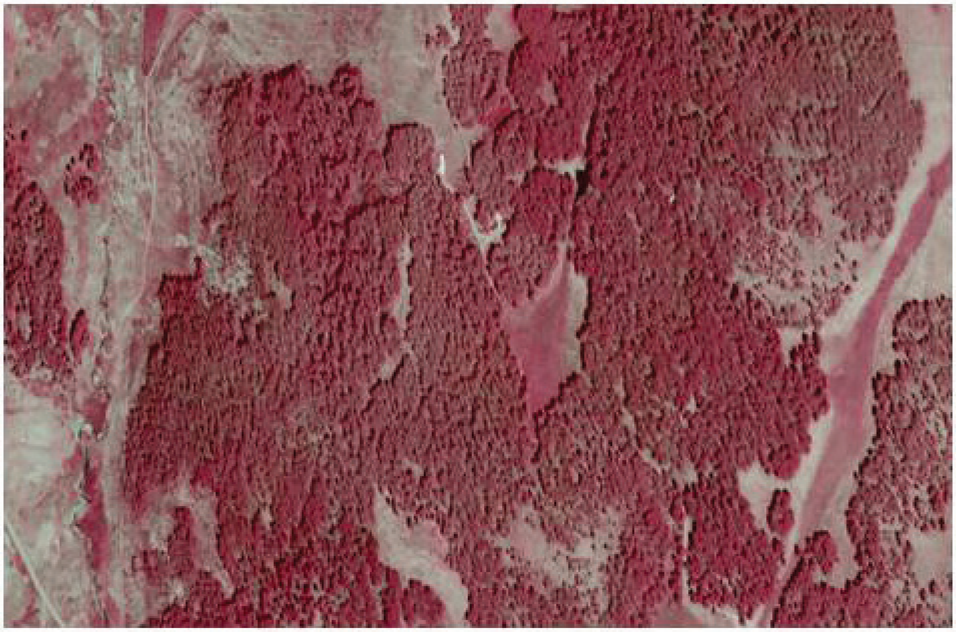


Fig. 2. Un-delineated portion of an NAIP image of the Cedar Mountain study area showing aspen.

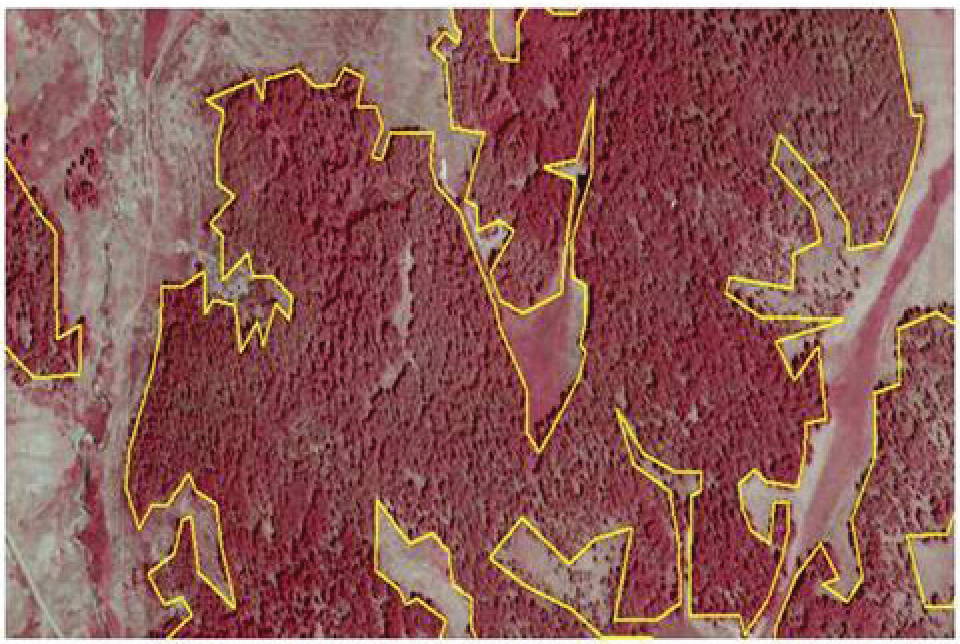


Fig. 3. Delineated portion of an NAIP image of the Cedar Mountain study area showing aspen.

to manually delineate aspen cover from non-aspen cover (see Appendix A) (figs. 2 and 3). This process created a map of aspen (including persistent and seral) with no further separation into different stand classes. The resulting aspen map was used to mask out areas that were not aspen. Subsequent digital classification of TM imagery was conducted only within these areas to further differentiate aspen from non-aspen as well as to identify the three different stand classes.

One consideration in using this technique is that preliminary ground surveying is tremendously helpful. Although the resolution of NAIP is exceptional (1 m), species with similar spectral and textural characteristics to aspen can still be difficult to separate. Ground surveys help resolve areas of uncertainty when delineating aspen stands.

For the Cedar Mountain aspen stand type classification, training and validation data were collected via ground-based field work in the summer of 2008 to match the remotely sensed data with the time period for peak aspen foliage. Ground-based reference data points consisted of 90 x 90 m (approximately 1 ha) plots that were distributed according to a 900-m systematic grid generated in ArcGIS 3.1 (Appendix B). Using the 2500-m elevation contour as the study boundary, the systematic grid was established within the aspen woodland and aspen/conifer cover classes from the 2005 Southwest Regional Gap (SWReGAP) landcover project (fig. 4). (The SWReGAP data were the best available land cover data for aspen on Cedar Mountain at the time that the sampling protocol was developed.) We did not perform

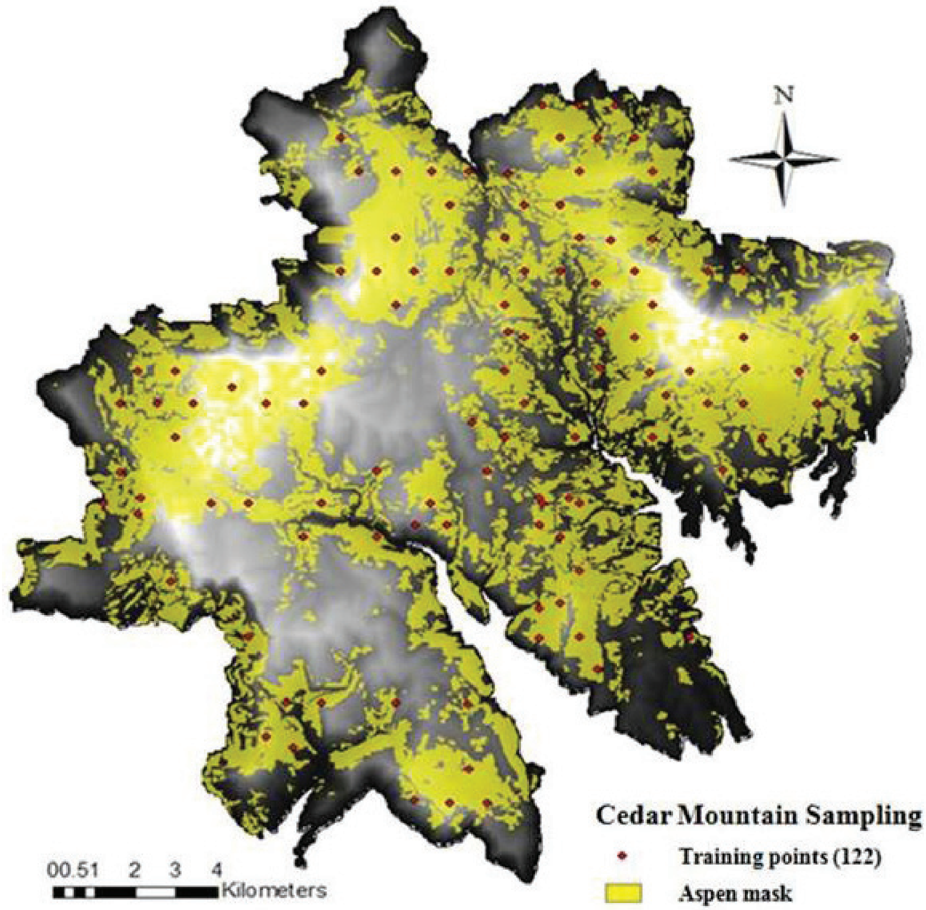


Fig. 4. Map of the Cedar Mountain study area with training data overlaying the delineated aspen cover mask.

the manual delineation of aspen cover until after the 2008 field season; therefore, this layer was not used for determining sampling points.) Each point was verified for contiguous aspen cover (at least 50% aspen or mixed aspen-conifer cover within 90 x 90 m plot) using NAIP imagery. Any point not meeting this criterion was discarded from the plot selection process.

