Accuracy of Airborne Lidar-Derived Elevation: Empirical Assessment and Error Budget

Michael E. Hodgson and Patrick Bresnahan

Abstract

As part of a countywide large-scale mapping effort for Richland County, South Carolina, an accuracy assessment of a recently acquired lidar-derived data set was conducted. Airborne lidar (2-m nominal posting) was collected at a flying height of 1207 meters above ground level (AGL) using an Optech ALTM (Airborne Laser Terrain Mapper) 1210 system. Unique to this study are the reference point elevations. Rather than using an interpolation approach for gathering observed elevations at reference points, the x-y coordinates of lidar points were located in the field and these elevations were surveyed. Using both total-station-based and rapid-static GPS techniques, observed vertical heights were measured at each reference lidar posting. The variability of vertical accuracy was evaluated for six land-cover categories. Root-mean-squared error (RMSE) values ranged from a low of 17 to 19 cm (pavement, low grass, and evergreen forests) to a high of 26 cm (deciduous forests). The unique error assessment of lidar postings also allowed for the creation of an error budget model. The observed lidar elevation error was decomposed into errors from lidar system measurements, horizontal displacement, interpolation error, and surveyor error. A cross-validation approach was used to assess the observed interpolated lidar elevation error for each field-verified reference point. In order of decreasing importance, the lidar system measurements were the dominant source of error followed by interpolation error, horizontal displacement error, and surveyor error. Observed elevation error in steeper slopes (e.g., 25°) was estimated to be twice as large as those on low slopes (e.g., 1.5°).

Introduction

The use of airborne lidar (LIght Detection And Ranging) sensors for topographic mapping is rapidly becoming a standard practice in the aeroservice community. Counties are collecting such data for a variety of management purposes—stormwater assessment, flood control, visualization, etc. Several efforts are underway to collect lidar data statewide, primarily for flood-plain mapping associated with the Federal Emergency and Management Agency's (FEMA's) flood mitigation efforts. Also, several federal agencies in the United States, Great Britain, and other countries are using airborne lidar for topographic mapping applications.

While a general understanding of the relative accuracy of lidar is known, too few empirical studies exist for assessing the accuracy of digital elevation models (DEMs) created from these data. During the initial years of lidar mapping efforts (i.e., 1995 to 2000), most aeroservice companies would routinely quote accuracies of 15 cm RMSE. Most would now agree such accuracy is only achievable under the most ideal circumstances (e.g., low altitude collections, flat terrain, minimal or no surface vegetation or obstructions, much human analysis, etc.). A few empirical studies have been conducted to date and suggest accuracies of 26 cm to 153 cm root-mean-squared error (RMSE) for large-scale mapping applications (Adams and Chandler, 2002; Bowen and Waltermine, 2002; Hodgson *et al.*, 2003). There is a need for numerous research efforts in quantifying the accuracy of lidar data incorporating various platform parameters and environmental conditions—collection, processing, geography.

In this article, the elevation error was assessed for a spatially dense (2-m nominal postings) lidar dataset collected in Richland County, South Carolina. The analysis tested the hypothesis that mean elevation error would vary between land-cover categories while holding terrain slope constant. The uniqueness of this lidar accuracy assessment is that the reference data were collected at actual lidar points rather than between lidar points, requiring spatial interpolation (Figure 1). This approach allowed for an assessment of observed lidar error without the additional error introduced by typical spatial interpolation. An error budget for observed lidar elevations was created, which included lidar system, horizontal, interpolation, and surveyor errors. Results from this analysis could be used to estimate observable elevation errors from similar lidar datasets in other environmental conditions.

Background

Sources of Positional (*X-Y-Z*) Error

The primary sources of positional error in the lidar collection process are associated with the Global Positioning System (GPS) equipment onboard the aircraft, the inertial navigation unit (INU) for estimating positions between GPS fixes, and the inertial measurement unit (IMU) for monitoring the pointing direction of the laser. The horizontal (*X-Y*) error is typically much greater than the vertical error. Assessing the horizontal accuracy of lidar observations is also problematic. Conventional assessments of horizontal error involve multiple overflights over building corners with flat roofs. Most lasers used in the commercial lidar sensors are similar and have a divergence from 0.2 to 0.33 mr. This divergence, along with scan angle and flying height, defines the lidar footprint (typically between 24 cm to 60 cm). Smaller footprints are more likely to pass through breaks in forest canopy.

