EXHIBIT 5

United States
Department of
Agriculture

Forest Service

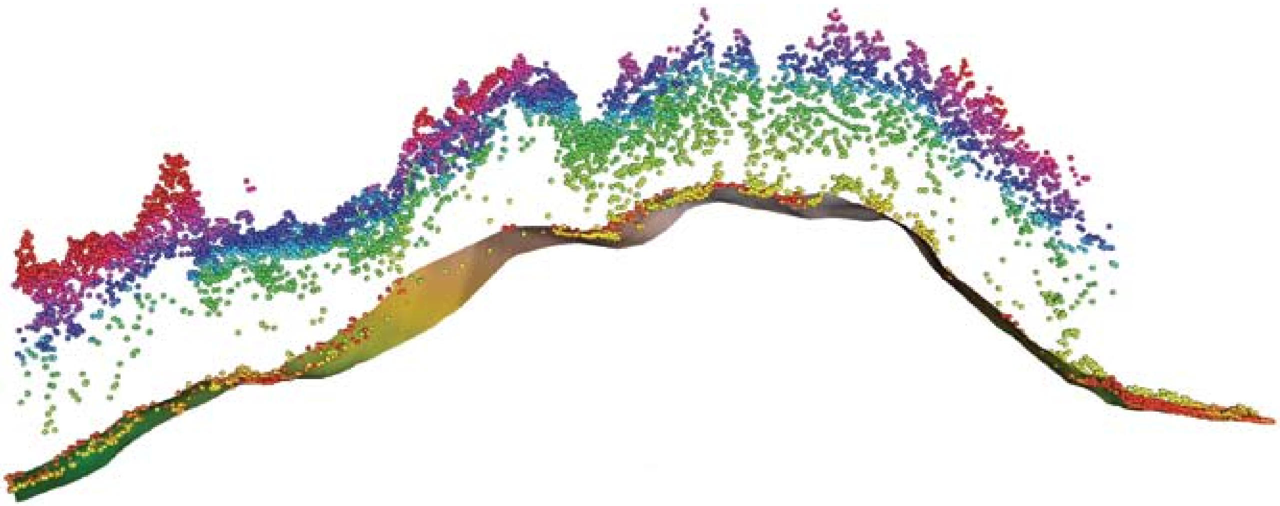
Pacific Northwest
Research Station

General Technical Report
PNW-GTR-768

July 2008

A Guide to LIDAR Data Acquisition and Processing for the Forests of the Pacific Northwest

Demetrios Gatziolis and Hans-Erik Andersen



The **Forest Service** of the U.S. Department of Agriculture is dedicated to the principle of multiple use management of the Nation's forest resources for sustained yields of wood, water, forage, wildlife, and recreation. Through forestry research, cooperation with the States and private forest owners, and management of the National Forests and National Grasslands, it strives—as directed by Congress—to provide increasingly greater service to a growing Nation.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, 1400 Independence Avenue, SW, Washington, DC 20250-9410 or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

Authors

Demetrios Gatziolis is a research forester, Forestry Sciences Laboratory, 620 SW Main, Suite 400, Portland, OR 97205; **Hans-Erik Andersen** is a research forester, Forestry Sciences laboratory, 3301 C St., Suite 200, Anchorage, AK 99503.

Abstract

Gatziolis, Demetrios; Andersen, Hans-Erik. 2008. A guide to LIDAR data acquisition and processing for the forests of the Pacific Northwest. Gen. Tech. Rep. PNW-GTR-768. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 32 p.

Light detection and ranging (LIDAR) is an emerging remote-sensing technology with promising potential to assist in mapping, monitoring, and assessment of forest resources. Continuous technological advancement and substantial reductions in data acquisition cost have enabled acquisition of laser data over entire states and regions. These developments have triggered an explosion of interest in LIDAR technology. Despite a growing body of peer-reviewed literature documenting the merits of LIDAR for forest assessment, management, and planning, there seems to be little information describing in detail the acquisition, quality assessment, and processing of laser data for forestry applications. This report addresses this information deficit by providing a foundational knowledge base containing answers to the most frequently asked questions.

Keywords: LIDAR, Pacific Northwest, FIA, forest inventory, laser, absolute and relative accuracy, precision, registration, stand penetration, DEM, canopy surface, resolution, data storage, data quality assessment, topography, scanning.

Contents

Introduction

Light detection and ranging (LIDAR), also known as airborne laser scanning (ALS), is an emerging remote sensing technology with promising potential to assisting mapping, monitoring, and assessment of forest resources. Compared to traditional analog or digital passive optical remote sensing, LIDAR offers tangible advantages, including nearly perfect registration of spatially distributed data and the ability to penetrate the vertical profile of a forest canopy and quantify its structure. LIDAR has been used in many parts of the world to successfully assess height and size of individual trees or, at the stand level, to estimate canopy closure, volume, and biomass of forest stands; to assess wildlife habitat; and to quantify stand susceptibility to fire (Andersen et al. 2005, Hinsley et al. 2006, Means et al. 2000, Naesset 2002, Persson et al. 2002, Popescu and Zhao 2007). Continuous technological advancement and competition among vendors in the United States have resulted in substantial reductions in data acquisition cost and have enabled acquisition of spatially complete laser data over entire states and regions. The U.S. Geological Survey has recently announced a plan to coordinate the acquisition of LIDAR data at a national scale (Stoker et al. 2007). Laser scanning data are regularly acquired over several national forests in Western States. These developments have triggered an explosion of interest in LIDAR technology. Despite a growing body of peer-reviewed literature documenting the merits of LIDAR for forest assessment, management, and planning, there seems to be a void in information describing issues related to the acquisition and processing of laser data. In the past year alone, the authors have received numerous requests for guidance on the technical specifications of planned data acquisitions, on instructions on how to perform data quality assessment, and on whether scanning data can be used to meet specific objectives. This article addresses this information deficit by providing a foundational knowledge base containing answers to the most frequently asked questions.

LIDAR offers nearly perfect registration of spatially distributed data and the ability to penetrate the vertical profile of a forest canopy.

LIDAR Systems

A LIDAR system operating from an airborne platform comprises a set of instruments: the laser device; an inertial navigational measurement unit (IMU), which continuously records the aircraft's attitude vectors (orientation); a high-precision airborne global positioning system (GPS) unit, which records the three-dimensional position of the aircraft; and a computer interface that manages communication among devices and data storage. The system also requires that a GPS base station installed at a known location on the ground and in the vicinity (within 50 km) of the aircraft, operate simultaneously in order to differentially correct, and thus improve the precision of, the airborne GPS data.

A LIDAR system comprises the laser device, an inertial navigational measurement unit, a high-precision airborne global positioning system, and a computer interface.

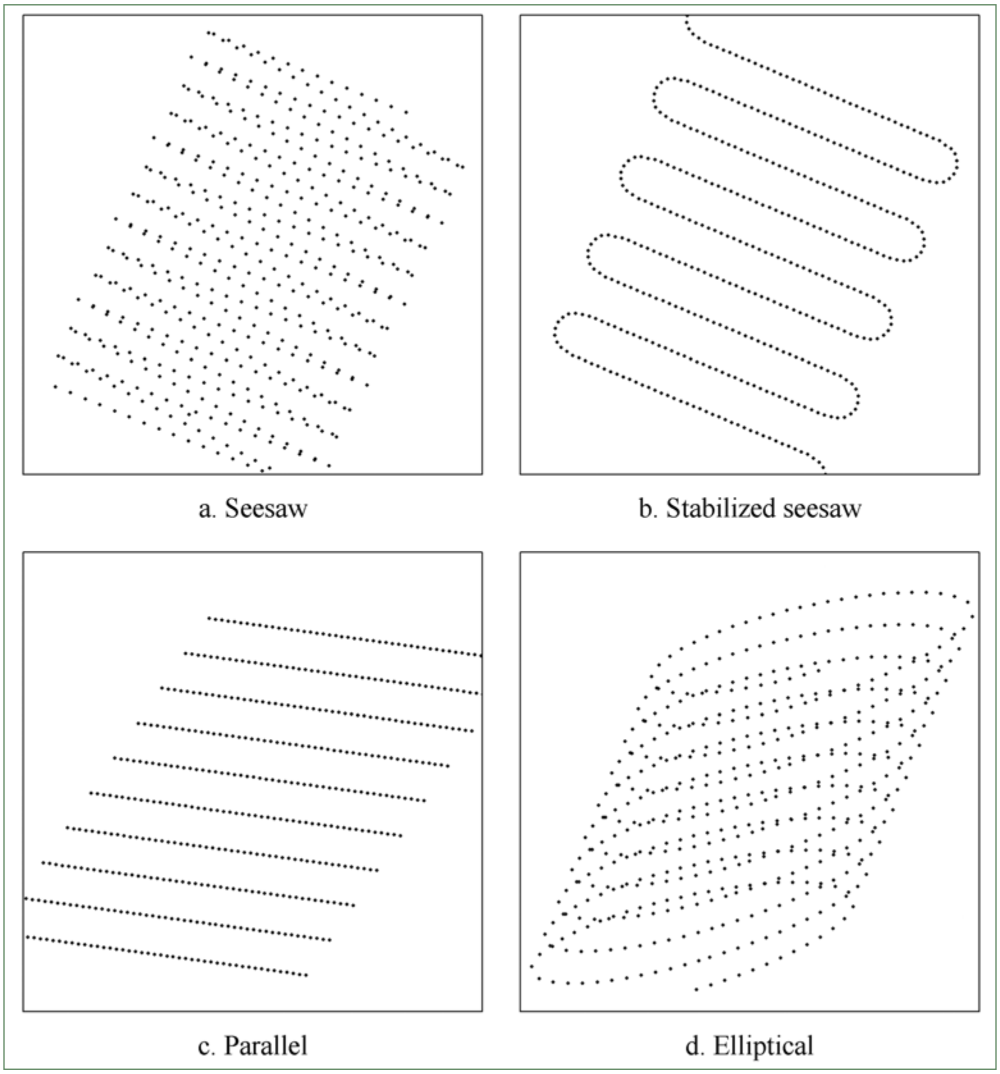
The laser device emits pulses (or beams) of light to determine the range to a distant target. The distance to the target is determined by precisely measuring the time delay between the emission of the pulse and the detection of the reflected (backscattered) signal. In topographic mapping and forestry applications, the wavelength of the pulses is in the near-infrared part of the spectrum, typically between 1040 and 1065 nm. There are two types of LIDAR acquisition differentiated by how backscattered laser energy is quantified and recorded by the system's receiver. With **waveform** LIDAR, the energy reflected back to the sensor is recorded as a (nearly) continuous signal. With **discrete-return**, **small-footprint** LIDAR, reflected energy is quantized at amplitude intervals and is recorded at precisely referenced points in time and space. Popular alternatives to the term "point" include "return" and "echo." The energy amplitude pertaining to each return is known as intensity. This article addresses only small-footprint, discrete-return LIDAR.

System Specifications

LIDAR systems have been evolving for more than a decade, and will likely continue to evolve even faster in the years to come. Hence, when planning data acquisition, it is essential to obtain specifications of currently available systems. Such specifications will determine both data acquisition costs and, quite likely, the feasibility of projects the acquired data are expected to support. Baltsavias (1999a) provided a good (if somewhat out-of-date) overview of the basic engineering and geometric concepts underlying airborne laser scanning, and Baltsavias (1999b) illustrated the variability in specifications among commercial systems. The major operational specifications of a LIDAR system are outlined below:

- **Scanning frequency** is the number of pulses or beams emitted by the laser instrument in 1 second. Older instruments emitted a few thousand pulses per second. Modern systems support frequencies of up to 167 kHz (167,000 pulses per second). Sometimes they can be operated at lower-than-maximum frequencies, typically 100 kHz or 71 kHz, but seldom at low frequencies, say, 10 kHz. The scanning frequency is directly related to the density of discrete returns obtained. Thus a system operating at 150 kHz onboard an aircraft flying at constant speed at a standard height above a target will generate a much higher number of returns than when operating at 71 kHz. Equivalently, a high-frequency system can generate desired return densities by operating on an aircraft that flies higher and faster than an aircraft carrying a lower frequency system, thereby reducing flying time and acquisition costs.