A total of 122 training points from the systematic grid were randomly selected and visited during the 2008 summer season and assigned to one of the three aspen stand classes. At each sample point, ocular estimates of overstory canopy cover were collected for cover class designation. The designation of sample points yielded 50, 50, and 22 training points for healthy, damaged, and seral classes, respectively.

Based on previous image classification efforts using Landsat TM imagery (Reese and others 2002; Lowry and others 2007), we utilized a Classification and Regression Tree (CART) analysis to produce a spatially explicit discrete classification of aspen stand types for the Cedar Mountain area (Appendix C). Decision tree classifiers (Breiman and others, 1984) are particularly relevant for remote sensing applications as they are non-parametric classifiers, requiring no prior assumptions of normality, and they readily accept categorical and continuous datasets.

The mapping procedure utilizing the decision tree classifier is presented in fig. 5 for the Cedar Mountain aspen stand classification. Using Erdas Imagine software, the National Land-Cover Dataset mapping tool (Homer and others 2004) was used to extract values from the predictor layers at each of the training sample locations (table 1) for each aspen stand type. The training data, therefore, consisted of a data matrix of observations (rows) and variables (columns). The variables consisted of the 13 predictors extracted for each observation and the dependent variable that categorized each observation into the three aspen stand classes. The training data matrix was imported into the data-mining, decision tree software See5 (RuleQuest Research 2004). As a preliminary assessment of map quality, 20% of the available training data were withheld from the decision tree model generation and used for validation. All 13 predictor layers were used in the model (see Appendix C). The output land-cover map was compared with field photos and observations to determine accuracy. Once the final model was selected based on preliminary accuracy assessments, the final model was generated using 100% of the available training points and predictor layers. The final map was validated by a systematic 445-point grid using NAIP imagery and visually examined for general accuracy and distribution of cover classes.

The compilation of an error matrix (i.e., confusion matrix) is considered the standard form for reporting site-specific errors (Congalton 1991). An error matrix identifies the overall accuracy of the image as well as errors for each class (i.e., user and producer accuracy). The Kappa (κ) or KHAT statistic was used as a measure of agreement between model

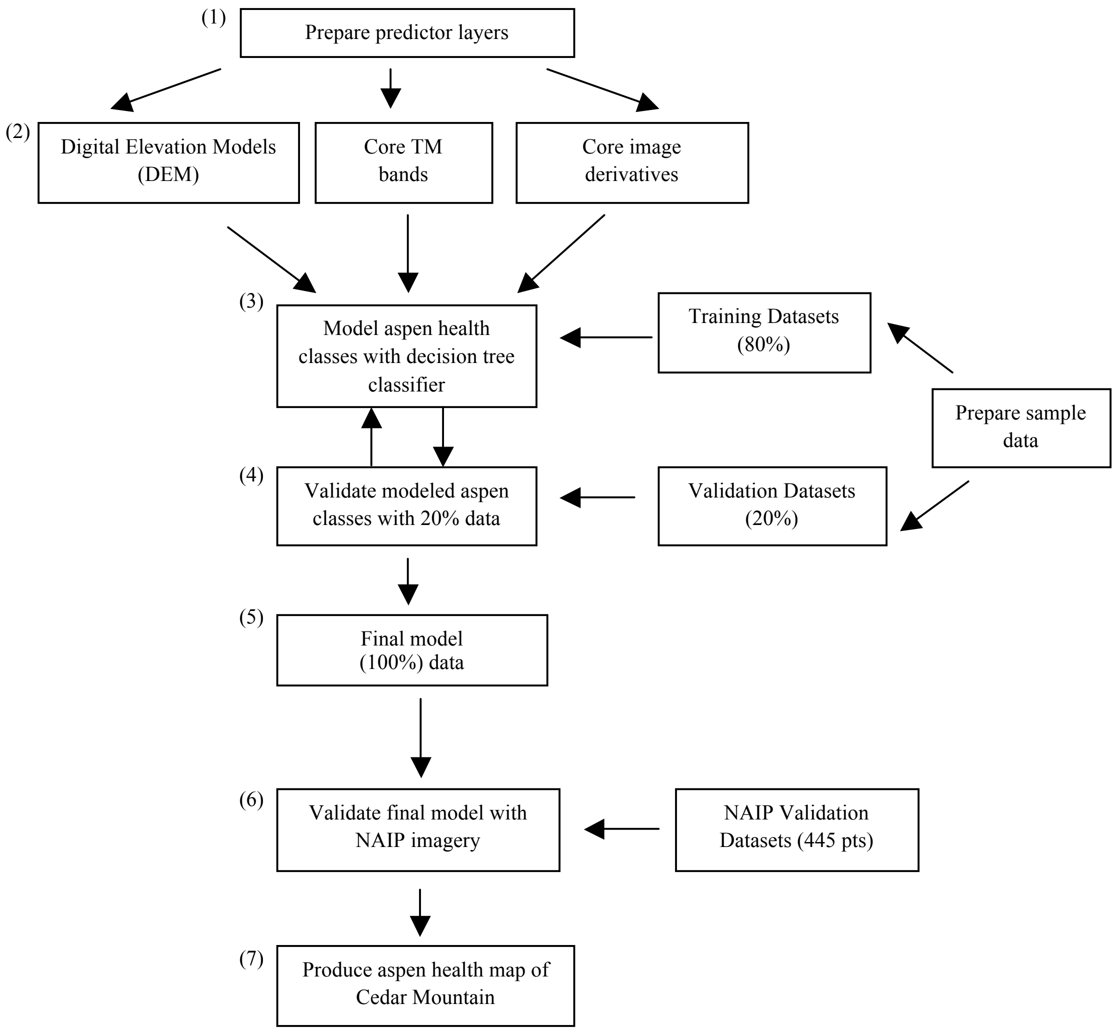


Fig. 5. Outline of the process to classify aspen stand types.

predictions and reality (Congalton 1991) or to determine if the values contained in the error matrix represent a result significantly better than random (Jensen 1996). If conducted properly (fig. 5), κ values greater than 0.80 (i.e., 80%) represent strong agreement or accuracy between the classification map and the ground reference information. κ values between 0.40 and 0.80 indicate moderate agreement and values less than 0.40 indicate poor agreement.

For the Cedar Mountain study, a rigorous systematic, site-by-site design was implemented that compared the classified image against high-resolution, color infrared NAIP imagery (reference data) for agreement. To accomplish this task, a 500-m systematic grid was produced within the NAIP-based aspen mask as a preliminary validation set (Appendix D). Each point needed to satisfy the requirement of at least a 50% canopy cover for aspen, the same requirement implemented for the initial 122 points used for model training. Any point that did not satisfy the requirement was removed from the validation set. A total of 445 points satisfied the requirement and were used as the validation/reference dataset (table 4). Once the validation set was established, each of the 445 reference points were classified into one of the three aspen stand classes based on NAIP imagery stand characteristics, field observations, and site photo points. Next, the validation points were superimposed onto the final classification and compared for agreement. From this assessment, an error matrix containing overall validation results (sum of diagonals), user and producer accuracy results, and a KHAT statistic for Cedar Mountain were reported.