M.E. Hodgson is with the Department of Geography, University of South Carolina, Columbia, SC 29208 (hodgsonm@sc.edu).

P. Bresnahan is with the Department of Information Technology, Richland County South Carolina, 2020 Hampton Street, Suite 3014, Columbia, SC 29204.

Photogrammetric Engineering & Remote Sensing
Vol. 70, No. 3, March 2004, pp. 331–339.

0099-1112/04/7003–0331/$3.00/0

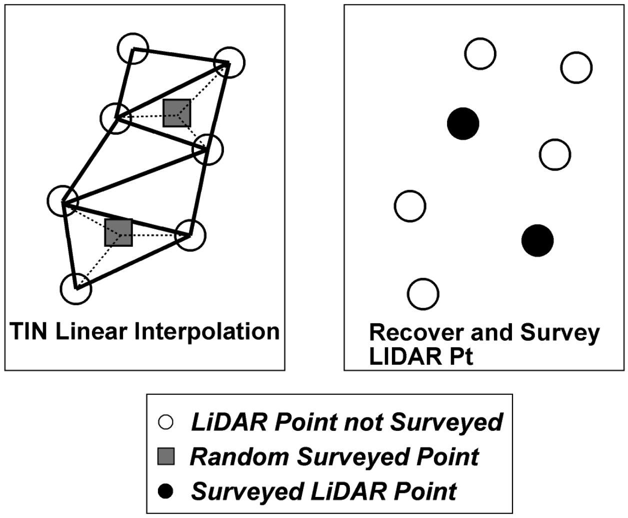

Figure 1. Linear interpolation of reference points introduces additional error in the observed lidar error whereas surveying the lidar point locations does not.

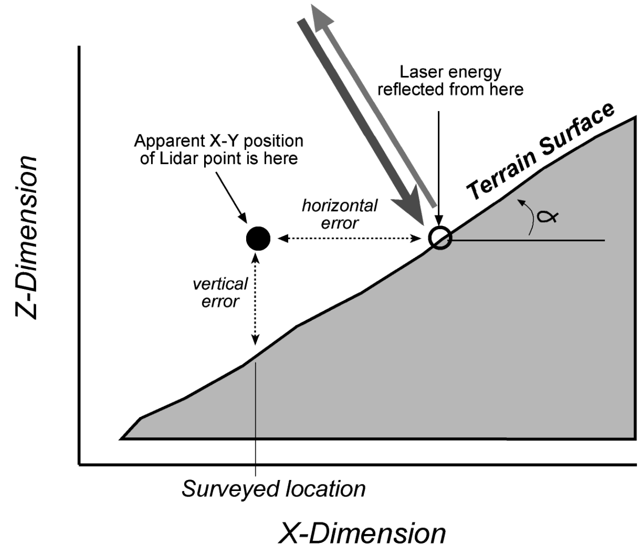

Figure 2. Illustration of the effects of terrain slope on observable elevation error.

Lidar Point Labeling Error

A single lidar pulse of approximately 200 cm in length is emitted toward a surface location at some 10,000 to 70,000 times per second. Most lidar sensors record the return energy as a waveform of multiple pulses, and then unique returns are identified (e.g., four or five returns). In some instances, the magnitude of the pulse is also recorded and is referred to as *intensity*. From this set of lidar returns, both automated and manual methods are used to identify or label each return as a "ground" return, "vegetation" return, "building" return, or other. This process is sometimes referred to as "vegetation removal." Automated methods are based on neighborhood operators that iteratively identify the lowest points within each neighborhood and add them to a candidate set of "ground" returns. Subsequent iterations refine the candidate set by adding additional returns that are also "low" or exhibit some angular deflection from a surface modeled by the current candidate set of points. The exact neighborhood operators and parameters vary by lidar mapping vendor and are confidential. Many vendors have adopted the Terrasolid software for their automated labeling of lidar returns (Schuckman, personal communication, 2002).

A human operator will analyze the candidate set of "ground" returns to further improve the accuracy of labeling. Typically, the human analyst will visualize small areas as a three-dimensional cloud of lidar points overlain on available digital orthophotography. Thus, the process of further labeling the points is locally adaptive and subjective.