- **Scanning pattern** is the spatial arrangement of pulse returns that would be expected from a flat surface and depends on the mechanism used to direct pulses across the flight line. Of the four scanning patterns supported by instruments used in acquiring laser data for forestry applications, the **seesaw** pattern (fig. 1a) and its **stabilized** equivalent (fig. 1b) are the most common. In these two patterns, the pulse is directed across the scanning swath by an oscillating mirror, and returns are continuously generated in both directions of the scan. Although this configuration is designed to preserve the spacing between returns, in practice, pulse density is not uniform and returns tend to “bunch up” at the end of the swath because of mirror deceleration. The nonuniform spacing of returns can be partially


a. Seesaw

b. Stabilized seesaw

c. Parallel

d. Elliptical

Figure 1—Nadir view of theoretical scanning patterns of LIDAR instruments.

mitigated, but not eliminated, with the use of galvanometers. In the **parallel line** pattern (fig. 1c), a rotating polygonal mirror directs pulses along parallel lines across the swath, and data are generated in one direction of the scan only. The **elliptical** pattern (fig. 1d) is generated via a rotating mirror that revolves about an axis perpendicular to the rotation plane.

- **Beam divergence**. Unlike a true laser system, the trajectories of photons in a beam emitted from a LIDAR instrument deviate slightly from the beam propagation line (axis) and form a narrow cone rather than the thin cylinder typical of true laser systems. The term “beam divergence” refers to the increase in beam diameter that occurs as the distance between the laser instrument and a plane that intersects the beam axis increases. Typical beam divergence settings range from 0.1 to 1.0 millirad. At 0.3 millirad, the diameter of the beam at a distance of 1000 m from the instrument is approximately 30 cm (fig. 2). Because the total amount of pulse energy remains constant regardless of the beam divergence, at a larger beam divergence, the pulse energy is spread over a larger area, leading to a lower signal-to-noise ratio.
- **Scanning angle** is the angle the beam axis is directed away from the “focal” plane of the LIDAR instrument (fig. 3) It should not be confused

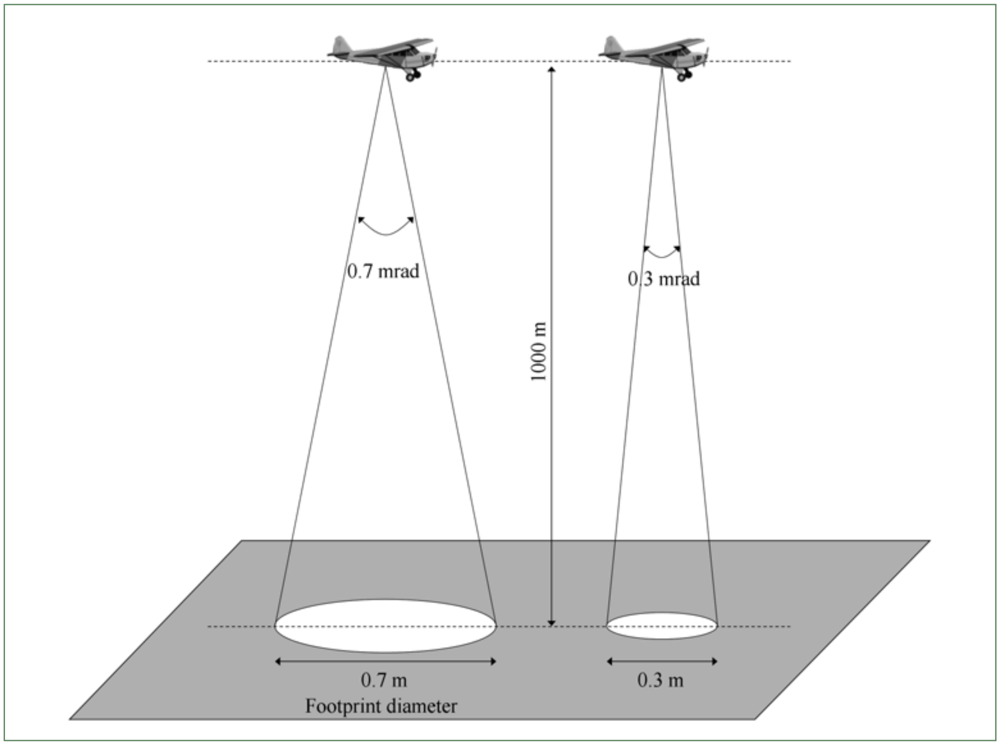


Figure 2—Illustration of LIDAR beam divergence. Horizontal and vertical distances are drawn in different scales.

with the angle formed between the beam axis vector and a vertical plane (nadir view), because the latter angle is affected by the attitude of the aircraft. The maximum angle supported by most systems does not exceed 15 degrees. The angle is recorded as positive toward the starboard and negative toward the port side of the aircraft. The combination of scanning angle and aboveground flight height determines the **scanning swath** (fig. 3).

- **Footprint diameter** is the diameter of a beam intercepted by a plane positioned perpendicularly to the beam axis at a distance from the instrument equal to the nominal flight height (fig. 2). It is thus a function of both beam divergence and the above-target flight height. The distribution of pulse energy is not uniform over the extent of the footprint. It decreases radially from the center and can be approximated by a two-dimensional Gaussian distribution.

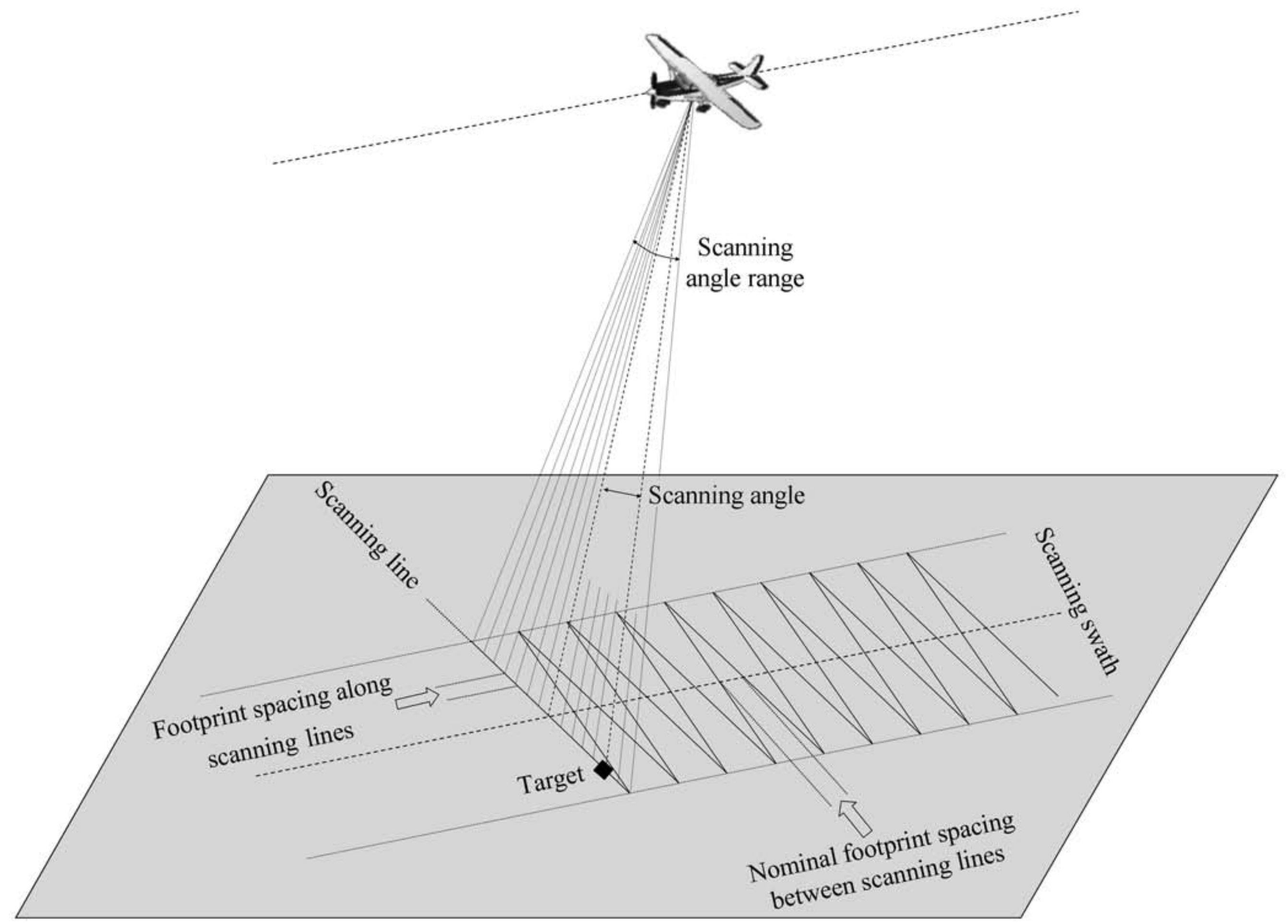


Figure 3—Illustration of scanning attributes of LIDAR data acquisition. Aircraft flying parallel to the ground and seesaw scanning pattern are assumed.

- **Pulse length** is the duration of the pulse, in nanoseconds (ns). Along with discretization settings (below), it determines the range resolution of the pulse in multiple return systems, or the minimum distance between consecutive returns from a pulse.
- **Number of returns (per beam/pulse)** is the maximum number of individual returns that can be extracted from a single beam. Certain systems can identify either the first or the first and last returns. Most modern systems can identify multiple returns (e.g., up to five) from a single beam.
- **Footprint spacing** is the nominal distance between the centers of consecutive beams along and between the scanning lines (fig. 3), which, along with the beam divergence, determines the spatial resolution of LIDAR data. The footprint spacing is a function of scanning frequency, the aboveground flight height, and the velocity of the aircraft.
- **Discretization settings** are specifications integral to the processing of the backscattered energy of a pulse to identify individual returns (fig. 4). They are system-specific and proprietary, and sometimes are referred to as **digitization settings**. They control the minimum energy amplitude necessary to produce a return and, along with the pulse length, determine the minimal distance between consecutive returns (discretization tolerance) from

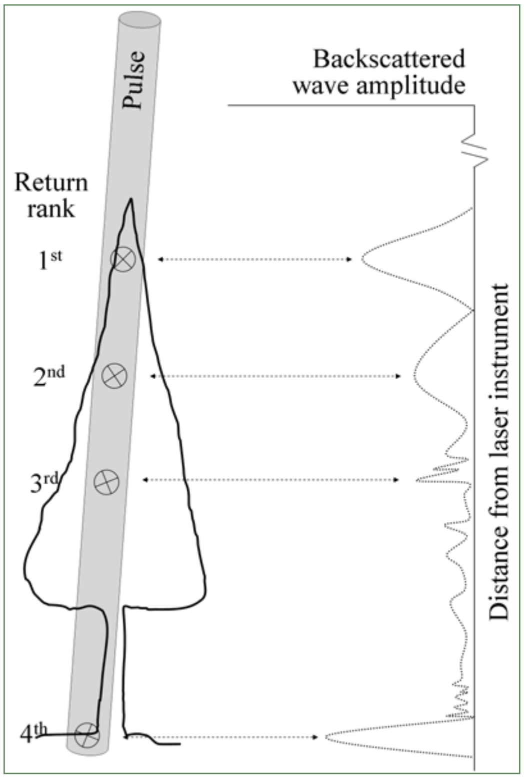


Figure 4—Illustration of the discretization process used to identify individual returns by processing the backscattered energy of a laser pulse.

the same pulse. Modern instruments can process the energy-backscatter pertaining to a single beam and identify up to six returns, but the majority support only up to four. The optimal settings for forestry applications likely depend on acquisition objectives and vegetation structure.