Results

The final map product (fig. 6) presents the distribution of healthy, damaged, and seral aspen stand types for Cedar Mountain and retains the 30-m pixel resolution consistent with all predictor layers used in the model, with a minimum mapping unit of approximately 0.40 ha (1 acre). Healthy aspen represented the most abundant cover type with an estimated 49% (5960 ha) of the total aspen cover, followed by damaged aspen 35% (4210 ha) and seral aspen 19% (1968 ha) (table 3).

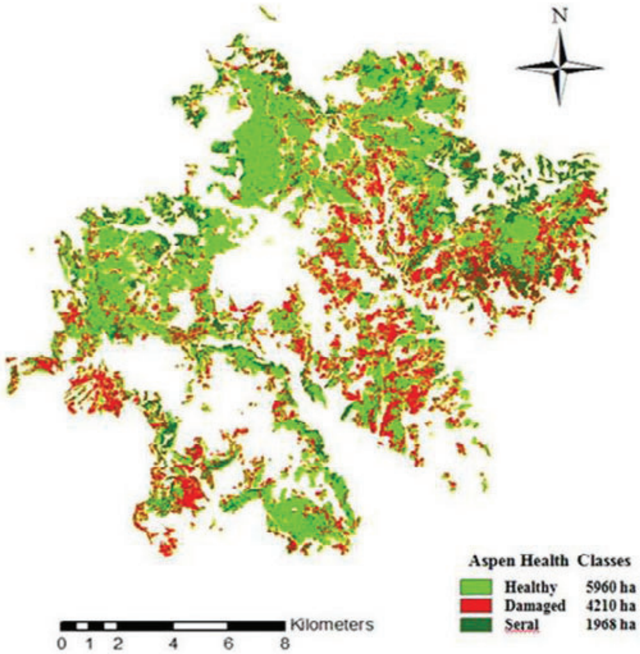


Fig. 6. Final aspen stand type classification map containing three mapped classes for the Cedar Mountain study area.

Table 3. Number of pixels and hectares as well as the percentage of the total aspen cover classified to each aspen cover class.

Aspen cover class	# of pixels	Area (ha)	% aspen cover
Healthy	66,230	5 960	49
Damaged	46,783	4 210	35
Seral	21,875	1 968	16

Table 4. Error matrix generated in See5 utilizing 20% of training/reference data. Red numbers indicate errors between classes.

	Reference data					
	Healthy	**Damaged**	**Seral**	**Totals**	**UA%**	**CE %**
Healthy	9	0	2	11	81.8	18.2
Damaged	2	6	1	9	66.7	33.3
Seral	1	0	6	7	85.7	14.3
Totals	12	6	9	27		
PA%	75.0	100.0	66.7			
EO%	25.0	0.0	33.3			
	errors	**%**	**Overall accuracy**			
Overall error	6	22.3%	77.8%			

Note: UA, user's accuracy; PA, producer's accuracy; EO, errors of omission; CE, errors of commission.

As previously stated, numerous assessments for goodness of fit (i.e., error matrices) were conducted within the See5 software that utilized 20% of the reference data to gauge the effectiveness of the model. This procedure was repeated for each model that was produced until the model of choice was selected based on the best validation results. The See5 generated error matrix for the final model is presented in table 4.

The final model selected utilized all 13 predictor datasets. Subsequently, the final model was re-run utilizing 100% of the training data. The final model was validated using a 445-point, NAIP-based independent accuracy assessment. The overall map accuracy for the final model was 81.3% with a Kappa of 69% (table 5). Healthy aspen stands had the highest user and producer accuracy (86.3% and 83.3%, respectively). Healthy stands represented the most abundant aspen cover type in the map at 49%. From a user's perspective, healthy aspen stands were most often confused with damaged stands (approximately 10% error rate with damaged stands, and approximately 3% error rate with seral stands). The damaged aspen cover class received the next highest user and producer accuracy measure (77.4% and 80.0%, respectively). The user accuracy measures suggest that damaged cover classes were most often confused with healthy stands (18%), and less so with seral stands (4.5%). The confusion between healthy and damaged stands was expected given the ambiguous nature associated with classifying the aspen health gradient. Lastly, the seral cover type received the lowest user and producer accuracy measures (73.4% and 77%, respectively). These findings may largely be due to the paucity of training points (22). Seral stands represented a small portion of the Cedar Mountain landscape (16%), yet they exhibited a wide array of stand characteristics that proved difficult to represent in the model.

Discussion

While remote sensing scientists have been utilizing satellite-based sensors to map land cover for over 30 years, it has not been until recently that efforts have been made to map aspen ecosystems; specifically, with regard to aspen decline. Increased awareness of aspen decline has increased the interest in aspen, and advances in remote sensing techniques now make it more feasible to assess aspen ecosystems. In this study, remote sensing and GIS methods were developed to determine the extent of aspen decline which, in turn, should help the land managers evaluate landscapes for restoration purposes. Cedar Mountain was chosen as the study area as this area exhibits considerable loss of aspen cover in the past decade. Although the methodology discussed in this report was successful on Cedar Mountain, it is only one of many viable approaches. Other potential classification methods are an unsupervised classification (non rule-based) (Jensen

Table 5. Error matrix for classification of aspen stands on Cedar Mountain. Red numbers indicate errors between classes.

	Reference data					
	Healthy	**Damaged**	**Seral**	**Totals**	**UA%**	**CE %**
Healthy	195	24	7	226	86.3	13.7
Damaged	28	120	7	155	77.4	22.6
Seral	11	6	47	64	73.4	26.6
Totals	234	150	61	362		
PA%	83.3	80.0	77.0			
EO%	16.7	20.0	23.0			
	errors	**%**	**Overall accuracy**			**KHAT**
Overall error	83	18.7%	81.3%			69%

Note: UA, user's accuracy; PA, producer's accuracy; EO, errors of omission; CE, errors of commission.

1996), classification based on spectral mixture analysis (Small 2001), hybrid unsupervised-supervised classifications, stratification regression models (Pereira and Itami 1991), and random forests classification (Gislason and others 2006). Time, cost, analytical skill, and objectives need to be considered when choosing a classification method. Analytically, the key point to consider is that all land cover classes should be as homogeneous as possible to increase accuracy of classification. More importantly, reducing variability within aspen cover classes will enhance the applicability of the product when used to locate sites.

In this study, an important objective was to develop mapping methods with procedures that could be transferred and independently applied to other areas experiencing aspen decline. In this study, the CART approach was found to be time-efficient and straight forward, making it a transferable option. The output provides a spatial resource that meets the needs of land managers for Cedar Mountain, but it should also work for land managers in other areas with aspen decline.

Although the layers that were selected (table 1) for the final CART analysis produced an effective model to map aspen stand classes for Cedar Mountain, this selection is not universally applicable. Datasets that may also be useful are the Normalized Difference Vegetation Index, various soil datasets, and solar radiation derivatives. To explore options further, the application of a Random Forest analysis (Breiman 2001) can be an informative means to examine the relative importance of various datasets or which predictor layers explain the most variability in the analysis. Selecting datasets that contribute the greatest predictive power to the model may produce the best results. We experimented with various combinations of the predictor data in an effort to create a more parsimonious (simple) and accurate model.