In this study the term $RMSE_{LIDAR\ System}$ refers to the elevation error at lidar points that includes error from both the laser pulse height measurement and point labeling process.

Mapping Sources of Error

A well-known characteristic of observed elevation error for terrain mapping is the relationship with *terrain slope* (Maling, 1989). Even if the *elevation* of a surface observation is measured without error, the horizontal error in the observation may introduce "apparent" error in the elevation value from a user's perspective. The introduction of elevation error based on horizontal error is only true for inclined slopes (Figure 2). The *maximum* amount of elevation error introduced is a function of surface slope: i.e.,

$$\textit{Elevation Error} = \tan\alpha \times \textit{Horizontal Displacement}. \qquad (1)$$

For instance, the additional elevation error introduced for a point with a 100-cm horizontal error on a 10° slope can be up to ±18 cm. The maximum error will only occur if the displacement is perpendicular to the contour line. The error will be zero for displacement parallel with the contour. In practice, the displacement direction for any single point is unknown and assumed to be random for the entire set of points.

Early work in topographic mapping has demonstrated the relationship between observation *point density* and the accuracy of the derived DEM (MacEachren and Davidson, 1987). As the density of observation points increases, the accuracy of the resulting DEM increases. The density of lidar "ground" returns depends on the nominal posting (i.e., spacing between lidar pulses at nadir), land-cover type, and point-labeling approach. This density concept resulted in a maximum 5-m posting criteria within FEMA's guidelines for using lidar data to construct DEMs in the floodplain mapping process. FEMA also suggested a minimum percentage of "data voids," areas where the distance to the nearest lidar observation is greater than some threshold.

Emipirical Assessments of Lidar Error

The state of North Carolina is currently collecting airborne lidar data for mapping flood hazards under FEMA's floodplain mapping efforts (http://www.ncfloodmaps.com, last accessed 23 October 2003). The project includes an accuracy evaluation on a county-by-county basis. Lidar data were collected at a nominal post spacing of 4.5 m and then interpolated to a 5-m cell size DEM. The target vertical accuracy for their data is 20 cm (RMSE) for coastal counties (composed largely of "flat" terrain) and 25 cm for inland counties (composed largely of rolling or hilly terrain). At least 20 reference points are collected in each of five land-cover categories (grass, weeds/crop, scrub, forest, and built-up). A variation on the reporting of accuracy is the "95 percent RMSE calculation" report. In this 95 percent report, the state removes 5 percent of observations in the accuracy assessment that have the highest errors. For the 41 counties studied in the first phase of the North Carolina program, the overall accuracy based on the 95 percent RMSE calculation was 15.15 cm.

Hodgson *et al.* (2003) have reported accuracies from a 3.4-m nominal post spacing lidar dataset covering a watershed in the piedmont of North Carolina. The uniqueness of this study was that the lidar data were collected in leaf-*on* conditions (summer month). This study found that the accuracy was significantly different between land-cover categories. Elevation error with the lidar data ranged from 33 cm (low grass) to 153 cm (scrub/shrub). Errors in low grass and high grass were much smaller than those in the more heavily vegetated canopies, except for pine forests. Elevation error was only correlated with increasing slope for the scrub/shrub land cover. The minimal correlation with slope was suspected because mean absolute error in slope ranged from only 1.7° to 4.8° by land-cover category.

In a coastal mudslide study area, Adams and Chandler (2002) found a lidar-derived elevation accuracy of 26 cm (RMSE). The lidar data were less sensitive to terrain slope than was a DEM derived from digital photogrammetry. The authors also found a bias in the lidar data that tended to slightly underpredict elevation.

Bowen and Waltermine (2002) assessed the accuracy of lidar-derived elevation data and how accuracy may vary by topography. Using data collected in a western river corridor, they found an overall accuracy of 43 cm (RMSE). Reference data in the form of transects were used to identify a weakness of vegetation removal algorithms in variable terrain compared with flat terrain.