Data Attributes

Small-footprint LIDAR data comprise a set of return coordinates in three dimensions with each return usually carrying attribute values that relate either to that return or to the pulse from which the return was generated.

- **Pulse density** is a direct function of the footprint spacing (described above) over a hypothetical flat plane: pulse density = 1/(footprint spacing^2). This is the most consistent measure of the spatial resolution of a LIDAR data set.
- **Return density** is the most common term used in describing a data set, and is often confused with **pulse density**. It is the mean number of returns in the data set present (in two dimensions) in a unit square area, typically 1 m^2. With the exception of single-return systems, return density is controlled by the specifications and operation mode of a LIDAR system and by the target scanned. Assuming that all other specifications remain the same, the return density generated by a four-return-per-square-meter-capable system over a forest stand will be much higher than the density generated over a nearby pasture (fig. 5), because in the latter case, virtually all the energy returned falls within a single quantum (distance class). Because of this scene-dependent variability, users should specify a minimum pulse density for a given acquisition, instead of return density.
- **Return intensity** or simply intensity, is an attribute that describes the strength of the beam backscattering pertaining to the return in question. It depends on the reflectance properties of the target, and hence it can potentially be used in target discrimination. Its utility for object classification is often reduced because of its dependence on bidirectional reflectance distribution function effects, the distance (range) to the laser instrument, the total number of returns identified in the parent beam, the rank of the return (first, second, etc.) in the parent beam, and the receiver's gain factor. The latter term describes the scaling of the receiver's sensitivity designed to prevent hardware damage in the event that it receives an extraordinarily high amount of backscattered energy as can occur with high reflectivity targets. Such reduction in sensor sensitivity is practically instantaneous. The reverse scaling, an increase in sensitivity in the presence of continuously weak energy backscattering, usually takes several seconds. The presence

Small-footprint LIDAR data comprise a set of return coordinates in three dimensions with each return usually carrying attribute values that relate either to that return or to the pulse from which the return was generated.

of an isolated, single high-reflectivity target scanned in one flight line can thus lead to substantial discrepancy in the mean intensity of returns on the overlapping part of two adjacent flight lines. Additional, object-independent variability in intensity values is introduced by suspected fluctuations in the energy emitted by the laser instrument. Personal communication with scientists involved in LIDAR research have revealed that these fluctuations can sometimes amount to 30 percent of the mean pulse energy and that they are likely more pronounced for high-frequency systems. Although LIDAR instruments currently do not record gain factors and energy output levels, persistent user requests to enable their logging could facilitate intensity normalization in the future and thus improve intensity-based classification of objects. [1] Return intensity is recorded in 8 bits (values 1 to 255), 12 bits (1 to 1023), and less often as a fraction in the 0 to 1 range or in 16 bits (1 to 65535).

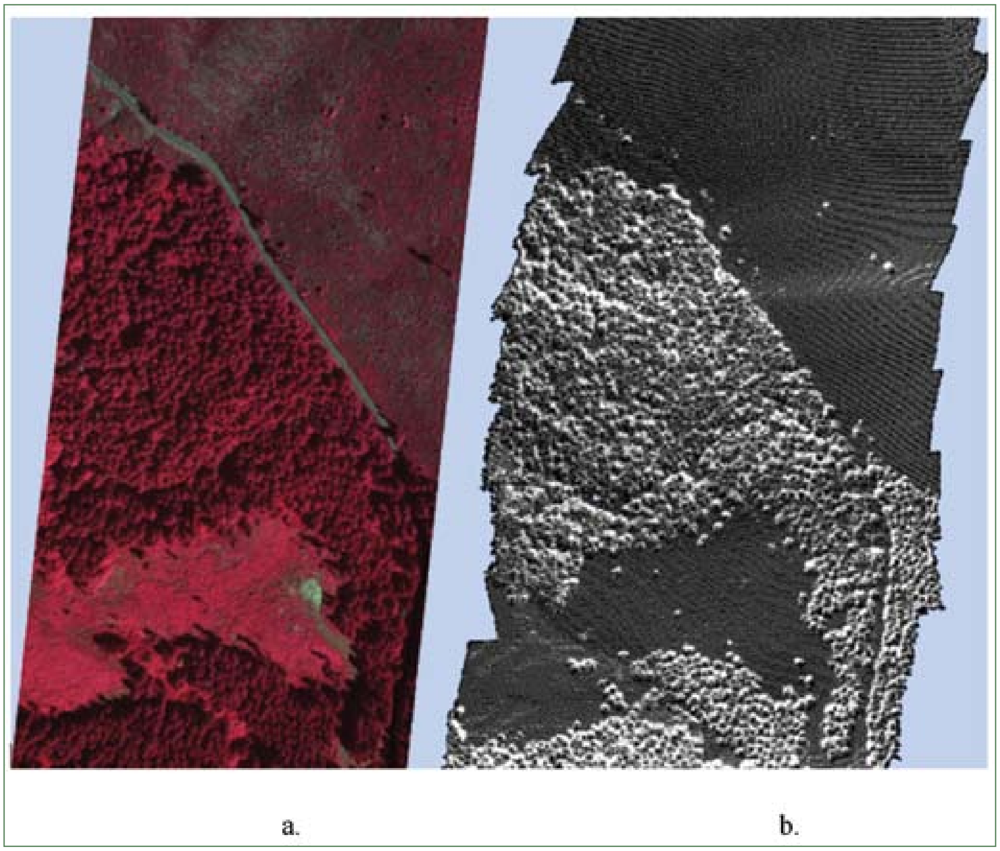


Figure 5—(a) False color (near infrared, red, green) digital aerial photograph and (b) corresponding gray-scale raster of LIDAR returns per square meter, with lighter tones depicting higher return count.

---

[1] Hyppä, J. Plenary session, 2007 International Society for Photogrammetry and Remote Sensing Workshop, Espoo, Finland.

- **Return number** refers to the rank of a return among those generated from one beam. It is meaningful only for systems that support multiple returns per beam. The return number should not be confused with the **number of returns**, a beam attribute.
- Attributes that a return inherits from its parent beam include the **scan angle**, usually recorded in degrees; the **end-of-scan-line**, a binary (true/false) attribute indicating whether the parent beam marked the edge of a scanning line; and those sometimes assigned at the data postprocessing phase such as indices to **flight lines** or **classification** schemes, and **GPS time**, an indication of the precise time that a pulse was emitted. Provided sufficient precision is used for storing GPS time, this attribute can be used as a unique identifier for a pulse.

Additional information is usually organized in the form of metadata, and often contains spatial geographic information system (GIS) layers with the spatial extent of the data acquisition, flight lines, the date and time range, the model and characteristics of the LIDAR instrument, etc.

Data Storage

The LIDAR data files are very large and can quickly fill up computer hard drives. The need for efficient access to and storage of scan data, coupled with the absence of a universal format standard, has led developers of LIDAR software to implement their own, proprietary storage format, which, with few exceptions, pay little attention to enabling import/export options. Only recently a file format (LAS) endorsed by the American Society for Photogrammetry and Remote Sensing (ASPRS) has been gaining popularity and support. As revealed in personal communications with several LIDAR data vendors across the United States in the last 2 years, the lack of significant progress in format standardization has prompted data delivery requests in ASCII (text) format in more than two-thirds of all acquisitions. Data delivered in most of those acquisitions consisted of X, Y, and Z coordinates and intensity only. This preference for the text format is rooted in the fact that, unlike any binary alternative, the contents of text files are easily accessible via a text editor. Assuming delimited format (text, space, tab, etc.) and that each file line carries data for one return, the data can be easily imported into popular databases and subsequently queried, merged, grouped into subsets, and rearranged as needed.

However, ASCII text is a poor format choice from the standpoint of data storage efficiency. To illustrate this issue, consider a LIDAR data file comprising a modest 1 million returns with coordinates of two-digit (centimeter) precision (universal transverse mercator projection) and 8- bit intensity being the only return attribute.

The size of this file will be approximately 32,134,000 bytes in text format and only 14,000,024 bytes in binary format (24 bytes are used to describe a transformation of scale in return coordinates from a two-decimal real number to a long integer), a gain in storage efficiency by a factor of 2.3. If the same file were to include all data attributes mentioned in the previous section, its size in text format would be approximately 64,094,000 bytes, and in LAS binary format it would be 28,000,227, a storage efficiency gain of also 2.3. In either file configuration (intensity only vs. all attributes), the time required for reading from or writing to the file in text format would be, depending on the hardware configuration of the computer, nearly an order of magnitude longer than for binary format. Efficiency in accessing files is important in research efforts and in applications that require files to be read multiple times.

Efficiency in accessing files is important in research efforts and in applications that require files to be read multiple times.

A less-evident implication of the file format is realized when considering how LIDAR data are organized in individual files. A LIDAR data file would typically contain returns either from a rectangular portion of the acquisition area, sometimes referred to as "bin" or "tile," or from individual flight lines (fig. 6). Compared to files representing smaller bins, files corresponding to larger ones will have a smaller percentage of returns near the borders of the bin, and thus introduce fewer

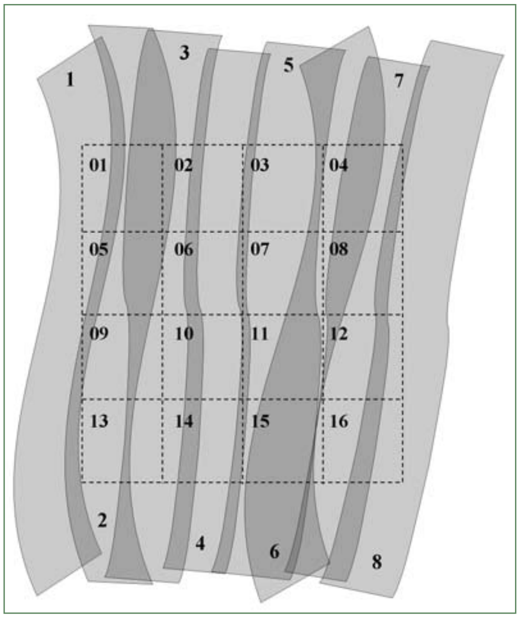


Figure 6—Illustration of the spatial extent of data in LIDAR files. Shaded, single-digit-numbered stripes correspond to individual LIDAR files with data from one flight line. Odd- and even-numbered stripes are flown in opposite directions. Darker shading shows stripe overlap. Dash-outlined, two-digit-numbered rectangles correspond to files containing returns from multiple flight lines.

discontinuities or artifacts in data derivatives and metrics calculated along the bin borders. Assuming an interest in minimizing border effects, maximum bin, and therefore file, sizes should be targeted. Table 1 shows the limits in file size and corresponding bin area imposed by a 32-bit computer operating system for various data storage formats and return density configurations. All data from an acquisition with a mere 100 million returns can be stored in just one binary file. If, instead, text format is preferred, the data would have to be split into two or more files. Note that switching from a 32- to a 64-bit operating system would eliminate this issue, as the file size supported by 64-bit operating systems is practically unlimited.

Data acquisition planning should be based on a careful evaluation of the project objectives while considering potential limitations imposed by budget constraints, availability of LIDAR instruments with specific capabilities, terrain, and vegetation structure and phenology.

Data Acquisition Considerations

Data acquisition planning should be based on a careful evaluation of the project objectives while considering potential limitations imposed by budget constraints, availability of LIDAR instruments with specific capabilities, terrain, and vegetation structure and phenology. Often, acquisition planning is challenging, as it involves many decisions among equally appealing or contrasting tradeoffs. The discussion below provides a synthesis of LIDAR data analysis objectives and their relation to system specifications and acquisition parameters.