An additional factor to consider in a classification is the use of core-image (Landsat TM) derivatives and ancillary datasets. For this project, the addition of core-image derivatives (BGW) and ancillary datasets (slope, aspect, and elevation) to supplement core-image multispectral data (Landsat TM) was found to increase classification accuracy. In general, classification techniques using both spectral and ancillary data lead to greater overall accuracy, precision, and class distinctions (Trotter 1991; Jensen 1996; Lowry and others 2007).

Lastly, studies have shown that the addition of multi-seasonal imagery can increase the power to discriminate between pure aspen and aspen/conifer (seral) classes (Heide 2002; Lowry and others 2007). Special considerations when implementing this option are the cost of imagery, availability of multiple dates of cloud free imagery, inherent effects of elevational gradients and phenology, and snow cover on the imagery. At the time of the Cedar Mountain aspen stand classification, limited funding and lack of cloud free imagery prevented this project from including multi-seasonal imagery. However, multi-seasonal imagery can capture phenological differences in aspen stands throughout a growing season, potentially providing a valuable dataset to improve the overall product.

One challenge of mapping any natural landscape with remotely sensed imagery is the large spectral, environmental, and biological diversity that typifies many areas. Successful mapping often entails collecting a substantial number of field samples to properly train and validate output maps. Mapping aspen decline and conifer encroachment into individual cover types is similarly difficult in terms of collecting enough training samples to account for the complex gradient of health and diversity found in aspen stands. A purposive collection of training samples can be effective; however, it is not a valid way to assess accuracy.

Sample designs for reference data collection vary considerably. Systematic grids, and systematic grids with a random sub-sampling of grid points like the ones used in this study, provide an objective acquisition of samples across the landscape; however, they often under sample rare cover classes. Completely random sample designs can also be implemented that are statistically defensible but that also tend to under sample rare cover classes. Hybrid designs that integrate systematic, random, and stratified designs all exist. In general, random or stratified random sample designs produce the best results for remote sensing purposes (Congalton 1988a).

Ground-based sample data collection generally provides the most reliable option to reduce potential confusion but is often expensive and time consuming. If funding is limited, utilizing NAIP imagery or another high-resolution imagery source (e.g., Google Earth) as a surrogate for ground-truthing efforts can yield equally successful, but maybe not as accurate, results at a fraction of the cost. This is particularly helpful in the development of independent validation datasets (i.e., sets that are separate from the training data), and it allows for the complete utilization of training data for model development (Congalton 1988b). This option, as implemented in the Cedar Mountain application, was cost-effective and provided a thorough means of validating both abundant and rare cover classes as well as the environmental gradient between the two classes. If possible, independent validation datasets should be an integral component of all remote sensing classifications.

Finally, the task of designing and acquiring unbiased training samples and independent accuracy assessment datasets for most land cover mapping projects is difficult to achieve, especially in a project that is focused on characterizing stand classes within aspen. Improvements to the design and methodologies would contain a sample design that addresses the variation present in the individual, including rare or diverse aspen stands (e.g., damaged and seral aspen). For this project, a random selection within a stratified sample design based on cover types would likely have increased the performance and accuracy of the model.

The cost to classify an aspen-dominated landscape by aspen stand type varies depending on the availability and quality of imagery, remote sensing analyst skill level, computer resources and software licenses, and the level of precision needed to meet project objectives. Landsat TM data and NAIP imagery for Utah are free of charge to the end user. Taxpayers assume the costs of these data. Other forms of high-resolution aerial photography or satellite imagery may or may not have additional fees. If no field-based efforts are implemented, the quality of NAIP imagery must be sufficient to allow photo interpretation or identification of aspen stand type based on the classification scheme. If a field season is included, wages for one to two technicians for approximately one to two months should be expected, depending on the size of the study area. Also, depending on skill level and time spent testing and developing methodologies, wages for a remote sensing analyst and possibly a photo interpreter should be expected for one to three months. Excluding experimentation with methodologies and techniques for this type of classification, an experienced remote sensing analyst and photo interpreter could complete a similar project in size and scope, including a field season, in approximately six to eight months.

Conclusion

The objective of this study was to develop transferable remote sensing and GIS methodologies and techniques used to map areas in the Intermountain West experiencing aspen decline. High-resolution aerial photography (NAIP imagery), multispectral satellite imagery (Landsat TM), core-image derivatives, and ancillary datasets were used in a CART analysis that successfully mapped three aspen stand types for Cedar Mountain in southern Utah with an overall accuracy of 81.3% using an NAIP-based independent accuracy assessment. To this end, this report can serve land and natural resource managers as a technical guide for using remote sensing and GIS technologies for aspen monitoring and restoration activities.

References

Baker, F.S. 1925. Aspen in the central Rocky Mountain Region. Bulletin 1291. Washington, DC: U.S. Department of Agriculture. 47 p.

Bartos, D.L.; Shepperd, W.D., comps. 2010. The aspen mortality summit; December 18-19, 2006; Salt Lake City, UT. Proc. RMRS-P-60www. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 21 p. Available: http://www.fs.fed.us/rm/pubs/rmrs_p060.html.

Bartos, D.L. 2001. Landscape dynamics of aspen and conifer forests. In: Shepperd,W.D.; Binkley, D.; Bartos, D.L.; Stohlgren, T.J.; Eskew, L.G., comps. Sustaining aspen in western landscapes: symposium proceedings; June 13-15, 2000; Grand Junction, CO. RMRS-P-18. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 5-14.

Bartos, D.L.; Campbell, R.B. 1998. Decline of quaking aspen in the Interior West—examples from Utah. Rangelands. 20:17-24.

Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. 1984. Classification and regression trees. Belmont, CA: Wadsworth, Inc. 358 p.

Breiman, L. 2001. Random forests. Machine Learning. 45:5-32.

Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment. 37:35-46.

Congalton, R.G. 1988a. A comparison of sampling schemes used in generating error matrices for assessing the accuracy maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing. 54:593-600.

Congalton, R.G. 1988b. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing. 54:587-592.

Crist, E.P.; Cicone, R.C. 1984. Application of the tasseled cap to simulate thematic mapper data. Photogrammetric Engineering and Remote Sensing. 50:343-352.

Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. 2006. Random forests for land cover classification. Pattern Recognition Letters. 27:294-300.

Heide, S.C. 2002. Comparison of methods to detect conifer encroachment into aspen stands using Landsat 7 ETM+ satellite imagery. Thesis. Moscow, ID: University of Idaho. 84 p.

Hogg, E.H.; Schwarz, A.G. 1999. Tree-ring analysis of declining aspen stands in west central Saskatchewan. Edmonton, Alberta: Canadian Forest Service, Northern Forestry Centre. Inf. Rep. NOR-X-359.

Homer, C.; Huang, C.; Yang, L.; Wylie, B.; Coan, M. 2004. Development of a 2001 national land-cover database for the United States. Photogrammetric Engineering and Remote Sensing. 70(7):829-840.

Jensen, J.R. 1996. Introduction to digital image processing of remotely sensed data. In: Jensen, J.R., ed. Introductory digital image processing: a remote sensing perspective. Prentice-Hall, Inc., New Jersey: 1-2.

Kurzel, B.P.; Veblen, T.T.; Kulakowski, D. 2007. A typology of stand structure and dynamics of quaking aspen in northwestern Colorado. Forest Ecology and Management. 252:176-190.

LaMalfa, E.M.; Ryle, R. 2008. Differential snowpack accumulation and water dynamics in aspen and conifer communities: implications for water yield and ecosystem function. Ecosystems. 11:569-581.

Loope, L.L. 1971. Dynamics of forest communities in Grand Teton National Park. Naturalist. 22:39-47.

Lowry, J.; Ramsey, R.D.; Thomas, K.; and others. 2007. Mapping moderate-scale land-cover over very large geographic areas within a collaborative framework: a case study of the southwest Regional Gap Analysis Project (SWReGAP). Remote Sensing of Environment. 108: 59-73.