A few other studies have focused on the effects of lidar postprocessing methods on elevation accuracy. Lloyd and Atkinson (2002) found that kriging with a trend model (i.e., universal kriging) may be more accurate when the density of lidar postings decreases. In a study on lidar point-labeling research, a 17-cm (RMSE) accuracy was observed in a grass and cereal crop land cover (Cobby *et al.*, 2001). Elevation accuracy from the lidar data was found to decrease in a densely wooded environment. Cowen *et al.* (2000) found that dense canopy cover can have a profound effect on the percentage of lidar "shots" reaching the ground. Dowman and Fischer (2001) have also conducted research on the effects of data editing of lidar multiple return data on final elevation accuracy. These two authors found considerable (e.g., up to 4 m) error over *open ground* when using all of the lidar returns. More recently, Raber *et al.* (2002) developed an adaptive method for more accurately identifying lidar returns from the ground (rather than above-ground obstructions).

Summary of Error Sources

Digital elevation models (DEMs) or triangulated irregular networks (TINs) produced from lidar observations are created from lidar returns labeled as "ground" observations. These lidar-derived "ground" elevations contain error from three categories of sources:

- elevation error from the sensor system measurement,
- horizontal error from the sensor system measurement, and
- labeling error process of identifying a "ground" return from other types of returns (e.g., canopy top, intermediate vegetation, building top).

The error in the measured elevation of a lidar point is the cumulative product of the sensor/platform sources, such as analysis of waveform, identification of the return position in the pulse length, and position error in the GPS/INU control. Horizontal error is a function of the same factors but often dominated by flying height. Horizontal error is often reported to be approximately 1/1000th of the flying height AGL on most systems. The horizontal error will introduce additional elevation error in the *use* of the observations. The lidar "ground" return set contains both omission and commission errors. Some points in the set are incorrectly labeled as "ground" returns, and some true "ground" returns are omitted. From a user's perspective, the point set is treated as if there are no errors or occasionally as if there are only elevation errors. The use of lidar observations typically involves the interpolation of an irregular point set to a DEM, which may introduce additional elevation error. Finally, any assessment of elevation error with reference data introduces additional "apparent" error from the surveying of the reference data.

In this study, we approach the data from a user's perspective and treat the point set as if it is labeled correctly and is in the correct horizontal position. Observable elevation errors are assessed and then assumed to result from the remaining factors. Finally, an error budget is created from errors in the lidar system, horizontal error, surveyor error, and interpolation error.

Methodology

Lidar Data Collection

Airborne lidar data were collected for the 2000-km^2 area comprising Richland County, South Carolina at a nominal 2-m by 2-m posting. Oriented perpendicular to the fall line, the topography of the County includes rolling hills of mixed deciduous and coniferous vegetation in the northwest that transitions through urban/suburban features to agricultural pine and bottomland hardwoods of the Congaree Swamp in the southeast. Data were collected during leaf-*off* conditions from 01 through 22 March 2000. A total of 115 flightlines of lidar data were required to completely cover the County. Flight altitude was 1207 m (3960 feet) AGL. Simultaneous color video imagery and black-and-white orthophotography were also collected and used to guide the selection of stable and known ground cover conditions (e.g., no cars in parking lots, no animals in fields).

The contracted vendor, MD Atlantic Technologies, collected and postprocessed all data. Automated postprocessing of the lidar returns using waveform analysis and spatial filters resulted in separate files. The delivered products included lidar return files for first vegetation, last vegetation, first ground, and last ground. The last ground elevation file represents the "bald earth" and was the subject of this accuracy assessment. The entire file of lidar returns for Richland County contained over one billion points where 250 million points were labeled as "ground" returns.

Reference Data

A stratified random sampling approach was used to collect reference data (654 points) for assessing the accuracy of the lidar. Unique to this study is the survey of elevations at the same *X-Y* location as the lidar ground points rather than random *X-Y* points between lidar points. Reference points were stratified based on proximity to existing high accuracy geodetic monuments and land-cover categories. Thirteen unique study sites were developed as a sampling of the 115 lidar flightlines. Random points within the target land-cover categories were selected. Survey data were collected during December of 2000.

