CONSIDERATIONS FOR PLANNING, ACQUIRING, AND PROCESSING LIDAR DATA FOR FORESTRY APPLICATIONS

Robert J. McGaughey
Research Forester
USDA Forest Service
Pacific Northwest Research Station
University of Washington
PO Box 352100
Seattle, WA 98195-2100
bmcgaughey@fs.fed.us

Hans-Erik Andersen
Research Scientist
University of Washington
College of Forest Resources
PO Box 352100
Seattle, WA 98195-2100
hanserik@u.washington.edu

Stephen E. Reutebuch
Research Forester
USDA Forest Service
Pacific Northwest Research Station
University of Washington
PO Box 352100
Seattle, WA 98195-2100
sreutebuch@fs.fed.us

Abstract

Airborne laser scanning of forests provides accurate terrain models and, at the same time, estimates of multiple resource inventory variables through active sensing of three-dimensional (3D) forest vegetation structure. Researchers in forestry and remote sensing have developed analysis methods to characterize vegetation using airborne laser scanning data that can be applied to large geographic areas and a wide range of forest and non-forest vegetation types. Large-area data acquisitions are becoming more common as federal, state and local government agencies contract for large-area data coverage. Unfortunately, many such acquisitions are obtained using specifications designed for terrain mapping, often resulting in datasets that do not contain key information needed for vegetation measurement and analysis. Standards and specifications for airborne laser scanning missions designed for topographic mapping exist. However, similar guidelines for missions aimed at vegetation measurement and monitoring have not been developed.

This paper discusses the requirements for airborne laser scanning data used for topographic surveys and vegetation measurement and highlights deliverables, specific to forestry applications that should be included in data acquisition contracts. We describe the types of data products that can be expected from an airborne laser scanning mission and the amount of data that must be managed and stored. We present an overview of the current state-of-the-art in data visualization and processing, with emphasis on the analytical methods currently employed to characterize vegetation structure using airborne laser scanning data. Finally, five simple, easily understood data products are identified that would help insure that forestry needs are considered when multi-resource airborne laser scanning missions are flown.

Introduction

Airborne laser scanning data have proven to be a good source of information for describing the ground surface and characterizing the size and extent of man-made features such as urban areas and buildings. The technology has gained a strong foothold in mapping operations traditionally dominated by photogrammetric techniques. However, forestry-related applications are limited. Activity within the research community focuses on extracting descriptions of vegetation and other useful information from point clouds. Currently, there are very limited commercial services that use airborne laser scanning data to create information products, other than bare-ground surface and canopy surface models, that are specific to forestry or that characterize land cover characteristics over large areas. Airborne laser scanning has tremendous potential to improve the extent and quality of information describing the ground surface and vegetation characteristics over large areas. Using products derived from airborne laser scanning data, managers will be able to make more informed decisions and will have more confidence that their decisions can be implemented on the ground.

LIDAR Overview

There are several varieties of airborne light detection and ranging (LIDAR) systems; in this paper we will focus on the most common terrain mapping system—discrete return, small-footprint LIDAR (i.e., typical laser beam diameter on the ground in the range of 0.2—1.0 m). These systems have been developed over the last 15 years to map terrain (Wehr and Lohr, 1999). An airborne laser scanning system, mounted in either a fixed- or rotary-wing aircraft, consists of four basic components:

- Laser scanner,
- Global Positioning System (GPS) mounted in the aircraft and positioned on the ground,
- Inertial Measurement Unit (IMU),
- On-board computer to control the equipment and monitor mission status.

Laser scanners designed for terrain mapping emit near-infrared (NIR) laser pulses at a high frequency (typically 25,000 to 100,000 per second). The position and attitude of the laser scanner unit at the time each pulse is emitted are determined from flight data collected by the GPS and IMU units. The range or distance between the scanner and an object that reflects the pulse is computed using the time it takes for the pulse to travel from the scanner, to the object, and back to the scanner. A precise coordinate is computed for each reflection point using the position and attitude of the scanner and the direction and distance traveled by the pulse from the scanner to the object.