Manis, G.; Lowry, J.; Ramsey, R.D. 2001. Preclassification: an ecologically predictive land form model. Gap Analysis Program Bulletin No. 10, December 2001. Moscow, ID: U.S. Geological Survey, GAP Analysis Program. Available: www.gapanalysisprogram.com/bulletins/10/preclassification.htm.

Mueggler, W.F. 1988. Aspen community types of the Intermountain Region. GTR-INT-250. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 135 p.

Pereira, J.M.C.; Itami, R.M. 1991. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. Photogrammetric Engineering and Remote Sensing. 57:1475-1486.

Preston, R.J. 1976. North American trees. Iowa State University Press, Ames. 399 p.

Reese, H.M.; Lillesand, T.M.; Nagel, D.E.; Stewart, J.S.; Goldmann, R.A.; Simmons, T.E. 2002. Statewide land-cover derived from multiseasonal Landsat TM data: a retrospective of the WISCLAND project. Remote Sensing of Environment. 82:224-237.

RuleQuest Research. 2004. [Online.] Available: http://www.rulequest.com/index.html. [Accessed October 10, 2007].

Ryel, R.J. 2004. Hydraulic redistribution. Progress in Botany. 65:413-435.

Schier, G.A. 1975. Deterioration of aspen clones in the middle Rocky Mountains. RP-INT-170. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 14 p.

Sexton, J.O.; Ramsey, R.D.; Bartos, D.L. 2006. Habitone analysis of quaking aspen in the Utah Book Cliffs: effects of site water demand and conifer cover. Ecological Modeling. 198:301-311.

Sinclair, W.A.; Lyon, H.H. 2005. Diseases of trees and shrubs. Cornell University Press, Ithaca, New York. 660 p.

Small, C. 2001. Estimation of urban vegetation abundance by spectral mixture analysis. International Journal of Remote Sensing. 22:1305-1334.

Strand, E.K.; Vierling, L.A.; Bunting, S.C.; Gessler, P.E. 2009. Quantifying successional rates in western aspen woodlands: current condition, future predictions. Forest Ecology and Management. 257:1705-1715.

Trotter, C.M. 1991. Remotely-sensed data as an information source for geographical information systems in natural resource management: a review. International Journal of Geographic Information Systems. 5:225-239.

USGS Global Visualization Viewer. 2008. Landsat Science: MRLC/MTBS Reflectance. Available: http://glovis.usgs.gov/ImgViewer/Java2ImgViewer.html.

Utah Automated Geographic Reference Center. 2008. 2006 NAIP 1 Meter Color Orthophotography. Available: http://gis.utah.gov/naip2006.

Worrall, J.J.; Egeland, L.; Eager, T.; Mask, R.A.; Johnson, E.W.; Kemp, P.A.; Shepperd, W.D. 2008. Rapid mortality of *Populus tremuloides* in southwestern Colorado, USA. Forest Ecology and Management. 255:686-696.

Appendix A. Predictor Layer Preparation

Core-image datasets

For Cedar Mountain, only one Landsat Thematic Mapper (TM) scene was needed to cover the study area. Landsat TM reflectance data (path 34, row 38) was acquired for June 26, 2008. The image was reprojected using the North American Datum 1983 Universal Transverse Mercator (UTM) Zone 12N and nearest neighbor resampling intensity. The following steps were taken to download and prepare Landsat TM imagery layers:

a) To acquire Landsat TM datasets, access USGS Global Visualization Viewer at http://glovis.usgs.gov/.

b) Select "Landsat MRLC" from first drop down box, then select the "MRLC/MTBC reflectance" option from the second drop down box. This imagery contains reflectance values that have been pre-processed for atmospheric corrections and can be used without further radiometric enhancement.

c) Next, click the collection button, then select "Landsat Archive" and choose the "Landsat 4-5 TM" imagery option.

d) Locate study area by either entering coordinates or panning map with cursor. Scan imagery for clarity and cloud free dates during peak growing months (June to August) and select the best scenes.

e) Once imagery is selected, click the "Add" button, followed by the "Download" button. Users will be asked to register if they are not already. Click "Start Download."

f) Download and save file(s) in a folder. After downloading, extract (unzip) files.

g) Open Erdas Imagine 9.1. In a viewer, open file(s) containing the imagery (select TIFF file type), and select the reflectance file (i.e.,_refl.tif.)

h) Click the "Dataprep" button on the main toolbar in Erdas. Select "Reproject Image." Select "Datum" and "Projection" for study area. Use 30 x 30 cell size and select the nearest neighbor resampling method.

i) Repeat steps "g" and "h" for all predictor layers.

Image derived datasets

a) In the same file that Landsat TM data was stored and extracted, there are three to four other files that can be selected. The _tc.tif file contains the Brightness, Greenness, and Wetness (BGW) tasseled cap transformation. Select and repeat steps "g" and "h."

Digital Elevation Model (DEM) derived datasets

a) To acquire 30-m DEMs for Utah, access the Utah Geographic Information System (GIS) Portal website (http://agrc.its.state.ut.us/). Click "Gis data" link on the upper tool bar. Next, select "Download data for SGID" → "Elevation/Terrain" → "10, 30, 90 DEM "→ "30m DEM" → "Statewide (zip) 30_m_DEM.zip" and "download." Extract files once downloaded.

b) In Erdas viewer, add the DEM dataset for study area. To derive slope and aspect datasets from the DEMs, select "Interpreter" from the main toolbar, then "Topographic analysis," and select either "Slope" or "Aspect." Other datasets are also available to produce from the list.

c) Next, select the DEM file as the input file, then name and save the output file in a folder (i.e., predictor layers). In the same box, select "Degrees" or "Percent," then click "OK."

d) For ArcGIS, select "ArcToolbox" → "Spatial Analyst" → "Surface" → "Aspect," "Slope," etc.

e) Resample and reproject datasets following steps "g" and "h" in the first section.

Summary of predictor layers

a) Multi band predictors:

1. Landsat TM 5 reflectance (bands 1 through 5 and 7 for Summer 2008)

b) Single band predictors:

1. 2008 Summer Brightness
2. 2008 Summer Greenness
3. 2008 Summer Brightness
4. Elevation—continuous (integer)
5. Aspect—continuous moisture index (integer)
6. Slope—continuous (integer)
7. Landform—categorical 9 class

Acquisition and delineation of aspen cover using National Agriculture Imagery Program (NAIP) imagery

a) For study sites in Utah, access available NAIP imagery from the Utah GIS Portal http://agrc.its.state.ut.us/ web site. Click GIS Data link on the upper tool bar. Next, select Aerial Imagery, then the year and type of aerial photography that meets project objectives. Download the selected imagery. This will bring up a list of available zip files. Select files that cover the study area and place them into a folder. Extract/unzip those files.

b) Open ArcGIS 9.2. Click the "Add data" button and select the NAIP files. Determine appropriate projection. Pyramids may need to be constructed. (Note: Cedar Mountain was North American Datum 1983 UTM Zone 12N).

c) A shapefile needs to be created in order to begin digitizing. Select the ArcCatalog button. Find or create a folder. In the folder, right click and select "Shapefile." Name shapefile (e.g., aspen mask). Select "Polygon" as the feature type. Click "Edit." To properly project the shapefile, select "Import" and select a form of imagery that has previously been reprojected. Click "OK." Click "OK" again. Close ArcCatalog.

d) In ArcMap, set the scale around the 1:6000 range for optimal resolution. This will depend on the quality and resolution of the imagery. Next, click the "Add" button and select the newly created shapefile (i.e., aspen mask).