Land-Cover Categories

As discussed previously, overstory may have a profound effect on either the density of nearby lidar "ground" points and/or may attenuate the lidar signal. Both the density of overstory and height of overstory may result in a reduction in elevation accuracy. When developing digital elevation models for flood insurance rate mapping (FIRM) applications, FEMA has argued for using stratified reference points based on fundamental land-cover categories as selected here. In Richland County, South Carolina, the dominant forest types are evergreen, deciduous, mixed forests, brush/low trees (i.e., a scrubby environment), and bottomland hardwoods. Six land-cover categories were used to stratify the surveyed lidar points. The land-cover categories in this study included pavement, low grass, high grass, brush/low trees, evergreen, and deciduous (Figure 3). The differences between low grass

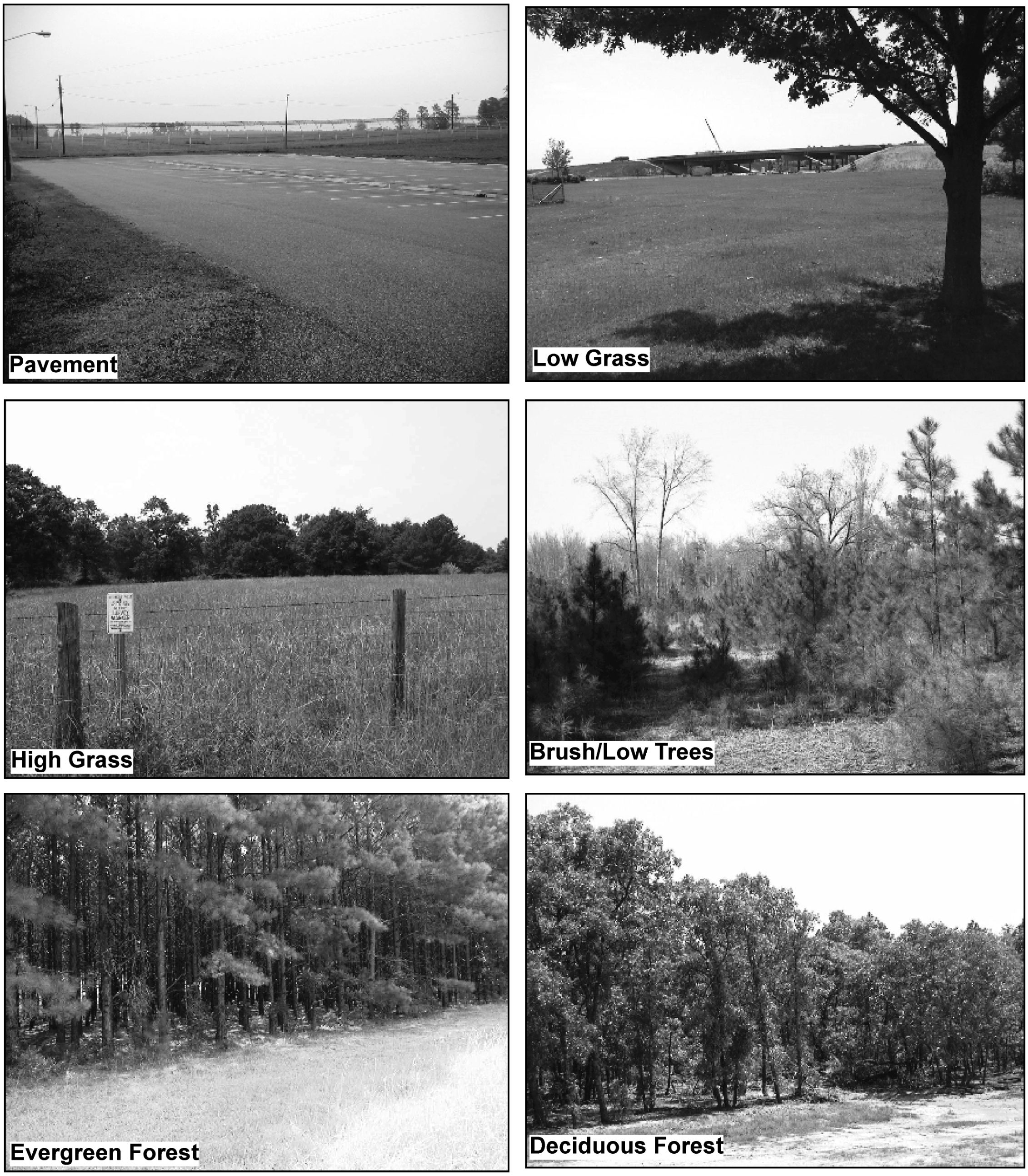

Figure 3. Example photographs of each of the land-cover categories—pavement, low grass, high grass, brush/low tree environment (in foreground), evergreen forest, and deciduous forest. Low grass cover is 8 cm high or less while high grass cover may range up to 90 cm in height. Photos were taken in May of 2001 while the lidar data were collected in March of 2000.