Table 1—Attributes of a single LIDAR data file

Operating system	File format	Return attributes	Number of returns	Density (returns /m^2) 1	4	8
				- - - - Bin area (ha) - - - -		
32-bit	Text	X, Y, Z, intensity	133,658,000	13,366	3,341	1,671
	Binary	X, Y, Z, intensity	306,783,000	30,678	7,670	3,835
	Text	All	67,010,000	6,701	1,675	838
	Binary	All	153,392,000	15,339	3,835	1,917
64-bit	Any	Any	Practically unlimited			

Note: Text format assumes universal transverse mercator coordinates with 2-digit precision and 8-bit intensity.

Quantification of forest structure and assessment of tree height and volume via LIDAR data is typically performed either at the individual tree or plot/stand level. There is general agreement among researchers that the identification of individual trees requires a minimal return density of approximately four returns per square meter. This density often implicitly assumes systems that support multiple returns per pulse. High-scanning-frequency systems can achieve this density when using aircraft that fly high and fast to reduce acquisition costs. However, two data sets with equal return density acquired over the same area by instruments operating at different scanning frequencies can have very different return distributions in three dimensions. This is in part because the energy carried by a single pulse in high-scanning-frequency systems, and therefore its ability to penetrate vegetation, is much lower than the energy of a pulse in a slower system. The high-scanning-frequency system should be expected to generate proportionally more returns from the upper part of the canopy. Conversely, the low-scanning-frequency system will likely have a higher proportion of returns from the understory or the ground. In forest stands with tall and very dense vegetation, common in the Pacific Northwest (PNW), it is likely that a lower frequency system could generate more ground returns than a faster system, even where the overall density of the faster system is much greater than the density generated by the slower system. Although these assumptions have not been tested formally, they are indirectly supported by the fact that the proportion of ground-to-total returns generated by low-frequency laser systems mounted on low-flying aircraft or helicopters over high-density tropical forests (Clark et al. 2004) is much higher than the one achieved over comparably dense PNW forests scanned by high-frequency systems flying approximately 1000 meters above terrain (Gatziolis 2007). Hence, where a precise and accurate description of the ground surface under dense vegetation is important, the option of using a lower scanning frequency setting should be seriously considered.

Although the three-dimensional distribution of returns is also affected by the discretization process, the absence of specific information on the settings of alternative systems usually precludes a meaningful evaluation of comparative advantages offered by each system. Fine sensitivity in pulse discretization, that is to set the minimum that the amplitude of the backscattered pulse energy would need to exceed for a return to be identified to a low value (fig. 4), will tend to produce returns closer to the top of the canopy and support a more detailed description of vegetation surfaces, including leader stems typical of many conifer species. Fine sensitivity, though, is associated with lower positional precision of returns from lower vegetation strata. Fine distance tolerances between consecutive returns from a given beam would tend to produce more returns from the upper layers of tall, dense, and healthy vegetation at the expense of fewer returns from the ground. Coarse distance tolerances would prevent ground returns where the magnitude of the tolerance exceeds the mean vegetation height.

As stated in the “Data Attributes” section, the local return density would differ among vegetation types and structures. To avoid misunderstandings, data acquisition requests should specifically mention the minimum pulse density (not return density) that is acceptable over a particular forest or land type. Data vendors with experience in local acquisitions will likely be able to assess the flight height above ground and aircraft speed for which their laser instrument can meet the requested pulse density over a forest type of interest.

To determine the preferred beam divergence setting (wide vs. narrow), one should consider how this setting affects the interaction of the beam with vegetation. In wide divergence, the canopy volume illuminated (or sampled) by the beam is larger than in narrow divergence. The ratio of canopy volume "sampled" by each setting is actually the square of the divergence ratio (table 2). Because wide divergence affords a more comprehensive coverage of the canopy, sampling plot- or stand-based metrics computed from wide divergence LIDAR data would likely be more robust and exhibit lower variance than those computed using narrow divergence data. Because, as stated previously, in wide divergence the cross-section pulse energy is spread over a larger area, the reduced photon density leads to a lower signal-to-noise ratio in backscattered energy. This reduces the three-dimensional precision of returns, and often causes backscattering from leader stems in coniferous species or from the ground in dense stands to be of an amplitude too weak to be identified as returns during the pulse discretization process. Recent research in the Pacific Northwest has demonstrated that the accuracy of LIDAR-based individual conifer tree height measurements obtained at a narrow beam divergence (0.33 m) setting are significantly more accurate than those obtained at a wide divergence setting (0.8 m) (Andersen et al. 2006).

Table 2—Vegetation volume illuminated by two LIDAR beam divergence settings[a]

	Beam cross section area at distance (m) from LIDAR instrument			Vegetation volume
	970	985	1000	
	Square meters			*Cubic meters*
Wide divergence (0.7 millirad)	0.362	0.373	0.385	11.204
Narrow divergence (0.3 millirad)	0.067	0.069	0.071	2.058
Ratio				5.444[b]

[a] Flight height assumed to be 1000 m above vegetation, and vegetation is 30 m tall.

[b] Result (5.444) = $(0.7/0.3)^2$.

Before discussing how scanning angles affect the fidelity of LIDAR data, it is necessary to describe a phenomenon known as "path reflectance" or "multipathing," a term borrowed from optical remote sensing and GPS technologies. Energy propagating through a medium is subject to attenuation, or reduction in both density and amplitude owing to scattering, absorption, and reflectance. The attenuation of near-infrared pulses, although minimal in the atmosphere, can be substantial in porous and heterogeneous media, such as the forest. Of interest here, is the change in propagation (or path) direction that sometimes occurs when the beam hits the surface of objects with considerable mass, such as tree branches, trunks, rocks, or the ground. How often a change of direction occurs is believed to correlate to the angle of incidence and, to a lesser extent, to the distance the pulse travels through the canopy. As mentioned earlier, the LIDAR instrument records the pulse direction vectors and the time difference between pulse emission and return. The identification of individual returns performed by processing the backscattered energy is based on the premise that the pulse has traveled along a straight line, an assumption that is true for the majority of the pulses. For pulses with one or more changes in direction, the actual path remains unknown. Returns from these pulses are recorded as originating further away from the LIDAR instrument than their true, albeit unknown, locations (fig. 7).

As the direction of the beam deviates progressively more from nadir, the beam's angle of incidence upon the ground increases. This causes the ground to behave less as a diffuse and more like a directional reflector, thereby facilitating path reflectance. The distance the pulse travels through the canopy increases too. In acquisitions on a slope, both of these measures can increase even further. For example,

Figure 7—Illustration of LIDAR path reflectance. Black line represents the beam propagation anticipated by the LIDAR instrument. Red line represents the actual beam propagation. Letters A, B, and C indicate the locus of anticipated return, first, and second beam inflection points, respectively.

the incidence angle to the ground of a beam emitted with a scanning angle of 12 degrees down a 100-percent slope will be 57 degrees. The distance traveled by the beam through 30-m-tall vegetation in such terrain would be 38.95 m, an increase of nearly 30 percent compared to the distance the beam would travel through the same canopy at zero degrees scanning angle. Therefore, the canopy penetration rates for laser pulses in a forest area are directly related to the off-nadir scanning angles of the laser pulses.

Experience has determined that LIDAR data artifacts attributed to multipathing, such as returns located well below the ground, proliferate when scanning angles exceed 12 to 14 degrees over dense forest stands. For that reason, some data vendors tend to remove from delivered data sets returns from beams at large scanning angles. Others are reluctant to eliminate those returns, especially where the resulting return density nears the minimum specified. Maximum absolute scanning angles of 12 degrees for flat or moderate-slope areas and 10 degrees for areas with steeper slopes should produce scan data with minimal scan-angle-related artifacts.

Decisions on acquisition timing should be based on the seasonal progression and phenological stage of vegetation while considering local weather patterns and terrain.

Decisions on acquisition timing should be based on the seasonal progression and phenological stage of vegetation while considering local weather patterns and terrain. In the absence of disturbances, seasonal progression determines the development stage and density of foliage and, hence, canopy penetrability by LIDAR pulses for deciduous cover types. In practice, flight timing has little effect on coniferous canopies unless there is a deciduous understory. Where extraction of a high-fidelity digital elevation model (DEM) is one of the products the laser acquisition is expected to provide, it is advisable to consider a leaf-off acquisition even if deciduous species are only present as shrubs or brush. Note that often the extraction of a DEM, if not of importance by itself, is an intermediate LIDAR data analysis step necessary for the assessment of many stand or tree parameters, including canopy/crown cover, tree height, and canopy base height.

Seasonal progression can be essential in research efforts designed to achieve near-simultaneous collection of laser and field data. Sometimes it would be necessary to “grow” the field data forward or backward in time to ensure a close temporal match with the laser data. Seasonality also affects the values of indices or output of established models that assess vegetation parameters by assuming, often implicitly, a certain vertical vegetation structure that is seasonally dependent. Many of the models used to estimate stand volume, basal area, and mean height, and even indices of fuel accumulation and stand susceptibility to fire, rely on the proportion of total returns present in selected quantiles or percentiles of vegetation height. Foliage development as the growing season progresses would tend to increase the percentage of returns at higher height percentiles, thereby altering model or index values. When uncertainty exists about whether assumptions related to foliage maturity are embedded in models expected to be used with the data, acquisitions toward the end of the growing season would be preferred.

Unlike most forms of optical remote sensing, acquisition of LIDAR data is practically independing of solar illumination and could, in fact, be done at night when lower windspeeds interfere less with aircraft handling.

Unlike most forms of optical remote sensing, acquisition of LIDAR data is practically independent of solar illumination and could, in fact, be done at night when lower windspeeds interfere less with aircraft handling. Wet, rainy days should not be considered because infrared light does not penetrate water vapor. Persistent, visibility-limiting fog or frequent rainfall may not permit acquisitions during certain, predetermined periods. Flying the aircraft at constant aboveground height so as to maintain uniform return density, a task easily (and safely) accomplished over flat areas, can be challenging over undulating or mountainous terrain, especially during windy or foggy days. These considerations indicate that, for example, LIDAR data acquisitions during leaf-off conditions in the coastal PNW region might be delayed if the acquisition window happens to coincide with a period of prolonged low-ceiling fog or continuous rain, conditions typical of the region during the winter months. Similarly, the summer window for obtaining scan data at high-elevation, snow-free conditions can shift and be of different length between years. Note also that as many LIDAR data vendors share aircraft and instruments, it is more convenient for them to plan a mission within a larger time window. Therefore, planning large time windows for acquisition will likely provide flexibility with logistical and weather-related limitations, but it may also hinder efforts to coordinate the flights with field data collection. Perhaps the best approach to optimizing the timing for laser data acquisition is to have, where possible, the data vendor be awarded the acquisition contract well ahead of the anticipated flight day(s). This allows ample time for the vendor and user to communicate concerning the number of flight days, season progression, and vegetation conditions (tree leaf-out, senescence, etc.), and ultimately optimize the acquisition timing.

Quality control for a LIDAR data set involves an evaluation of return coordinate accuracy and precision, compliance with acquisition specifications, and data spatial consistency and completeness.

Data Quality Control

Quality control for a LIDAR data set from a user's perspective involves an evaluation of return coordinate accuracy and precision, compliance with acquisition specifications, and data spatial consistency and completeness. Ideally, a report with quantitative estimates of these data quality measures would be part of every laser data delivery. Unfortunately, such reports are rare, and when they are produced, the information included is frequently selective or incomplete, in part because, as discussed below, the procedures involved in evaluating data quality are costly and time consuming, especially for data over mountainous forested areas.