e) To begin editing (digitizing), click the "Editor toolbar" button → "Editor Drop Down" → start editing. In the pop-up box, select the aspen mask shapefile. In the Editor bar, make sure the Task is "create a new feature" and the target is the "aspen mask" shapefile. Click "OK."

f) Next, in the Editor toolbar, click the pencil symbol and begin delineating (digitizing) all aspen cover from non-aspen cover.

g) Periodically, save your edits. To do this, in Editor toolbar, click the "Editor Drop Down" box and select "Stop editing" and save. Continue until finished.

h) For modeling purposes, converting the delineated aspen mask .shp file to an .img file will be necessary. In ArcToolbox, select "Conversion Tools" → "To Raster" → "Polygon to Raster." Select the aspen mask .shp file for input features, use Cell_Center for the cell assignment type, and select "30" for the cell size. All other specifications use default settings.

i) In order to properly apply the aspen mask, all of the individual polygons that compose the shapefile need to be combined into one. Open Raster Calculator. Select aspen mask .img file, multiply by "0" and add "1." The formula should be as follows: [aspen_mask_.img] * 0 + 1

j) The Raster calculation will provide a temporary layer named "Calculation" in the layers column. This needs to be made permanent. Right-click the temporary layer → "Data" → "Export Data." This brings up the Export Raster Data box. In this box, make sure a 30 x 30 cell size is selected and the output file (format) is an IMAGINE Image. Name the file, select output folder, and select "Save." This file is now ready to be used as a mask in the modeling procedures.

Area of Interest (AOI) and subsetting procedures

All operations are conducted in Erdas Imagine 9.1.

a) Display imagery containing study area in viewer. Select "File" → "New" → "AOI layer." Next, select "AOI" → "Tools." A toolbox will appear. Utilize tools/options to create broad boundary of study area. (Note: Boundary will be used to subset all other predictor layers. The NAIP-based aspen delineation will be used specifically in the CART model to reduce non-aspen cover.)

b) Once an AOI is established, save it in a folder. Next, click the "DataPrep" button on the main toolbar and select subset image.

c) Choose the reprojected imagery (.img) files as your input file. Create an output file name and place in a convenient folder. Next, select the AOI button/box at the bottom. Choose viewer. Select "OK."

d) Repeat procedure for all predictor layers used in the model.

Table A-1. Criteria used to distinguish aspen stand cover classes for the Cedar Mountain study area.

Healthy	Full crowns with little to no die-off (<25% overstory mortality)
Damaged	Consisted of dead and dying stands with considerable to full overstory die-off and/or foliage loss (25 to 100% overstory mortality)
Seral	Presence of aspen and at least 25% conifer cover within the plot

Note: Condition based on 90 x 90 m plot observation.

Table A-2. Example of a portion of an Excel spreadsheet used to convert sample point data into a .txt file. The first two columns are composed of coordinates and the third is the site stand type. This file is used as the dependant variable in the decision tree classifier.

	A	B	C
1	323017	4147207	1
2	319867	4149907	1
3	321667	4149907	1
4	326617	4150807	1
5	326167	4151707	1
6	328867	4151707	1
7	325717	4152607	1
8	321217	4154407	1
9	317167	4155307	1
10	318067	4155307	1
11	322531	4155314	1
12	325267	4155307	1
13	326167	4155307	1
14	316267	4157107	1
15	327967	4157107	1
16	315817	4158007	1

Table A-3. Portion of the attribute table for NAIP validation points indicating the "ID" or map cover class and the NAIP-based cover class.

FID	Shape *	Id	NAIP
0	Point	1	3
1	Point	1	1
2	Point	1	2
3	Point	1	3
4	Point	1	2
5	Point	1	2
6	Point	1	1
7	Point	1	1
8	Point	1	3
9	Point	1	1
10	Point	1	1
11	Point	1	1
12	Point	0	1
13	Point	1	1
14	Point	1	1
15	Point	1	3
16	Point	1	1
17	Point	1	1
18	Point	0	3
19	Point	0	1

Table A-4. Error matrix for classification of aspen stands on Cedar Mountain. Red numbers indicate errors between classes.

	Reference data					
	Healthy	Damaged	Seral	Totals	UA%	CE %
Healthy	195	24	7	226	86.3	13.7
Damaged	28	120	7	155	77.4	22.6
Seral	11	6	47	64	73.4	26.6
Totals	234	150	61	**362**		
PA%	83.3	80.0	77.0			
EO%	16.7	20.0	23.0			
	errors	**%**	**Overall accuracy**			**KHAT**
Overall error	83	18.7%	**81.3%**			69%

Note: UA, user's accuracy; PA, producer's accuracy; EO, errors of omission; EC, errors of commission.

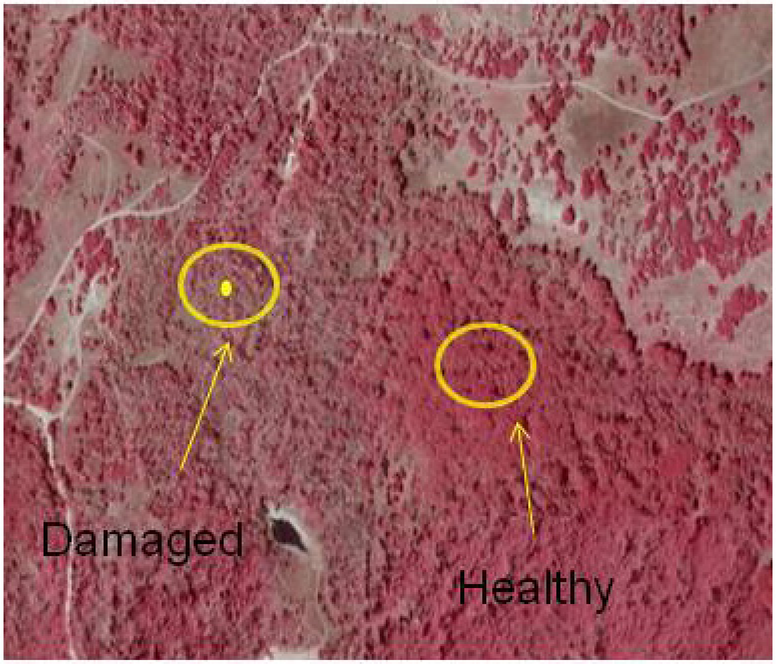


Fig. A-1. An example of deciphering aspen health stand classes based on NAIP canopy characteristics.

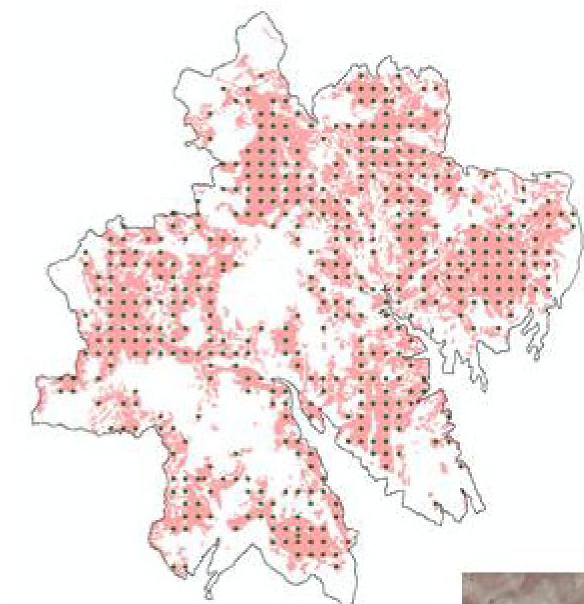


Fig. A-2. Map showing 500-m grid (446 points) generated in the delineated aspen mask that is used to validate the final aspen stand classification map for Cedar Mountain.