(less than 8 cm) and high grass (up to about 90 cm) are management and height. Because of the difficulty in controlling for different "mixtures" of mixed forests, only homogeneous evergreen and deciduous categories were used. Bottomland hardwoods are typically a very wet environment that absorbs much of the laser energy (few "ground" returns) and are a very difficult environment in which to establish surveys.

Survey Methodology at each Study Site

An independent surveying company was contracted to survey the elevation at each of the 654 reference lidar "ground" returns within the 13 unique study sites. Each study site contained an existing geodetic monument with both high vertical/horizontal control or with only vertical control (Figure 4). A *local azimuth* was established at each site with rapid-static GPS methods (using a Leica 9500 dual frequency 24-channel receiver). This local azimuth was required for establishing the orientation of the total station positioned on the monument to true north. Eight of the study sites had high order horizontal and vertical control. For the five study sites that only had vertical control at the monument, the horizontal coordinates were established using rapid-static GPS.

A stratified random approach was used to select lidar points in close proximity to the monumented sites (Figure 4).

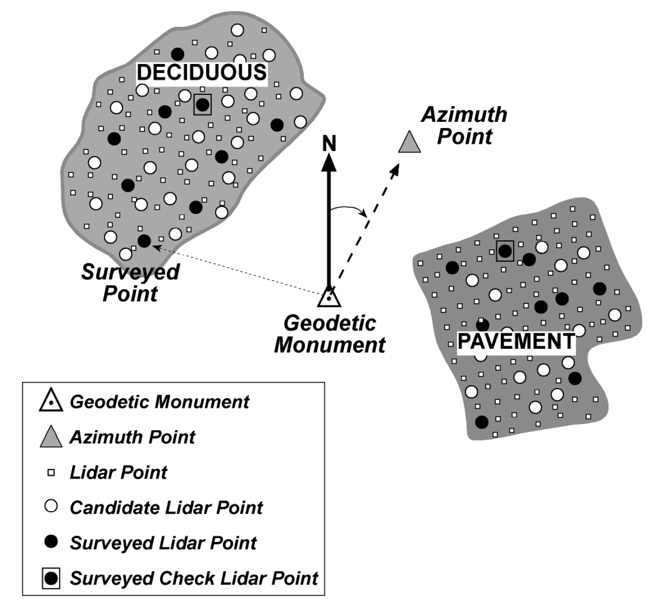

Figure 4. Planimetric diagram of a study site and how reference data were collected. The number of actual candidate and reference points is simplified for illustration purposes.

Prior to field surveying, 50 lidar "ground" returns points within each land-cover category surrounding a geodetic monument were randomly selected. Arranged by land-cover category, the *X-Y* coordinates of these randomly selected points were provided to the surveying company as "candidate" points to "recover" and survey. For instance, the study site depicted in Figure 4 is surrounded by two land-cover categories of interest. Thus, 100 total lidar locations were given to the surveyor. Although the *X-Y* coordinates of the lidar points were supplied to the surveyor, the corresponding lidar-derived elevations were withheld to avoid bias.

At each site, the surveying crew recovered at least 20 of the supplied candidate lidar point locations for each land-cover category. A Sokia Set total station (one-second accuracy) and survey rod were used as instruments. The survey rod had a three-second mounted prism and a wide shoe affixed to the bottom to ensure that the rod sat on the ground (rather than a pointed rod that pierced into the ground.). Each lidar *X-Y* position was located using the "stake-out" method. This method uses an iterative approach to "recover" the lidar *X-Y* position:

(1) The rodman moves in the specified direction and approximate distance from the total station and sets the rod level;
(2) The distance is measured (using a laser-range finder on the total station) and the direction is measured;
(3) If the surveyed *X-Y* position is not within 30 cm (1.0 foot) of the lidar *X-Y* position, the rodman moves closer in the direction of the provided position and re-measures (i.e., Step 2);
(4) If the surveyed and lidar *X-Y* positions are within 30 cm of each other, the elevation is surveyed.