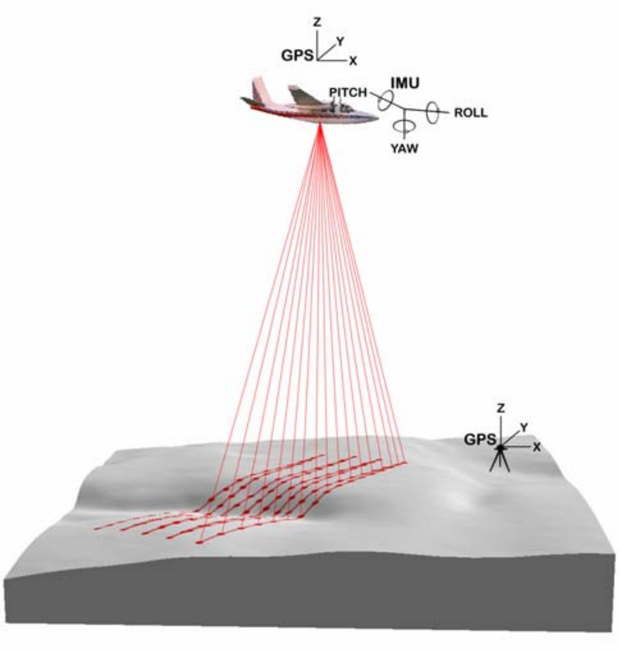

Figure 1. Schematic showing LIDAR data collection over bare ground.

A swath of terrain under the aircraft is surveyed through the lateral deflection of the laser pulses and the forward movement of the aircraft. The scanning pattern within the swath is established by an oscillating mirror or rotating prism which causes the pulses to sweep across in a consistent pattern below the aircraft (figure 1). Large areas are surveyed with a series of swaths that often overlap one another by 20 percent or more. The final pattern of pulse reflection points on the ground and the scanned swath width depend on the scanning mechanism settings and design (e.g. pulse rate, returns per pulse, and scanning angle), flying height and speed, and the shape of the topography.

Most LIDAR systems can distinguish 2-5 reflected signals, or returns, per laser pulse. Multiple returns occur when the pulse strikes a target that doesn't completely block the path of the pulse and the remaining portion of the pulse continues on to a lower object (figure 2). This situation frequently occurs in forest canopies that have small gaps between branches and foliage. Most terrain mapping missions are flown in leaf-off conditions to maximize the percentage of pulses that reach the ground surface. Projects that are designed to characterize canopy conditions are often flown in leaf-on conditions to maximize the number of returns from tree crowns and other vegetation.

Figure 2. Schematic showing multiple LIDAR returns resulting from the laser light being only partially obstructed by tree branches.

System manufacturers have worked hard to develop processing methods for distinguishing between laser reflections from the ground surface (terrain measurements) and vegetation surface. As a result, processed LIDAR data provide accurate measurements of the ground surface. LIDAR system manufacturers typically quote root mean squared (RMS) errors of 10-15 cm vertical and 50-100 cm horizontal for terrain mapping products under optimal conditions. In several studies the vertical accuracy of LIDAR terrain measurements was found to be in the range of 15-50 cm over ground and cover conditions ranging from flat areas with minimal vegetation (Pereira and Janssen, 1999) to moderately steep terrain with forest cover ranging from clearcuts to mature stands (Reutebuch et al., 2003; Kraus and Pfeifer, 1998).

LIDAR systems currently in operation can detect and separate vertical feature that are a minimum of 1-3 m apart. This means that systems cannot produce returns that represent objects closer than 1-3 m (vertically) and systems cannot detect the ground surface under low vegetation (less than 1-3 m) unless the vegetation is sparse (e.g. shrubs in leaf-off condition) enough to allow some laser pulses to reach the ground surface. For projects designed to map the terrain surface, penetration through vegetation to the ground surface is important and system settings are often adjusted to maximize pulse penetration. Unfortunately, the industry has less experience with vegetation mapping projects and the effects of system settings on the final data products are less well known.

Over the last decade LIDAR system capabilities have dramatically increased and data acquisition costs have correspondingly decreased as advances in IMUs, computing capability, and GPS technology have allowed LIDAR to move into the mainstream commercial terrain mapping sector. Today, several vendors market LIDAR systems, and several third-party vendors offer specialized LIDAR data processing software to support terrain mapping activities. Numerous LIDAR remote sensing firms offer a complete range of mapping services including the generation of digital terrain models, contour maps, extraction of infrastructure locations and characteristics, and delivery of raw data.