Return Coordinates

Although the positional accuracy and precision of LIDAR data far exceeds those afforded by traditional remotely sensed imagery, laser return coordinates do contain random and systematic errors. Both error types originate in one or more of the laser system's components. Random errors relate to noise in the computed location of the aircraft, in the recording of aircraft attitude and scanning angles, or in the recording of time between pulse emission and backscatter reception, which ultimately determines the distance (range) to the target. The magnitude of random errors can be calculated during system calibration. Because LIDAR systems get "out-of-tune" fairly quickly, periodic calibration is necessary. Assuming that the magnitude and distribution of errors from each laser system component is known, error propagation techniques can be used to estimate the nominal coordinate error and determine if it is acceptable for an application. It should be noted that the error estimates obtained via propagation techniques refer to returns from hard surfaces on flat terrain, conditions uncommon in forests of the Pacific Northwest. Hence, even in an optimally and recently calibrated system, random errors in derived return coordinates over forested landscapes can be far from negligible. Random errors are known to affect the absolute accuracy of return coordinates, but, perhaps against commonly held expectations, they also affect their relative accuracy. The latter term describes the integrity in the spatial arrangement of neighboring returns. As an example, the same amount of noise in the measurement of the aircraft's roll, one of the attitude vectors, will cause errors along the edges of the scanning swath to be larger than those in the middle, thereby degrading the relative accuracy of the return data.

Biases in GPS, aircraft attitude, scanning angle, and time measurements cause systematic errors.

Biases in GPS, aircraft attitude, scanning angle, and time measurements cause systematic errors. If erroneously measured, the three-dimensional offset between the onboard GPS unit and the pulse emission point, would cause errors that are independent of the above-terrain flight height, but that do depend on the flying direction. Errors in measurement of the scan angle can cause return-coordinate errors that increase with flying height and flight direction. Where the above-terrain flight height changes continuously and in the presence of slopes, the magnitude and type of error embedded in return coordinates is spatially variable.

As is evident from the brief discussion above, the amount of complexity in the assessment of the error budget in a laser data set is substantial. Experience suggests that the use of error propagation techniques and system calibration data alone tend to underestimate coordinate error in forested landscapes. More accurate assessments of the error budget require field surveying of the ground and of objects that are clearly discernible in visualizations of the return cloud, or in derivatives of the laser data. In mountainous areas or forests of heterogeneous structure, more surveying could be necessary. Whether the benefits from an accurate assessment of coordinate error warrant the expense to obtain it should be determined by the project objectives. The decision should consider that often the fidelity of data products actually depends on the relative accuracy of return coordinates, even in the presence of sizable, absolute errors.

Relative accuracy of LIDAR return coordinates—
The assessment of relative accuracy is possible only between data obtained in overlapping swaths. It has the advantage that it can be performed exclusively by data postprocessing in the office. Generally, the higher the swath overlap (over a minimum of 20 percent), the more precise the assessment would be. There are two approaches in computing relative accuracy, with different degrees of complexity. The first approach relies on computing raster, or triangular irregular network (TIN), surfaces of elevation or intensity for each (or part) of the two flight lines and then evaluating their spatial correspondence (Gruen and Akca 2005, Maas 2002, Okatani and Deguchi 2002). Often the evaluation is based on calculating the raster difference. A mean value other than zero or a skewed distribution for values of the difference raster is indicative of bias in the measurement of range. The presence of clusters of cells with values consistently positive or negative and of shape that resembles dominant landscape objects or structures indicate bias in the recording of aircraft location. Figure 8 illustrates the latter case showing the highest return digital surface models computed for the area common to two overlapping scanning swaths (8c and 8e) and their difference surface (8d). The 0.013-m mean value of the difference raster is reduced to 0.004 m when considering only returns from hard, impermeable surfaces (rooftops, paved road). This level of range error between scanning swaths is consistent with the magnitude anticipated by error propagation for range measurements. Unlike the high precision in range measurements, the difference raster in figure 8d shows substantial bias in two-dimensional (horizontal) space. The bias is evident in the vicinity of human-made structures. The difference raster has negative values in the southwest corner of a building (blue arrow), positive values on the opposite corner (red arrow), and nearly zero in between. The amount and direction of the horizontal discrepancy between the two surface rasters can be assessed approximately with measurements performed on a computer screen or, alternatively, by using scripts that compute the offset at which the normalized correlation of the surfaces is maximized. A more precise assessment can be accomplished by digitizing the boundaries of structures and then computing the rotation, translation, and scale adjustments required for aligning the boundary vector pairs in space. Likely the most precise assessment of positional accuracy among all methods

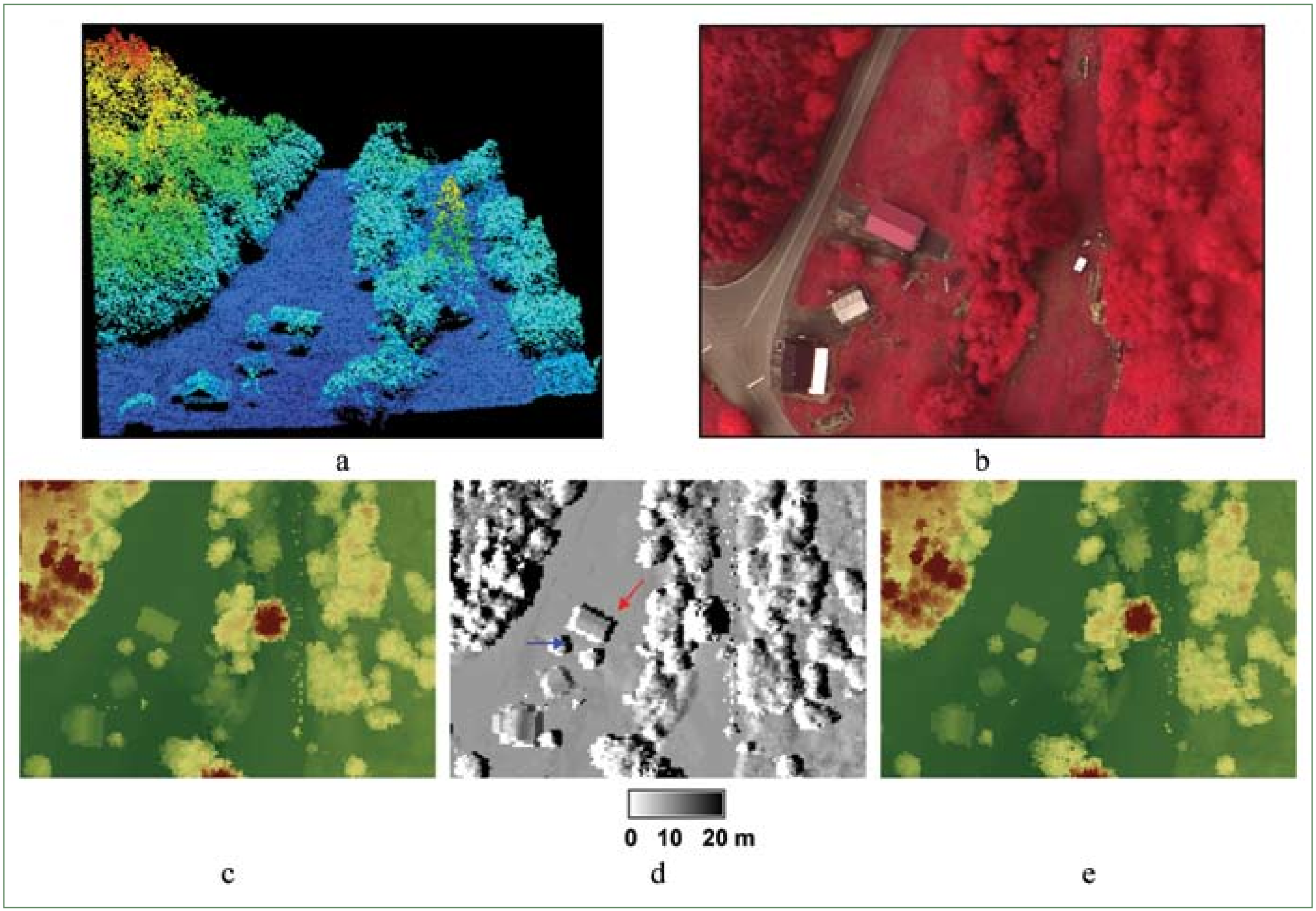


Figure 8—(a) Perspective view of return cloud colored by elevation, (b) false-color digital orthophotograph, (c, e) 1-m highest return rasters computed from the point clouds of two overlapping flight lines (darker tones denote higher elevation), and (d) difference raster computed by subtracting raster in (e) from raster in (c).

that rely on intermediate raster data derivatives can be achieved by comparing the two surfaces in the frequency domain (e.g., via Fourier transforms).

Raster-based evaluations of positional error in laser data have a shortcoming that is rooted in the form of the return data. Unlike typical remote sensing products, return clouds are abstract and cannot be easily compared to discrete objects. Consider, for instance, the edge of one of the roofs visible in figure 8a or 8b and the footprint of a pulse that hits that edge. Assuming negligible random error in the data, we would expect a return right on the edge of a roof if exactly half of the pulse’s footprint intersected the roof plane, or, in other words, if the center of the footprint would coincide with the edge line. This seldom is the case. Instead, there would be many returns with elevation value equal to that at the edge of the roof, some of them positioned (in two dimensions) over the roof plane and the others hovering in midair next to the structure. Higher return densities, narrower beam divergence settings (smaller footprints), longer linear features of objects, and a higher frequency of objects of a given size promotes more precise feature extraction for an object of a given size. Because the positional uncertainty of objects delineated on elevation raster surfaces is proportional to the area of a raster cell, determining the optimal resolution for the raster is of importance. Raster resolutions resulting in an average of four to six returns per cell are regarded as appropriate for evaluating, and ultimately improving, the relative accuracy of a laser data set.

Regardless of the exact methodology used, the assessment of relative accuracy in LIDAR data is susceptible to artifacts or errors introduced during the creation of the derived products, such as raster surfaces or object segmentations.

Regardless of the exact methodology used, the assessment of relative accuracy in LIDAR data is susceptible to artifacts or errors introduced during the creation of the derived products, such as raster surfaces or object (i.e., individual tree) segmentations. Often the algorithms used to produce these derivatives are complex and highly sensitive to the input parameters. These postprocessing errors sometimes could generate the impression that the data are of poor relative accuracy when in reality they are not. To avoid such pitfalls, sufficient experience and adequate understanding of the processes involved in generation and quality assessment of derived products is essential. Fortunately, the magnitude of the errors can be deduced with precision that increases proportionally to the number of objects identified and the area each occupies. However, prematurely increasing the area over which a relative accuracy evaluation is performed could lead to erroneous or biased results because, sooner or later, either the system's acquisition parameters will cease being stationary or the spatial distribution of objects will change.

The second approach used for assessing the relative accuracy of a laser data set requires computation of a surface for only one of the two overlapping scanning swaths. The other remains in its original return cloud form. Unlike the first approach, the highest-return surface is now represented as a triangulated irregular network. Spatial correspondence is evaluated by examining how well the network triangles fit to the return cloud (Lee et al. 2005). The advantage of this approach is that (a) it does not require decisions on the appropriate resolution of computed surface rasters; (b) it is found to work well with continuous canopies, sparse forest, or in the presence of human-made objects; and (c) it can be used over small areas. The drawback to the point-based methods is that there is some speculation on how to determine which returns should participate in the evaluation of fit for each network triangle. As this approach is an active research project, refinements in its application should be anticipated.

Absolute accuracy of return coordinates—

The evaluation of relative positional accuracy is usually followed by an evaluation of absolute accuracy using information provided by a high-precision field survey. In most acquisitions familiar to the authors in the PNW, the ground survey, if performed at all, is a single transect along a road free from overhanging vegetation.