Most of the lidar points were surveyed using a single shot. In some instances, a short traverse was required to determine heights from the total station to one or more lidar points. Candidate lidar reference points that were on logs, rocks, or other surface obstructions were discarded and another candidate reference point was used.

Eighteen "check" points (distributed throughout the 13 study sites) were used for estimating the accuracy of the survey methods (Figure 4). A second survey was made using a new total station setup. For these "back-checks," the total station was set up on another "monument" and the reference points were re-surveyed. Thus, for 18 reference points, there was a second set of elevations to calculate the survey-induced elevation difference. Although the differences in elevations measured from two different station setups is not strictly "error" (because "truth" is not known), it is treated here as a good estimate of surveyor/method introduced error. The estimate of surveyor error here does not include error in the monument position. Surveying the lidar points required somewhat different levels of effort for each land-cover category. Obviously, surveying elevations in the pavement or grass categories was not as labor intensive as the brush/low tree environment. The surveying crews varied their effort to maintain the same relative degree of confidence in their results. The assumption in this analysis is the resulting surveyor-induced error does not vary by land cover. Based on the back-check approach, the *observed survey error* (vertical error) in this study would be 3.1-cm RMSE.

Hypotheses

Following the work of others (Cobby *et al.*, 2001; Hodgson *et al.*, 2003; also see http://www.ncfloodmaps.com, last accessed 22 October 2003), it was assumed that the mean absolute error of lidar points will vary with land-cover type. The null hypothesis tested was that the mean absolute error for each land-cover category was equal. A one-way ANOVA and subsequent paired *t*-tests were used to test if land cover does affect mean absolute error. A second test was conducted comparing the mean signed error to 0.0. The research hypothesis explored in this second test was whether there was a tendency for the lidar data to under- or over-predict elevations within each land-cover category.

Error Budget

An error budget model was created to isolate the relative contribution from each error source. The mean errors from the different sources were estimated in the following steps:

(1) Estimate the *surveyor error*;
(2) Compute the *observed* elevation error at lidar *points*;
(3) Model elevation error caused for *horizontal* error and terrain *slope*;
(4) Compute the *elevation error* with surveying and horizontal errors removed (this is an estimate of the lidar sensor system error);
(5) Estimate the *interpolation induced* elevation error; and
(6) Estimate the elevation errors in *other terrain slopes* caused by horizontal displacement, interpolation, and lidar derived elevation at the point.

Error was always measured by subtracting the surveyed elevation from the lidar-derived elevation, resulting in positive errors for an over-prediction of elevation. Several measures of error were computed: mean signed error, mean absolute error, and RMSE. The RMSE was computed as

$$RMSE_{Observed\ LIDAR\ Pts} = \sqrt{\frac{\sum (Z_{LIDAR} - Z_{Survey})^2}{n}}. \qquad (2)$$

The cumulative error from *independent* error sources is not a simple additive form. For any single point, the accumulated errors may cancel each other, resulting in an apparent error of 0.0. The model presented here assumes that the errors from each source are independent. The cumulative error observed in a TIN or DEM (i.e., $RMSE_{observed\ in\ TIN}$) at each reference point in a typical accuracy assessment is derived as

$$RMSE_{observed\ in\ TIN} = \sqrt{RMSE^2_{LIDAR\ System} + RMSE^2_{Survey} + RMSE^2_{Horiz,Slope} + RMSE^2_{Interp}} \quad (3)$$

where $RMSE_{LIDAR\ System}$ is the elevation error from the lidar system measurement system and return labeling process, $RMSE_{Survey}$ is the surveyor error, $RMSE_{Horiz,Slope}$ is the elevation error introduced from horizontal displacement and associated slope effects, and $RMSE_{Interp}$ is the elevation error introduced through the interpolation process.

The $RMSE_{observed\ in\ TIN}$ values computed in this study are the equivalent values that are reported by practically all other studies of lidar elevation accuracies. This study did not require a spatial interpolation because the reference points coincided with the lidar ground posting. The spatial interpolation error contribution is, thus, 0.0 and the total error observed from the set of surveyed lidar points ($RMSE_{Observed\ LIDAR\ Pts}$) is

$$RMSE_{observed\ LIDAR\ Pts} = \sqrt{RMSE^2_{LIDAR\ System} + RMSE^2_{Survey} + RMSE^2_{Horiz,Slope}}. \quad (4)$$

If the survey error is relatively small when compared to the observed lidar error, it is interesting to note how little effect survey error has on evaluating the error solely from the lidar collection process (Figure 5). For example, if the survey error introduced in the reference data was 5 cm RMSE, the RMSE from horizontal displacement and slope was 0 cm, and the actual lidar elevation error was only 15 cm, the resulting observed error in the field would only be 15.8 cm. The estimate of surveyor introduced error (i.e., 3.1 cm) in this study was derived from the back-check surveyor methodology.