Data Specifications and Characteristics

Most commercial applications of LIDAR focus on mapping the ground surface. Missions are planned to produce pulse densities ranging from 0.1 to 1 pulse per square meter with the final density of ground returns dependent on the amount and density of vegetation cover and the presence of above-ground structures. For data used to characterize vegetation, different specifications are needed. Table 1 compares typical acquisition specifications for LIDAR missions designed to map topography and characterize vegetation. The actual specifications for a mission depend on the mission objectives and the desired accuracy of the final products. The values presented in Table 1 provide information useful when comparing the two types of LIDAR missions and are not intended for use in data acquisition contracts. In general, missions designed to characterize vegetation structure require 200 to 300 percent more flight time and cost significantly more per acre than missions designed to map topography.

Table 1. LIDAR data acquisition specifications for topographic and vegetation survey missions with fixed-wing aircraft.

	Topographic mapping	Vegetation characterization
Scan angle	±20-30 degrees	±10-15 degrees
Flying height[1]	2000 m	1000-1200 m
Pulse repetition frequency	10-70 kHz	30-100 kHz
Beam footprint on the ground	40-80 cm	20-50 cm
Swath width	1500-2000 m	400-600 m
Pulse spacing on the ground[2]	1-3 m	0.2-1.0 m

[1] Pilots attempt to maintain the height above ground but in areas with steep topography actual height above ground over the coverage area varies.

[2] Pulse spacing on the ground is for a single flight line. Areas covered by more than one flight line will have smaller pulse spacing and higher return density.

A variety of data products ranging from raw return data to highly processed surface models can be produced from a LIDAR mission. Raw return data includes all LIDAR returns with the exception of returns classified as outliers. Raw return data are typically delivered in either ASCII text format or an industry standard binary LAS format (ASPRS, 2005). Files, in either format, always include the return location (easting and northing) and elevation for each return. Additional information for each return may include the return number, total number of returns for the pulse, intensity (relative measure of the amount of energy reflected) for the return, pulse identifier (simple sequential number or GPS time), and scan angle for the pulse. If the acquisition contract specifies a bare-ground surface model, the deliverables may include either separate files containing the returns that were classified as bare-ground returns or additional fields in the “all returns” files that indicate that a return was classified as a bare-ground return.

Surface models describing either the ground surface or the upper surface of vegetation are commonly produced from LIDAR data. Ground surface models result from simple classification of LIDAR returns or they may be a combination of classification results and breaklines extracted from other data sources such as topographic surveys, road location and size data, or aerial photographs. Vegetation surface or canopy surface models are created from the subset of return data that includes all first returns that are not classified as bare-ground. Vegetation surface models may be normalized by subtracting the bare-ground elevation to create a vegetation height model.

Deliverables for a LIDAR mission should always include metadata describing the flight and instrument specifications during data acquisition, processing methods, projection information, and reference datum for elevations. The format and level of detail available in metadata varies by LIDAR operator. As a minimum, the metadata should include the coordinate system, datum, and ellipsoid information, the geoid used to compute return elevations, and a description of file formats and file naming conventions for all data products.

Absolute positional accuracy of return data, while it varies depending on flight parameters and GPS data quality, ranges from 30-100 cm. Vertical accuracy, dependent primarily on flying height, ranges from 10-25 cm. Accuracy of bare-ground surface products is assessed by computing the root mean square error (RMSE) for elevations in flat, open areas. RMSE values of 10 to 30 cm are common. In forested areas, the RMSE may be significantly higher due to the reduced density or complete absence of returns from the ground surface under dense vegetation.

LIDAR data are delivered in a variety of formats. Surface models, either bare-ground or vegetation surface, are usually delivered as a regular grid of elevations interpolated from the returns classified as bare-ground. In some cases, triangular irregular network (TIN) models that use the bare-ground returns as nodes are provided. Surface models are usually delivered as ASCII raster files, binary raster files, or ArcInfo* interchange files.

LIDAR datasets can be very large. High pulse densities combined with the presence of vegetation cover can result in up to 150,000 returns per hectare requiring up to 3.6 gigabytes of storage space per hectare. Current practices include delivery of data on DVD or external hard drives. Data users should be prepared to create backup copies of all data files regardless of the media and should realize that data may need to be reformatted and copied onto a faster storage device for analysis.