Where the area was inaccessible to vehicles, the surveyed transect was outside the acquisition boundary. The absolute accuracy report for those acquisitions was based on a comparison of transect coordinates with those of co-located (in two dimensions) or nearly co-located returns. Given that horizontal discrepancies between two point domains or coordinate systems cannot always be detected or measured along a transect, linear surveys can support evaluations of only absolute elevation accuracy. Discrepancies owing to a shift of the return cloud parallel to a plane that approximates the local ground surface will be missed. A three-dimensional evaluation would require intersecting transects or, alternatively, placing targets with characteristic shape and reflectance that clearly distinguish them from their surroundings within the acquisition area. Both of these options are costly and often practically difficult to implement. Note that absolute accuracy reports based on a single survey may not be representative of the acquisition area. Precise, kinematic-GPS-based, surveys of bare ground performed on multiple intersecting transects in a study area in coastal Oregon showed that although the relative accuracy between overlapping scanning swaths was often nearly constant throughout the acquisition area, the absolute accuracy varied substantially. In a handful of locations, it was almost an order of magnitude lower (worse) than the one mentioned in the report submitted with the laser data. The lower accuracy was attributed to aircraft location bias introduced when the aircraft was engaged in rapid ascending or descending.

The relative accuracy of LIDAR data is more important than the absolute accuracy in most forest applications.

The relative accuracy of LIDAR data is more important than the absolute accuracy in most forest applications, including estimation of density, structure, basal area, volume, etc. Given the cost of quality control, investments in evaluation of and measures to promote the relative accuracy will likely be more beneficial to the data user than those targeting the absolute accuracy. This general rule does not necessarily apply to data acquisitions for topographic mapping, research purposes, or studies investigating forest growth and change with time. Nevertheless, the reader should be aware that laser data with even suboptimal return coordinate accuracy registration will be far superior, in terms of registration and internal consistency, to any other form of remotely sensed data.

Spatial Completeness

The implications from lack of continuity or scanning uniformity in a LIDAR data acquisition differ. They can range from simply being sources of artifacts and local variability in data derivatives to precluding data analysis. Even sporadic discontinuities in the laser data could, for example, prevent a successful delineation of the drainage network or the computation of landform and vegetation structure metrics. Figure 9 shows a case from a laser acquisition, where, in the interest of minimizing costs, the sidelap of adjacent scanning swaths was set to only 10 percent. Thanks to changes in aircraft attitude, a noticeable portion of the acquisition area was not scanned.

The absence of scanning uniformity across the acquisition area is demonstrated by variability in pulse density or return density over the same or similar objects and vegetation. It occurs where the distance between adjacent flight lines or the flight height above the ground does not remain constant (fig. 10), or, as shown above (fig. 9), in the presence of pronounced instability in the aircraft attitude vectors. Local fluctuations in return density theoretically can be prevented over flat or undulating terrain if the percentage of sidelap between adjacent scanning swaths can be represented by the formula N × 100/(N+1), where N is an integer greater than 0. The value of the formula for N = 1 is 50 percent, the theoretical minimum sidelap that produces scanning uniformity. For N = 3, the amount of sidelap would be 75 percent. Larger N values produce unnecessarily high sidelap amounts. If the amount of scanning swath sidelap specified does not comply with the formula, scanning uniformity cannot be achieved. For instance, sidelap of 60 percent corresponds to N = 1.5, which is not an integer. In that case, half of the acquisition area will be

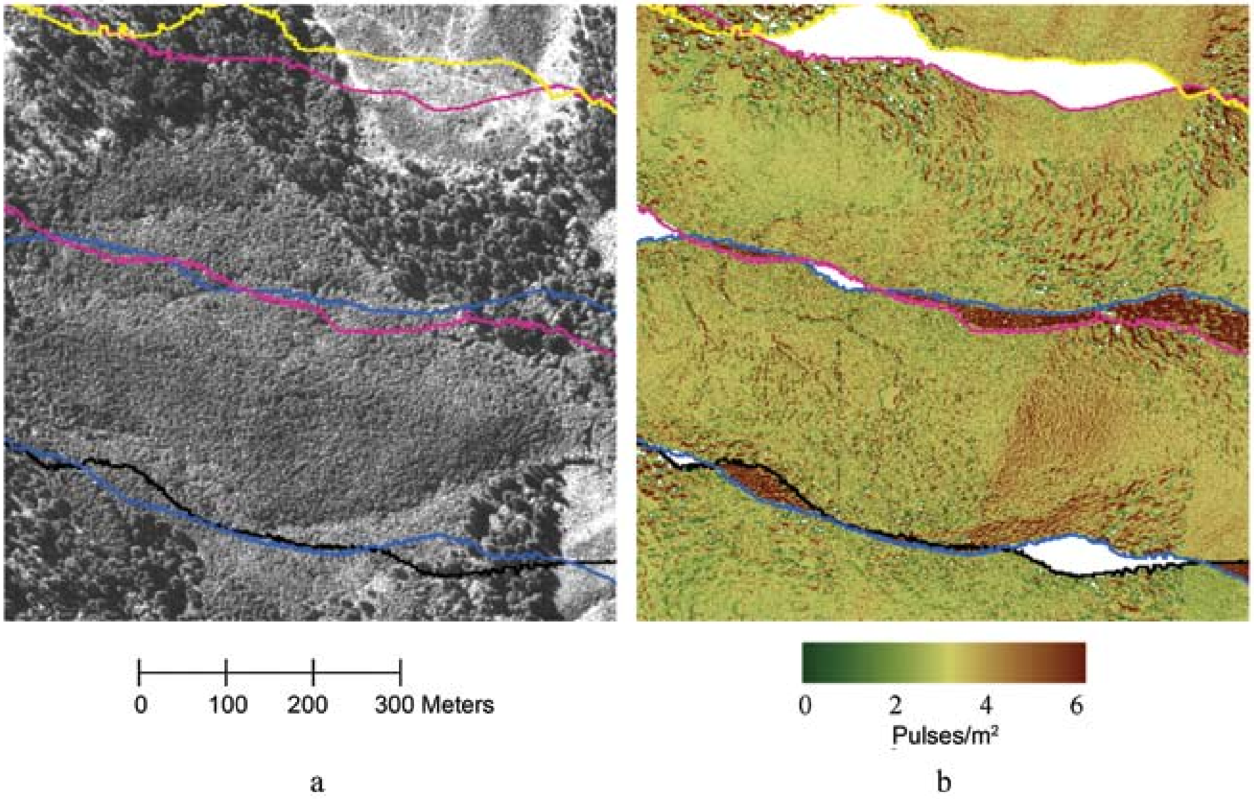


Figure 9—(a) Gray-scale panchromatic orthophotograph, and (b) raster of laser pulse density. Color lines denote the spatial extent of laser scanning swaths.

represented in two scanning swaths and the other half in three scanning swaths. At sidelap of 30 percent, 57 percent of the acquisition area will be represented in a single scanning swath and the remaining 43 percent in two scanning swaths. In practice, constant return density even over homogeneous vegetation is difficult to achieve, and some variability should be anticipated and tolerated.

Interesting insights on how aircraft attitude affects the laser scanning can be obtained by examining figure 10. It can be seen in the figure that a few flight lines almost coincide, whereas the others are regularly spaced. This is because persistent atmospheric turbulence during data acquisition that translated into nearly constant changes in aircraft pitch and roll had caused the scanner to miss many areas, despite the intended 50-percent swath sidelap. To the credit of the data vendor, the lack of spatial completeness was identified and certain lines were flown a second time. Although the additional flight lines covered all the areas initially missed, the minimum pulse density (2 pulses/m^2) requested in the acquisition specifications was not met everywhere. Areas scanned in swaths with excessive meandering, denoted by “I” in figure 10b, were more susceptible to not meeting the pulse density standard. Areas represented in many swaths had pulse density several times higher than the one requested.

Variability in aircraft pitch also affects the local pulse density, which when mapped, exhibits a striping effect oriented perpendicularly to the flight line. When the aircraft flies horizontally (zero pitch), the forward propagation of successive

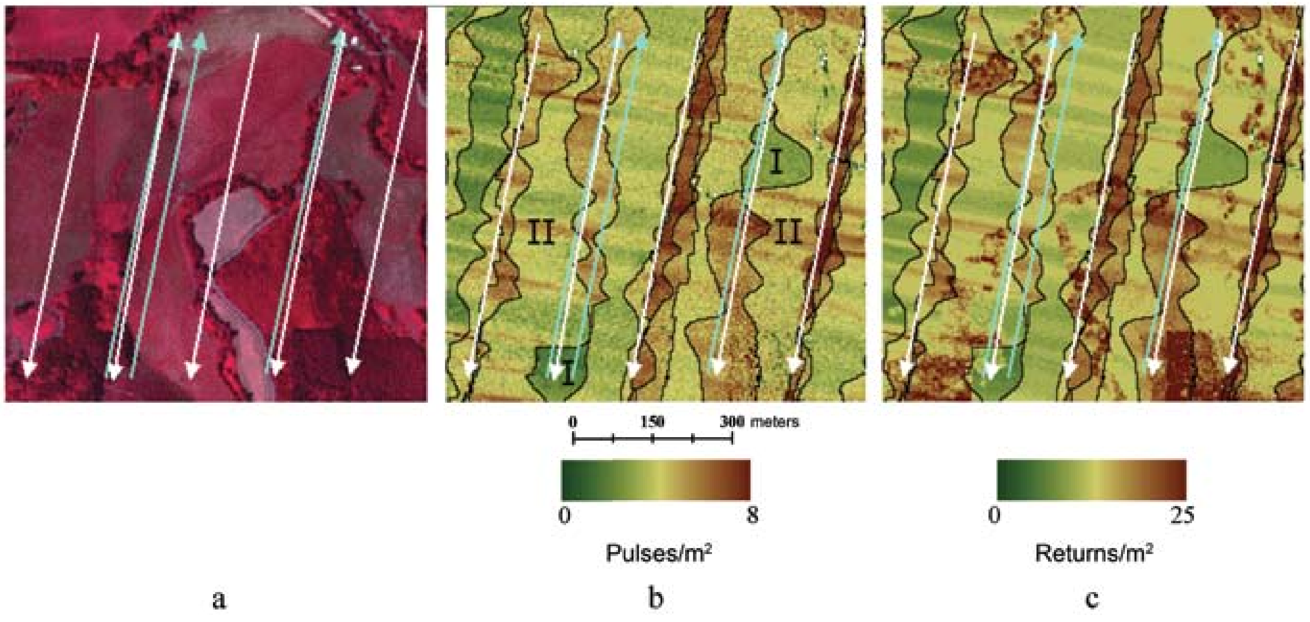


Figure 10—(a) False-color, high-resolution orthophotograph showing direction-colored flight lines of laser data acquisition mission, (b) raster of laser pulse density, and (c) raster of laser return density. Black lines on the rasters denote the spatial extent of individual scanning swaths. Areas identified with roman numerals demonstrate the effects of substantial instability in aircraft roll (I) and pitch (II).

scanning lines and the distance between them is constant, although some minimal variability could occur on steep ridges and valleys. A sudden reduction in pitch (aircraft dives) slows the forward propagation of the scanning lines and, if large enough, it can even reverse it momentarily. Similarly, a swift increase in pitch (aircraft ascent) accelerates the forward propagation of scanning lines and increases the distance between them. The result is sporadic, higher-than-average pulse density areas (figure 10c, II) usually followed, along the flight line, by below-average density areas. The combined effect of roll and pitch changes on the local return density for a single flight line is shown in figure 11. The area depicted is covered with continuous forest over rolling terrain and the overall laser return density is 4.8/m^2, higher than the 4 returns/m^2 anticipated. In this example, concentration of pulses in portions of the area owing to aircraft attitude dropped the return density well below the specification standard for approximately 60 percent of the area shown in the figure, with the reduction in density being more pronounced along the edges of the scanning swath. In the absence of pertinent documentation, the effects of locally variable pulse density of laser data derivatives can only be speculated. These effects are likely limited to areas with low pulse density and could perhaps be associated with an increase in the variance of biophysical parameter estimates computed at the individual tree level.