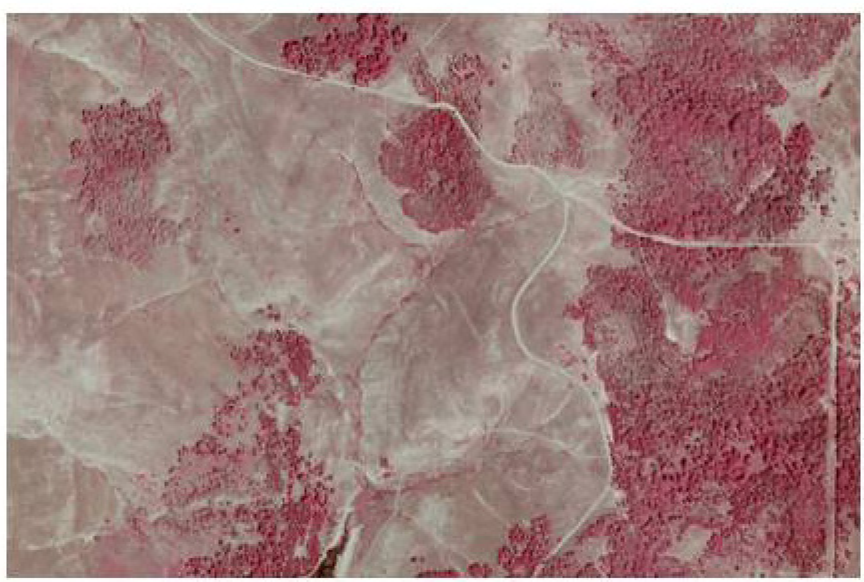


Fig. A-3. NAIP imagery used during the validation process.

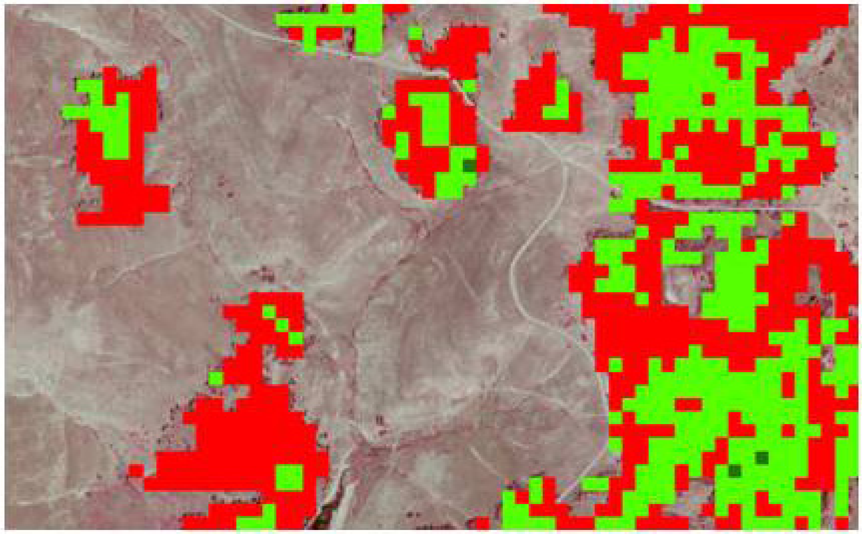


Fig. A-4. Same NAIP image as the base layer (fig. A-3) with the aspen stand type map overlain. Comparing the model map and the NAIP imagery was the core procedure used to generate accuracy assessments for the Cedar Mountain application.

Appendix B. Sample Point Generation

Sample point generation

a) A systematic grid can be established within the delineated aspen mask using Hawth's Tools. Download and install Hawth's Tools from http://www.spatialecology.com/htools/. This free extension for ArcMap provides a suite of useful sampling tools, including a tool to create systematic grids of user specified size. Once Hawth's Tools is installed, a new toolbar should appear in ArcMap. If it does not, check if the extension is active ("Tools" → "Extensions….") and make sure the toolbar is visible ("View" → "Toolbars").

b) Determine the desired grid size (Note: the Cedar Mountain application used a 900 m grid). In Hawth's tools, select "Sampling Tools" → "Generate Regular Points." Select aspen mask layer for the extent. Specify point spacing. Select either alignment or alternating rows and the output shapefile and folder. Click "OK."

c) Next, the points within the aspen mask need to be extracted. Click "Selection" → "Select by location" → "Grid generated" → "are completely within" → aspen mask .shp file. (Note: This file needs to be a vector file.) Select "Apply." This will select all points within the aspen mask.

d) These selected points need to be extracted. Right-click the grid layer and select "Data" →" Export Data." In the pop-up box, make sure "selected features" is in export box and that the coordinate system is the same as this layer. Give the file a name and click "OK."

e) Next, the points need coordinates for both modeling and field purposes (optional). In Hawth's Tools, select "Table Tools" → "Add XY" to Table (points). In the pop-up box, select the grid file just produced, click the Add new fields tab and enter names for the X and Y field (e.g., "X_coord" and "Y_coord"). Select the same Coordinate System as the layer's source data. Click "OK."

f) Open the attribute table for the grid data layer and examine the data. Notice the ID column contains only zeros for all points. For identification purposes, start the Editor (this allows the user to edit data in the attribute table), right-click the "ID" column and select Field Calculator. In the Fields box, select "Field identifier (FID)." This will add it to the ID = box. Once entered, add "+ 1" to "FID." The command should read: [FID] + 1. This will assign an ID number to each sample point, starting with 1. Save edits and stop editing.

g) Optional: Verify selected sample points for pure and/or seral aspen cover classes with National Agriculture Imagery Program (NAIP) imagery. Cedar Mountain maintained a requirement of at least 75% pure and/or seral aspen cover for a 90 x 90 m plot. (Note: Sample point was the center of plot.)

Sample point designation (cover classes)

a) Each sample point that was ground-truthed was assigned to one of three aspen stand classes based primarily on crown and overstory. On Cedar Mountain, each site was assigned to a cover class based on the criteria in table A-1.

b) Additional points were selected (based on NAIP imagery) and a classification was implemented after initial model classification to increase sample size for seral aspen. When using NAIP imagery for designation procedures, prior visitation of study area aids considerably in discriminating between different cover classes. Once the photo interpreter is trained, deciphering the three aspen stand classes is straightforward (fig. A-1).

Summary of samples

a) Ninety-four samples were visited on-site and classified to one of the three classes during 2008. An additional 28 points were acquired and designated using NAIP imagery. Fifty sample points were assigned to both "Healthy" and "Damaged" aspen stand classes, while 22 sampling points were assigned to the "Seral" aspen stand class for a total of 122 sample points used in the final model.

b) Once all sample points are assigned a cover class, this file needs to be converted to a .txt file for use in the National Land-Cover Dataset mapping tool. To do this, open the .dbf file for the sample points in a Microsoft Excel spreadsheet. Data should only consist of X and Y coordinates and cover class designation (see table A-2). Thus, removing the site ID and column titles from the spreadsheet is necessary. Save as a .txt file (table A-2).

Appendix C. Modeling

Introduction

Software and extension tools used in this application are ArcGIS 9.2, Erdas Imagine 9.1, See5 data mining software, and the National Land-Cover Dataset (NLCD) mapping tool (Homer and others 2004). These tools were used in conjunction to conduct the Classification and Regression Tree (CART) analysis using the 13 selected predictor layers for the model.