Data Processing and Visualization

The biggest hurdle when using LIDAR to support project-level analyses is the lack of software available for processing and analyzing the large datasets common to LIDAR acquisitions. Existing GIS and image processing systems cannot be used to efficiently process LIDAR return data. Limitations arising from the number of points in a data file or the presence of several returns with approximately the same XY location make it difficult to use the data in such applications. Two general schools of thought exist for LIDAR data analysis. The first uses the LIDAR return data as true XYZ points and employs custom developed programs to conduct analyses. The second converts the return data, either all returns or some subset of returns, into surfaces or images and then uses commercial image-processing or GIS software to conduct analyses. Both approaches have advantages. In general, point data are needed if the final product involves individual tree characteristics or plot-level comparison of field data and LIDAR data. Such analyses take advantage of returns from interior branches and foliage and returns from the ground surface beneath tree crowns. The surface (or image) approach is suitable for stand-level characterization of vegetation, analyses of gaps, and analyses involving vegetation characteristics over large land areas. Often the decision to use point data or convert the data into surface or images depends solely on the individual's ability to develop custom programs or scripts to process their data.

Hyyppa and others (2004) provide an excellent overview of the current algorithms and methods used to produce a variety of products from LIDAR data. They describe derivation of bare-ground and canopy surfaces, prediction of stand-level and individual-tree characteristics using LIDAR data, identification of individual trees using image processing methods, and efforts to fuse LIDAR data with aerial images. They also provide an extensive list of references covering a range of LIDAR-related topics.

The FUSION software system developed by the authors (McGaughey and Carson, 2003; McGaughey et al. 2004) provides visualization and analysis capabilities for LIDAR projects (Figure 3). The software is in the public domain, operates on all version of Microsoft Windows, and is currently distributed by the Remote Sensing Applications Center of the USDA Forest Service. FUSION consists of two main subsystems:

- interactive data exploration system (FUSION and LDV) and
- command-line processing system (PDQ viewer and analysis programs).

The interactive data exploration system consists of two main programs, FUSION and LDV (LIDAR data viewer). The primary interface, provided by FUSION, consists of a graphics display window and a control window. The FUSION display presents all project data using a 2D display typical of geographic information systems (GIS). It supports a variety of data types and formats but requires that all data be georeferenced using the same projection and units of measurement. Users interact with FUSION using the mouse or by directly inputting sample locations to select subsets of LIDAR data for display in LDV. LDV (figure 4) provides an interactive 3D visualization environment for the examination of spatially-explicit data subsets including LIDAR returns, images, surface models, and 3D objects (e.g. tree models). LDV provides a number of analysis tools that can be used to better understand the kinds of information that can be extracted from the LIDAR point cloud and to experiment with parameters that can be used with the command-line processing programs.

---

* Use of trade or firm names in this publication is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.