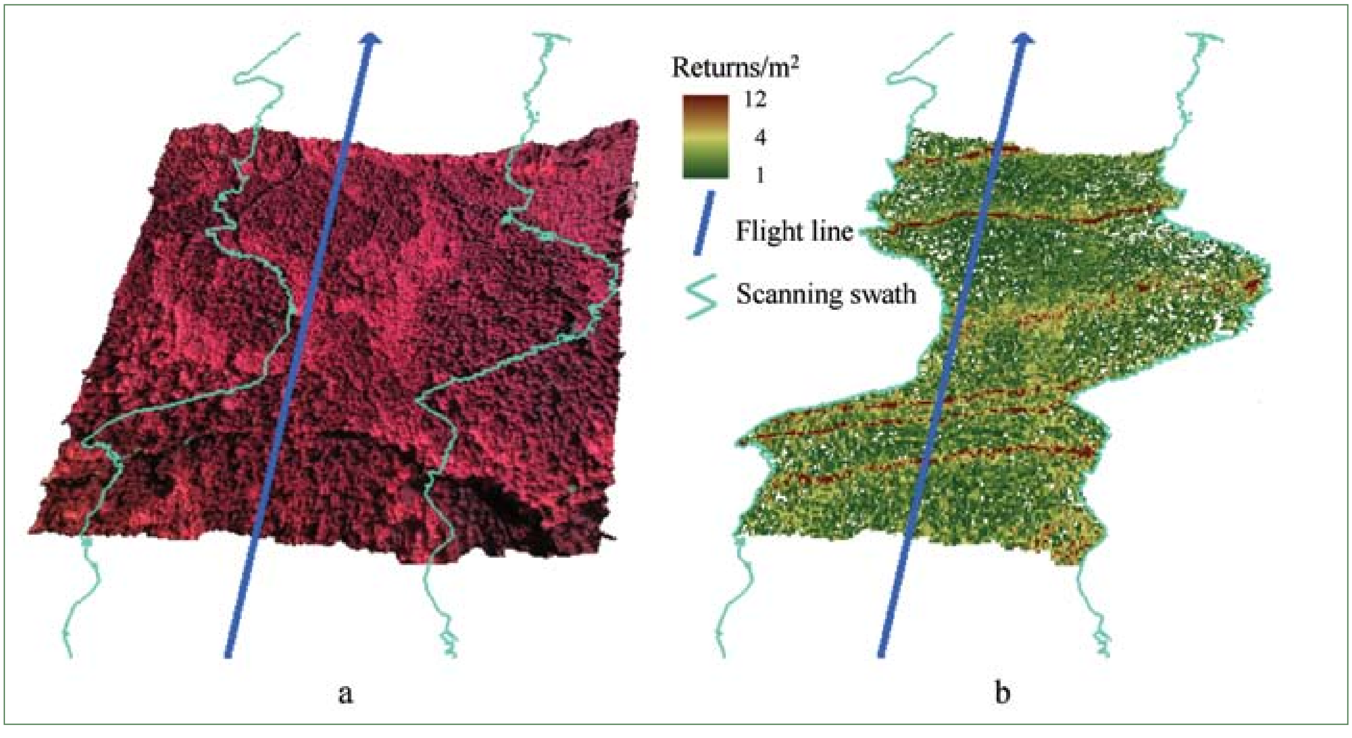


Figure 11—Perspective view of (a) a false-color, high-resolution orthophotograph of a forested area with a flight line and corresponding laser scanning swath superimposed, and (b) co-located raster of return density derived from the swath's laser data.

Another dimension in spatial completeness refers to the portion of returns that are removed by postprocessing prior to the data delivery. The majority of returns that are eliminated correspond to pulses with scanning angle exceeding the maximum specified, or those believed to be affected by multipathing and thus positioned below the true ground surface. Although the elimination of returns in the latter case is theoretically legitimate, the procedure used to identify them can contain flaws, at least in certain circumstances. It was found, for example, that proprietary algorithms used to classify points in the return cloud as above, on, or below the ground do not work particularly well in dense, coniferous forests in the PNW coastal region (Gatziolis 2007). Similar problems have been identified in the classification of returns acquired at leaf-on conditions over dense, multistory, deciduous stands. In those conditions, many returns eliminated as being below the ground because of multipathing, when compared with precise and accurate representations of terrain obtained by alternate means, were found to be actually above the ground. This topic is an active research field and new, better performing algorithms are introduced frequently. The legitimacy of returns flagged by the algorithm employed by a data vendor as belowground and eliminated from the delivered data set cannot be assessed by an alternative, perhaps superior algorithm, at a later time. Data users with interest in exploring alternative algorithms and data analysis techniques are advised to request from the vendors that all data initially flagged for removal from the delivered data set be organized in a second set to accompany the first.

Consistency of Tabular Return Attributes

Intensity is the most commonly requested and used return attribute.

Intensity is the most commonly requested and used return attribute. Often it is the only attribute delivered with the return coordinates. It is typically expressed as an integer, but sometimes is archived as a floating point number in the 0 to 1 range. The latter preference can introduce data consistency problems if the precision of the float numbers is not adequate or the scaling of the backscattered energy in the 0 to 1 range is data-tile or flight-line specific. Figure 12 demonstrates one such case showing the return intensity histograms from the overlapping portion of two adjacent flight lines, with the intensity values scaled linearly in the 0 to 1 range. Evidently, the shape of the histograms is the same, but the second histogram appears compressed. This is because the scaling of the intensity values was based on the maximum per flight line, with the maximum being much higher in one flight line than in the other. In the example of figure 12, the intensity values for the two flight lines can be adjusted via histogram matching. If, however, a delivered laser data set with inconsistently (flight-line-specific) scaled intensity is organized in tiles and without information on the membership of returns or pulses to individual flight lines, then

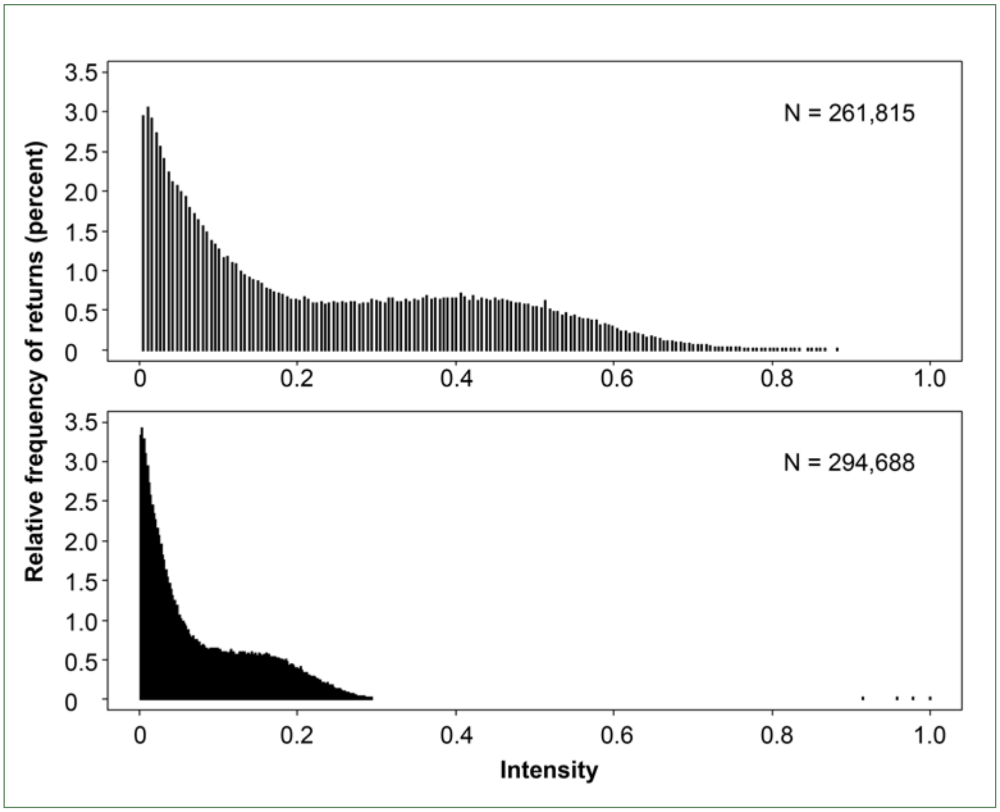


Figure 12—Distribution of intensity for returns in two overlapping scanning swaths.

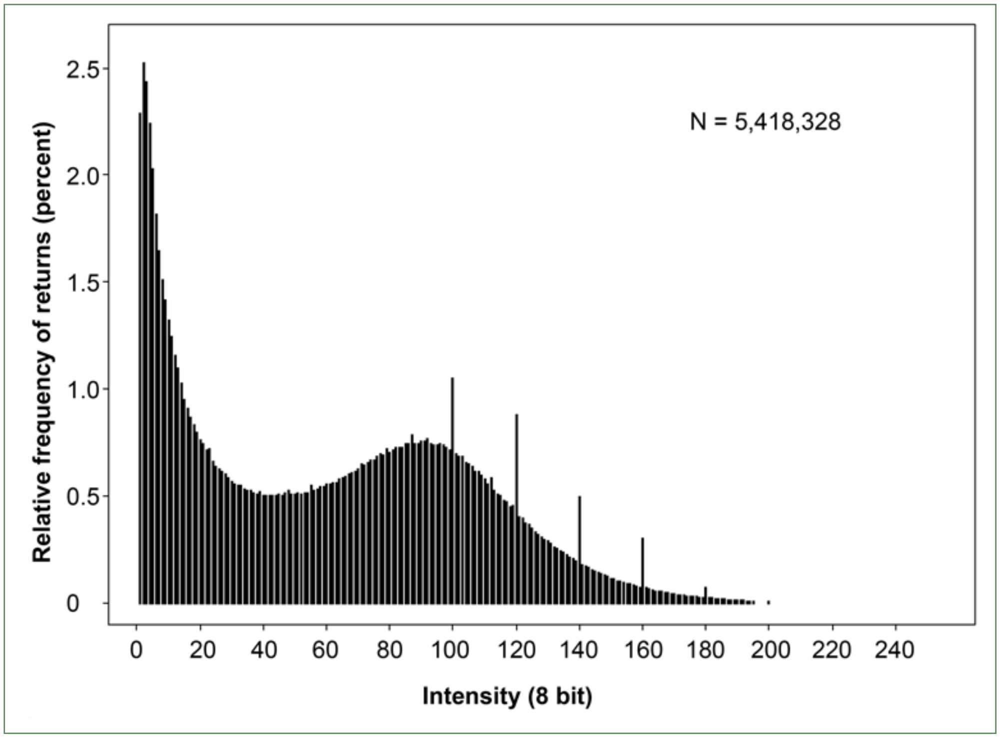


Figure 13—Histogram of 8-bit return intensity from 1 km^2 of forested land with homogeneous structure showing a few intensity values with abnormally high frequency.

there is little the analyst can do to restore the intensity data to a consistent scale and improve their utility.