Procedures

a) Download and install the NLCD mapping tool from www.mrlc.gov. This is an Erdas Imagine extension that interfaces with See5 data mining software to conduct CART analyses.

b) In Erdas, open the NLCD mapping tool and select the NLCD mapping Tool. In the Independent Variables box, select all available predictor layers (i.e., Landsat TM data, elevation, landform, etc.) for all the potential models. These datasets should be all in one folder for easy access (e.g., Predictor layers). For each model, select any combination of predictor layers and add them to the Independent File List. These datasets will be used to train this particular model.

c) In the Dependant Variable (.txt) box, select the sample point .txt file generated from the grid (Appendix B). Accept the 255 default values for the Ignore Values box. Under Sampling number, select "percent." Enter 80% for training and 20% for validation. This ratio can be altered. Select "random" for sampling method. Name the output name file (.names file). Select See5. Click "OK."

d) Open See5 data mining software. Select "File" → "Locate Data." Browse files for data file produced by the NLCD mapping tool. This file should be in the same folder as the .names file.

e) Click the "Classifier Construction Options" (second button from the left). This provides many options to manipulate the decision tree. For the Cedar Mountain applications, Boost was selected and 15 trials were employed, and the Global pruning box was unchecked. Experiment with the options. Select" OK" when finished.

f) A results box from the decision tree will appear. The .test data (i.e., error matrices) are presented at the bottom. Examine the results for overall performance.

g) Next, in the NLCD mapping tool, choose the "See5 Classifier" button. Select the generated .names file (step c). Select "tree" and the .tree file will automatically be entered. Select the NAIP based aspen mask .img file (i.e., aspen_mask.img) for the mask option. Lastly, name the output file (.img) and select a folder to store it in (all model output files should be stored in this folder). Check the Create Error or Confidence Layer and select "OK."

h) Open the output file (.img) in a new Erdas viewer to view the map of the generated model. Examine the map for general appearance and accuracy. Repeat steps "b" through "h" for each model until the best model is selected. Once a model is selected for the final product, repeat the modeling procedures using 100% of the data to train the model. Once produced, the next step is to employ an independent accuracy assessment using high-resolution imagery (see Appendix D).

See5 file descriptions (Rulequest 2004)

a) .data file: Contains the training cases from which See5 extracts rules. This is also produced from the CART Module Sampling tool, by "drilling" the dependent variable pixels through the specified predictor images. Required by See5 Software.

b) .test file: Produced from the CART Module Sampling tool but not used by Southwest Regional Gap (SWReGAP). This file, if populated, would contain a separate "test" set of cases to evaluate the rules generated from See5. The SWReGAP mapping procedures did not populate this file, and it was not used.

c) .names.hst file: Produced from the CART Module Sampling tool. Details the distribution of samples available within the dependent input, and those output to the *.data and *.test file. Not required by See5, but produced by CART Module Sampling tool.

d) .set file: Produced from See5 software. This file contains the settings for the classification tree run. For example, the third value "15" indicates the number of boosts used for boosting.

e) .tree file: Produced from the See5 software. This file contains the classification tree in "tree" format. This, along with the *.data and *.names files, are required by the CART Module Classifier tool to spatially apply the tree.

f) .out file: Output file generated by See5 and displayed when See5 classification tree model has run. This file provides a visual representation of the classification tree that is somewhat easier to interpret than the *.tree file.

Appendix D. Validation

Introduction

Two accuracy assessments were conducted during the modeling and validation procedures to examine model performance. The first accuracy assessment(s), described in Appendix C, was preliminary and was employed in the See5 data mining software (Classification Tree [CT]) during the modeling procedures. The CT model was run utilizing 80% of the reference samples while randomly selecting and withholding 20% of the reference points to validate the CT model. The validation works by intersecting the validation sample points with the CT modeled map to see if the generated map agrees with the validation points. The .txt, .dbf, and .shp files were examined for a kappa statistic, error matrices (including commission and omission errors), and spatial references to errors, respectively. This process was repeated until a final model was selected and run using 100% of the reference data to train the model.

Discussed next, the final model was validated by a thorough independent accuracy assessment that utilized high-resolution National Agriculture Imagery Program (NAIP) imagery as the reference source. High-resolution (1 m) NAIP imagery was selected since it is readily available, free of charge, offers great spatial resolution with color infrared options, and can serve as a highly reliable surrogate for on-site ground-truthing. In this application, a 500-m systematic grid (445 points) was established within the delineated aspen mask. Each of the generated reference points in the grid were designated into one of the three aspen stand classes, then compared against the final model (i.e., aspen stand classification map) to create standard error matrices and a Kappa statistic.

Validation procedures

Validation procedures are all conducted in ArcGIS 9.2.

a) If the Hawth's Tools extension was not downloaded earlier, download and install from http://www.spatialecology.com/htools/. Once Hawth's Tools is installed, a new toolbar should appear in ArcMap. If it does not, check if the extension is active ("Tools" → "Extensions….") and make sure the toolbar is visible ("View" → "Toolbars").

b) To create a systematic grid, repeat steps "a" through "c" in Appendix B under "Sample Point Generation." (Note: For Cedar Mountain, a 500-m systematic grid was produced within the aspen mask shapefile; fig. A-2.)

c) Add study area NAIP images to viewer. Examine each grid point independently and remove sites that do not meet the criteria of at least 50% aspen cover. (Note: Grid layer needs to be in editing mode ["Editor toolbar" → "Editor Drop Down" → "Start Editing" → "Select source containing grid layer" → Click "OK"].)

d) Once grid is established, re-examine each point and determine its aspen stand class based on NAIP imagery canopy characteristics (e.g., 1—Healthy, 2—Damaged, etc.). This process of classifying reference points needs to be done prior to validating the model. (Note: This procedure can simultaneously occur during step "c.") In order to classify each point, a new column needs to be added to the grid layer attribute table. Make sure grid layer is *not* in editing mode. Select the grid layer and open the attribute table. In the attribute box, select "Options" → "Add Field" → name the field (e.g., stand class or NAIP) and select "Short integer" for the class → Click "OK." Create a second column (e.g., ID) that will be used to monitor accuracy during the validation process (table A-3).

e) Once all reference points have been classified, validation can begin. Add the modeled "aspen stand type" map layer, classified NAIP-based reference points, and NAIP imagery layers, and make sure they are active in the layer column on the left, with the NAIP imagery as the base layer.

f) Develop a labeling system to keep track of correct/incorrect validations (e.g., 1 = correct, 0 = incorrect). Begin validating by comparing the modeled map against the NAIP-based reference points. (Note: Having the NAIP imagery readily available is helpful during this phase.) If the aspen layer correctly maps a given cover type, enter a "1" in the ID column for that reference point. Enter a "0" if it is incorrect (table A-3). Repeat for each reference point.

g) Construct an error matrix (table A-4) so that user and producer accuracy measures can be determined. Also, calculate the Kappa (κ) or KHAT statistic (Congalton 1991) based on the produced error matrix (fig. A-3). See Jensen (1996) for guidelines on constructing both error matrices and KHAT statistics.

The Rocky Mountain Research Station develops scientific information and technology to improve management, protection, and use of the forests and rangelands. Research is designed to meet the needs of the National Forest managers, Federal and State agencies, public and private organizations, academic institutions, industry, and individuals. Studies accelerate solutions to problems involving ecosystems, range, forests, water, recreation, fire, resource inventory, land reclamation, community sustainability, forest engineering technology, multiple use economics, wildlife and fish habitat, and forest insects and diseases. Studies are conducted cooperatively, and applications may be found worldwide.

Station Headquarters
Rocky Mountain Research Station
240 W Prospect Road
Fort Collins, CO 80526
(970) 498-1100

Research Locations

Flagstaff, Arizona
Fort Collins, Colorado
Boise, Idaho
Moscow, Idaho
Bozeman, Montana
Missoula, Montana
Reno, Nevada
Albuquerque, New Mexico
Rapid City, South Dakota
Logan, Utah
Ogden, Utah
Provo, Utah

Federal Recycling Program Printed on Recycled Paper