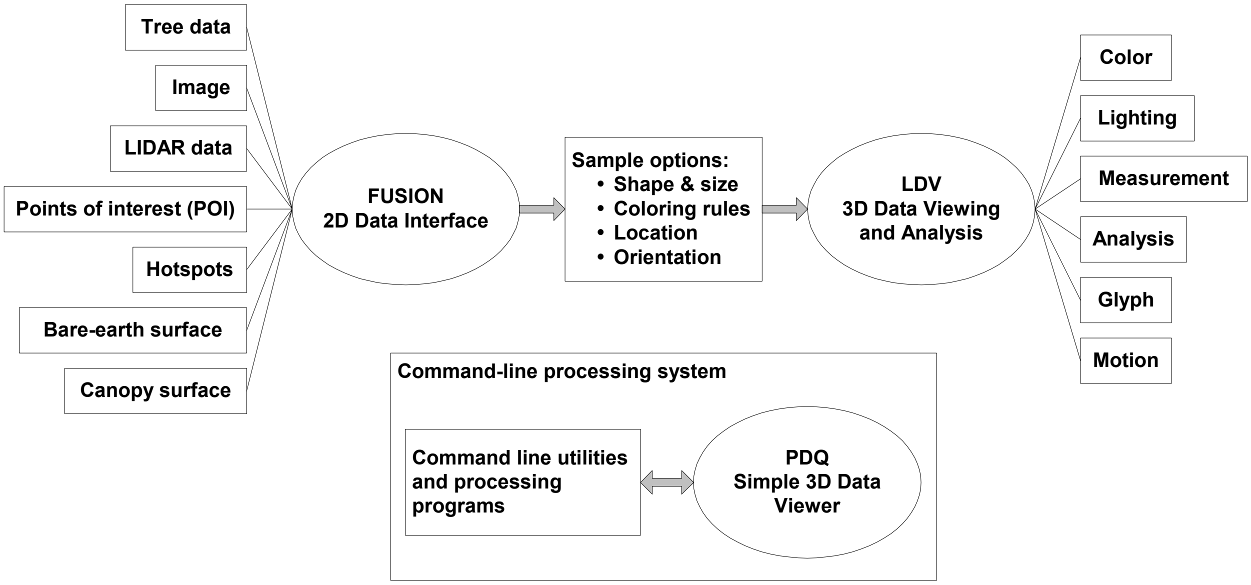

Figure 3. Schematic showing the overall structure of the FUSION software system.

The command-line processing system provides a set of analysis and processing programs that help users assess the overall quality and completeness of LIDAR data, develop a basic set of derived products, and automate processes involving the spatial matching of field plot data with LIDAR data samples. The command-line programs are designed to operate on project datasets containing multiple data tiles with each tile containing up to 20 million returns. Programs are included to filter "all-return" data to obtain bare-ground returns, create surface models from data files, produce canopy surface and canopy height models, produce layers containing percent cover estimates, and to produce images using the return intensity data. The CATALOG program, designed to provide information to help assess the overall quality and completeness of LIDAR acquisitions, combines many of these capabilities to produce an HTML summary for an entire project area. The summary includes a statistical summary for each tile in the project area, a geo-referenced image showing tile locations, geo-referenced images showing pulse and return densities, analyses to detect and highlight tiles that might contain anomalous returns (elevation outliers), and creation of a geo-referenced, project-wide intensity image.

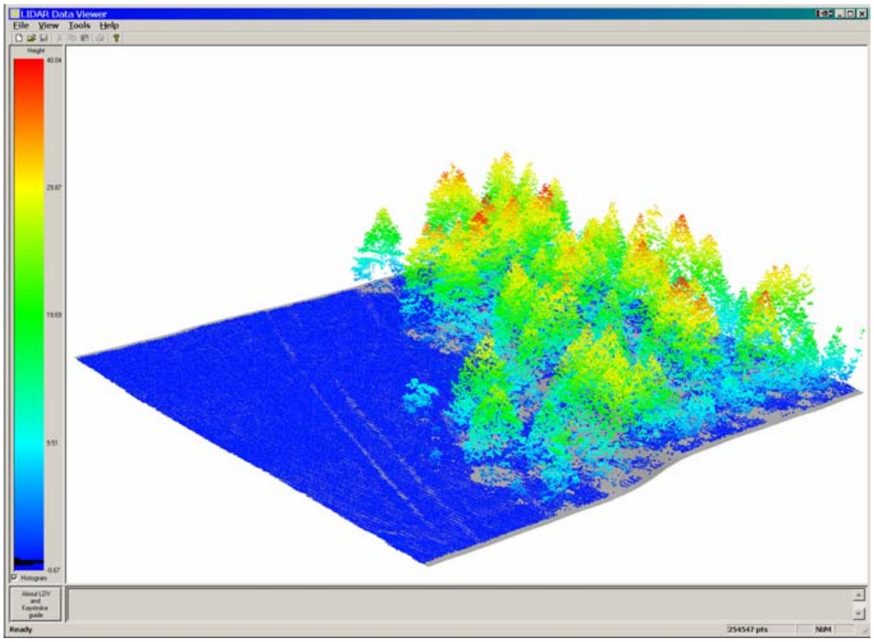

Figure 4. LDV allows interactive viewing of LIDAR point data, surface models, and other data sampled in FUSION.