A second issue with intensity data consistency, albeit of minimal implications on their utility, relates to the circumstantial reduction in the precision of intensity value recorded. It has been observed in nearly every data set examined by the authors. It does not seem to be associated with data acquisition conditions, the phenological stage of the vegetation, or the laser instrument used, and occurs only at higher intensity levels. This issue can perhaps be better understood by examining figure 13, which shows a histogram of return intensity values for 1 km^2 of structurally homogeneous, uneven-aged coniferous forest. Certain intensity values in the histogram, all pertaining to rounded numbers and of consistent increment between them (100, 120, …, 180), have frequency much higher than their neighboring values. About half of the returns in the histogram bins with higher-than-anticipated frequency were preceded chronologically by returns with very low intensity. One possible explanation for this phenomenon is an inability of the laser sensor to precisely quantify one or very few backscatter events among many others of consistently low energy. Another explanation is suboptimally performed discretization of the backscattered energy.

Inadequate precision has more serious implications in the recording of GPS time, a quantity that indicates the moment a pulse was emitted. The GPS time of a pulse or return can be used to determine the distance in three dimensions between the return and the antenna of the laser instrument, to establish the chronological succession of pulses and returns, or determine return membership in individual pulses. In older, lower frequency systems (< 10 kHz) four digits of precision in the recording of GPS time were adequate. For modern, high-frequency systems (> 100 kHz), at least six digits of precision are needed. Sometimes, an upgrade to a higher frequency system is not met by a proper modification in the precision of recorded GPS time. The effect of this omission is shown in table 3, which contains data from a laser acquisition performed by using a 71-kHz instrument. Recording GPS time with adequate precision (5 digits) allows us to deduce that the 17 returns belong to 13 different pulses (1001 to 1013), with four pulses represented by two returns each. Using inadequate precision (4 digits) would have indicated that the 17 returns belong to only two pulses (1001 and 1002) with each pulse generating 8 and 9 returns, respectively, a deduction that is absurd given that the system's configuration supported a maximum of four returns per pulse. Note that determining the precision of GPS time is easy when the data are stored in ASCII format. When examining binary data, it should be ensured that the data viewer or conversion-to-text utility employed supports the precision level anticipated. Otherwise lack of precision could be a software-introduced artifact.

Table 3—Laser return data showing effect of global positioning system (GPS) time precision in identifying return membership to unique pulses via data postprocessing

				5-digit precision		4-digit precision	
X	Y	Z	Intensity (8-bit)	GPS time	Derived pulse ID	GPS time	Derived pulse ID
804.60	359.55	219.12	76	37614.24936	1001	37614.2494	1001
804.13	359.99	222.21	18	37614.24938	1002	37614.2494	1001
803.90	359.58	219.23	56	37614.24938	1002	37614.2494	1001
803.23	359.66	219.72	52	37614.24939	1003	37614.2494	1001
802.55	359.71	219.99	61	37614.24940	1004	37614.2494	1001
801.88	359.79	220.43	57	37614.24942	1005	37614.2494	1001
801.14	359.74	219.94	64	37614.24943	1006	37614.2494	1001
800.42	359.73	219.83	75	37614.24944	1007	37614.2494	1001
800.02	360.26	223.60	12	37614.24946	1008	37614.2495	1002
799.69	359.71	219.54	43	37614.24946	1008	37614.2495	1002
799.26	360.19	222.94	4	37614.24947	1009	37614.2495	1002
798.96	359.69	219.28	49	37614.24947	1009	37614.2495	1002
798.48	360.08	222.06	7	37614.24949	1010	37614.2495	1002
798.25	359.70	219.30	45	37614.24949	1010	37614.2495	1002
797.55	359.72	219.32	61	37614.24950	1111	37614.2495	1002
796.84	359.74	219.37	63	37614.24951	1112	37614.2495	1002
796.13	359.75	219.31	56	37614.24953	1113	37614.2495	1002

Note: Data were from 71-kHz instrument and stored in ASCII format.

PNW-FIA Software for Quality Assessment of LIDAR Data

The PNW FIA Program has developed a collection of script executables compiled from code written in C programming language that enable numerous data quality assessment and analysis operations.

Most commercial software packages with LIDAR data analysis capabilities are costly and offer few or no options for data quality assessment. To address this limitation, PNW Forest Inventory and Analysis (FIA) Program has developed a collection of script executables compiled from code written in C programming language that enable numerous data quality assessment and analysis operations. Although in the current development stage the scripts can be executed only from a command line, they will be, prior to their public release, also organized in a graphical user interface (GUI).

During script development, emphasis was given to optimizations that expedite execution, efficient use of computer hardware resources (computer memory), and support for both 32- and 64-bit operating systems and Windows or Unix/Linux platforms. The optimizations have helped realize script execution speeds that are 3 to 10 times faster than those of equivalent, commercially available, software. Many of the scripts make implicit, transparent to the user, calls to R-routines, which are provided along with the executables. R (R Development Core Team 2007) is a widely utilized, well-documented, open-source, and freely available software package that is particularly popular among researchers in a variety of disciplines. Porting C script output directly to R institutes flexibility in data quality assessment efforts and subsequent analyses, simplifies otherwise complex analysis tasks, and enables informative graphical representation of script output. In Windows platforms, a number of the scripts have equivalent versions in the Fusion software suite, which is available from the Forest Service Remote Sensing Applications Center Web site (http://www.fs.fed.us/eng/rsac/fusion/). Fusion also contains an impressive and informative laser data visualization interface.

The scripts require that the input laser data are in the LAS format described in the "Data Storage" section. To ensure portability, routines converting popular ASCII formats to LAS are provided. Typical script output includes contingency matrices (pivot tables), histograms and distribution moments of all return attributes present in an LAS file, and custom-resolution rasters of return attributes in GridASCII format or as georeferenced portable network graphic (PNG) files. Tasks specific to quality assessment of laser data range from examining LAS fields for content validity, to evaluating the consistency of distance intervals between returns originating from the same pulse, to identifying and mapping the spatial distribution of pulses with missing (filtered) returns. A set of scripts enables subsetting an LAS file by using one or more attribute value ranges, or by two- and three-dimensional spatial constraints explicated via a bounding box, a circle, a binary image or raster functioning as a mask, or via a shapefile. Where the computational efficiency is not a concern, spatial and tabular constraints can alternatively be combined simultaneously into a complex, structured query language (SQL) query. Table 4 provides a brief description of the scripts and their functionality. A detailed version containing command line syntax, examples, data requirement, and the algorithm(s) employed in each script will be available in an upcoming publication.

Table 4—Functionality description of the Pacific Northwest Forest Inventory and Analysis LIDAR utilities

Type	Script name	Description
Portability	ASC2LAS	Converts laser data from text to binary LAS format.
	LAS2ASC	Converts laser data from binary LAS to text format.
Rasterization	LAS2Raster	Generates rasters of distribution moments for each return/pulse attribute in GridASCII or georeferenced portable network graphic (PNG) format.
Statistics	LAS_Stats	Generates histograms and ASCII tabular statistics for each return attribute present in an LAS file.
Aggregation	LAS_Merge	Combines multiple LAS files into one.
Subsetting	Subset_by_attribute	Extracts returns from an LAS file within a range of values for a single return attribute and saves them in a new LAS file.
	Subset_w_raster	Eliminates returns using a raster mask in GridASCII format.
	Subset_w_BMP	Eliminates returns using a bitmap mask.
	Subset_w_Shapefile	Distributes returns to separate LAS file according to attributes of a polygon shapefile.
	Subset_w_SQL	Extracts returns from an LAS file that meet complex tabular and spatial criteria provided as a structured query language (SQL) query.
Quality assessment	ID_complete_pulse	Splits an LAS into two new files, one containing data from pulses with missing returns, and the other containing the remaining (nonfiltered) returns.
	MinDist_bw_returns	Computes the minimum distance in three dimensions of returns that originate from a single pulse and have consecutive return numbers.
	Update_Header	Queries the contents of an LAS file and updates the information present in the file header if necessary.

References

Andersen, H.-E.; McGaughey, R.J.; Reutebuch, S.E. 2005. Estimating canopy fuel parameters using airborne LIDAR data. Remote Sensing of Environment. 94: 441-449.

Andersen, H.-E.; Reutebuch, S.E.; McGaughey, R.J. 2006. A rigorous assessment of tree height measurements obtained using airborne LIDAR and conventional field methods. Canadian Journal of Remote Sensing. 32(5): 355-366.

Baltsavias, E.P. 1999a. Airborne laser scanning: basic relations and formulas. ISPRS Journal of Photogrammetry and Remote Sensing. 54: 199-214.

Baltsavias, E.P. 1999b. Airborne laser scanning: existing system and firms and other resources. ISPRS Journal of Photogrammetry and Remote Sensing. 54: 164-198.

Clark, M.L.; Clark, D.B.; Roberts, D.A. 2004. Small-footprint LIDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sensing of Environment. 91: 68-89.

Gatziolis, D. 2007. LIDAR-derived site index in the the U.S. Pacific Northwest–challenges and opportunities. Espoo, Finland: International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences. 36 (Part 3/W52): 136-143.

Gruen, A.; Akca, D. 2005. Least squares 3D surface and curve matching. ISPRS Journal of Photogrammetry and Remote Sensing. 59: 151-174.

Hinsley, S.A.; Hill, R.A.; Bellamy, P.E.; Baltzer, H. 2006. The application of LIDAR in woodland bird ecology: climate, canopy structure and habitat quality. Photogrammetric Engineering and Remote Sensing. 72: 1399-1406.

Lee, J.; Yu, K.; Kim, Y.; Habib, A. 2005. Segmentation and extraction of linear features for detecting discrepancies between LIDAR strips. IEEE International Geoscience and Remote Sensing Symposium. Los Alamitos, CA: Institute of Electrical and Electronic Engineers. 7: 4954-4957.

Maas, H.-G. 2002. Methods for measuring height and planimetry discrepancies in airborne laserscanner data. Photogrammetric Engineering and Remote Sensing. 68(9): 933-940.

Means, J.; Acker, S.; Fitt, B.; Renslow, M.; Emerson, L.; Hendrix, C. 2000. Predicting forest stand characteristics with airborne scanning LIDAR. Photogrammetric Engineering and Remote Sensing. 66: 1367-1371.

Naesset, E. 2002. Predicting forest stand characteristics with airborne laser using a practical two-stage procedure and field data. Remote Sensing of Environment. 80: 88-99.

Okatani, I.S.; Deguchi, K. 2002. A method for fine registration of multiple view range images considering the measurement error properties. Computer Vision and Image Understanding. 87: 66–77.

Persson, A.; Holmgren, J.; Sodermann, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing. 68: 925-932.

Popescu, S.; Zhao, K. 2007. A voxel-based LIDAR method for estimating crown base height for deciduous and pine trees. Remote Sensing of Environment. doi:10.1016/j.rse.2007.06.011.

R Development Core Team. 2007. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org. (May 29, 2008).

Stoker, J.; Parrish, J.; Gisclair, D.; Harding, D.; Haugerud, R.; Flood, M.; Andersen, H.E.; Schuckman, K.; Maune, D.; Rooney, P.; Waters, K.; Habib, A.; Wiggins, E.; Ellingson, B.; Jones, B.; Nechero, S.; Nayegandhi, A.; Saultz, T.; Lee, G. 2007. Report of the first national LIDAR Initiative meeting. Open-File Report OF2007-1189. U.S. Department of the Interior, U.S. Geological Survey. http://www.lidar.cr.usgs.gov/presentations/NLImeetingReport-final.pdf. (June 2, 2008).

This Page Left Blank Intentionally

This Page Left Blank Intentionally

Pacific Northwest Research Station

Web site	http://www.fs.fed.us/pnw/
Telephone	(503) 808-2592
Publication requests	(503) 808-2138
FAX	(503) 808-2130
E-mail	pnw_pnwpubs@fs.fed.us
Mailing address	Publications Distribution Pacific Northwest Research Station P.O. Box 3890 Portland, OR 97208-3890

U.S. Department of Agriculture
Pacific Northwest Research Station
333 SW First Avenue
P.O. Box 3890
Portland, OR 97208-3890