LIDAR Data Products Useful for Forestry Applications

There are several simple, easily understood and widely recognized LIDAR-derived forest mapping products that many agencies and specialists within the resource management community would find useful. The following five could be generated easily for most LIDAR projects (Reutebuch et al. 2005):

1. High-resolution (1–5 m) bare-ground digital elevation model (DEM). These DEMs provide improved data for many applications including hydrologic and erosion process modeling, landscape modeling, road and harvest planning and design, and GIS (Figure 5B).
2. Canopy height models (CHM). CHMs provide spatially explicit vegetation height data over the landscape for estimation of growing stock, input for habitat and fire models, and any other resource planning activities where spatial arrangement and tree height are important considerations (Figure 5C).
3. Canopy cover or canopy density maps. These images provide a direct measurement of vegetation density by height aboveground. Figure 5D illustrates canopy cover where canopy height is greater than 2 m.
4. LIDAR intensity images. These high resolution images can be matched with existing orthophotographs and other digital imagery for change detection and monitoring over time. They also are useful in verifying the registration of LIDAR data with other geospatial data layers. As shown in Figure 5D, intensity data can be used in conjunction with CHMs to identify hardwood (brown) and conifer canopy areas (green) using data acquired in leaf-off conditions.
5. All-returns dataset. This archive of all the LIDAR returns and their associated reflectance intensity could be used for a wide range of specialized analysis and provides baseline data on current terrain and vegetation structure that could be used in the future for change detection and monitoring (e.g., crown expansion or dieback). At a minimum, this dataset should include pulse number, return number, east coordinate, north coordinate, elevation, and return intensity for each LIDAR return and metadata documenting the LIDAR mission flight parameters, sensor type and settings, GPS control, horizontal and vertical datum, coordinate units and projection, and date and time of mission. Ideally, all-return data files should be in the American Society for Photogrammetry and Remote Sensing LAS LIDAR data exchange format (ASPRS, 2005) or other formats that can be read by analysis and GIS software.

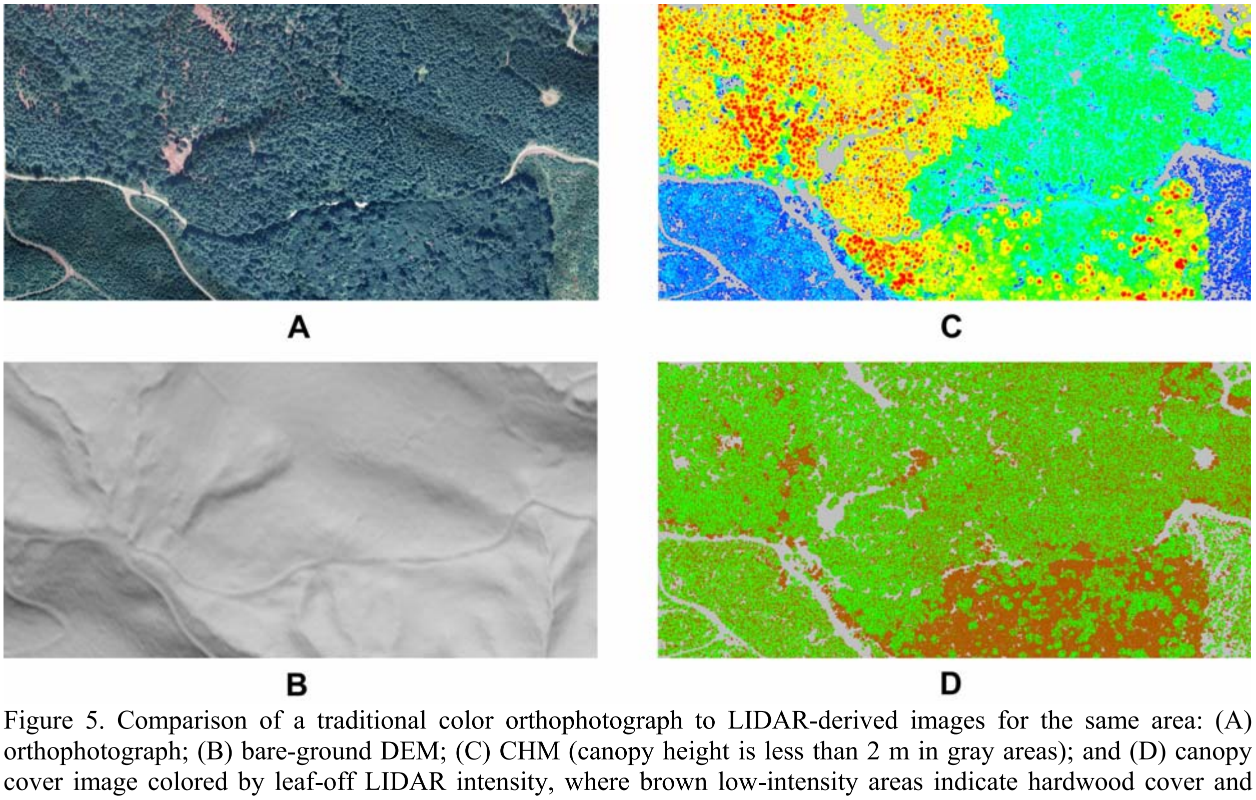

Figure 5. Comparison of a traditional color orthophotograph to LIDAR-derived images for the same area: (A) orthophotograph; (B) bare-ground DEM; (C) CHM (canopy height is less than 2 m in gray areas); and (D) canopy cover image colored by leaf-off LIDAR intensity, where brown low-intensity areas indicate hardwood cover and green high-intensity areas indicate conifer cover.

Conclusions

Over the last 5 years, numerous studies have shown that LIDAR data can provide high-resolution data for multiresource management and analyses including traditional forest inventory and more specialized single-use analysis (e.g., analysis of individual tree characteristics). Simultaneously, LIDAR has emerged as the leading technology for high-resolution terrain mapping, spurring the development of national guidelines and standards in this domain. It appears there is a similar need to develop national standards and guidelines for LIDAR data collection for forest vegetation measurement and monitoring to insure that the maximum value can be returned from future LIDAR projects over forested regions. LIDAR has the potential to revolutionize forest mensuration and monitoring. Systems operational today can provide high-density data that, combined with well described analysis processes, can produce vegetation height and structure information over large areas. Unlike passive remote sensing systems, LIDAR produces 3D measurements that are accurately geo-referenced and can be directly compared to plot measurements, ground survey data or measurable features such as building dimensions and tree heights.

References

ASPRS. 2005. ASPRS LIDAR Data Exchange Format Standard Version 1.1 March 7, 2005: http://www.lasformat.org/ (April 13, 2006).

Hyyppä, J.; Hyyppä, H.; Litkey, P.; Yu, X.; Haggrén, X. H.; Ronnholm, P.; Pyysalo, U.; Pitkänen, J.; Maltamo, M. 2004. Algorithms and methods of airborne laser scanning for forest measurements. International Archives of Photogrammetry and Remote Sensing. Vol. XXXVI Part 8/W2, Freiburg, Germany, pp. 82-89.

Kraus, K. and Pfeifer, N. 1998. Determination of terrain models in wooded areas with airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing. 53:193-203.

McGaughey, R. J.; Carson, W. W. 2003. Fusing LIDAR data, photographs, and other data using 2D and 3D visualization techniques. In: Proceedings of Terrain Data: Applications and Visualization – Making the Connection, October 28-30, 2003; Charleston, South Carolina: Bethesda, MD: American Society for Photogrammetry and Remote Sensing. pp. 16-24.

McGaughey, R.J., W.W Carson, S.E Reutebuch, and H.-E. Andersen. 2004. Direct measurement of individual tree characteristics from LIDAR data. In: Proceedings of the 2004 Annual ASPRS Conference, May 23-28 2004; Denver, Colorado: Bethesda, MD: American Society of Photogrammetry and Remote Sensing. Unpaginated CD-ROM.

Pereira, L. and Janssen, L. 1999. Suitability of laser data for DTM generation: a case study in the context of road planning and design. ISPRS Journal of Photogrammetry and Remote Sensing. 54:244-253.

Reutebuch, S.E.; McGaughey, R.J.; Andrsen, H.-E.; Carson, W.W. 2003. Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. Canadian Journal of Remote Sensing. 29:527-535.

Reutebuch, S.E., H.-E. Andersen, and R.J. McGaughey. Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory. Journal of Forestry 103(6): 286-292.

Wehr, A. and Lohr, U. 1999. Airborne laser scanning—an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing. 54:68–82.