Our goal was to estimate the elevation error associated with the lidar system process (including return labeling). By rearranging the terms in Equation 4, the RMSE from the lidar system process ($RMSE_{LIDAR\ System}$) is

$$RMSE_{LIDAR\ System} = \sqrt{RMSE^2_{observed\ LIDAR\ Pts} - RMSE^2_{Survey} - RMSE^2_{Horiz,Slope}}. \quad (5)$$

A simulation was performed to model the relationship between RMSE elevation, mean slope, standard deviation of slope values, and RMSE horizontal error. Horizontal error ($RMSE_{horizontal}$) was assumed to be *normally* distributed. Terrain slopes were assumed to be *normally* distributed with a range from 0 to 40° (mean slope). Standard deviation in slope (σ_{slope}) ranged from 0.5° to 1.0° of the mean slope. The direction of the horizontal error was assumed to be *uniformly* distributed between 0 and 90 degrees. Using a random set of over 17,000 observations, the following statistical relationship was found ($R^2 = 0.999$):

$$RMSE_{Horiz,Slope} = (0.689282633 * Tan(\overline{Slope}) * RMSE_{Horizontal}) + (0.006194908 * RMSE_{Horizontal} * \sigma_{Slope}) \quad (6)$$

The advertised horizontal error (RMSE) for the Optech 1210 instrument in this study is a function of altitude as 1/1000 of the flying AGL. For the 1207-m flying height AGL used in this study, the $RMSE_{horizontal}$ would be 120 cm.

Surface slope was derived from a triangulated irregular network (TIN) of the *surveyed* elevations at the lidar *reference* points. Mean surface slope by land-cover category ranged from 1.67° to 4.15° (brush and low trees). Assuming a 120-cm RMSE horizontal error, the $RMSE_{Horiz,Slope}$ values resulting from terrain slope would range from 2.5 cm (low grass points) to a high of only 6.2 cm (brush and low trees points).

Interpolation Error

To assess the additional amount of interpolation error introduced into a typical DEM created from lidar data, a crossvalidation approach was used. The crossvalidation approach estimated the elevation value *at* each lidar "ground" return point that was surveyed. This approach iteratively removes one point, interpolates the elevation at the location of the removed point using all remaining points in the set, and then processes another point in the same manner (Figure 6). Unlike other uses of the crossvalidation approach, the lidar observations included *surveyed* lidar observations. Only the error at the surveyed lidar observation (i.e., 654 points) in the crossvalidation approach were used to compute $RMSE_{Interp}$.

The interpolation algorithm used in this study was the commonly used linear interpolation across triangular faces in a TIN. This approach is often used by aeroservice companies

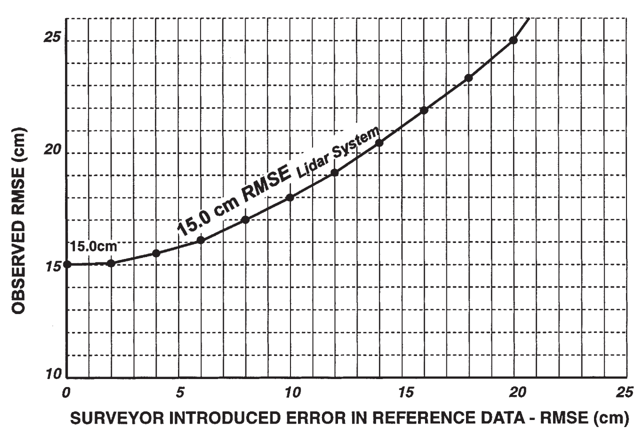

Figure 5. Graph of 15.0-cm observed error as a function of actual elevation error from the lidar system and error in the reference data. As surveyor error increases, the observed error in the lidar data will also increase at an increasing rate.

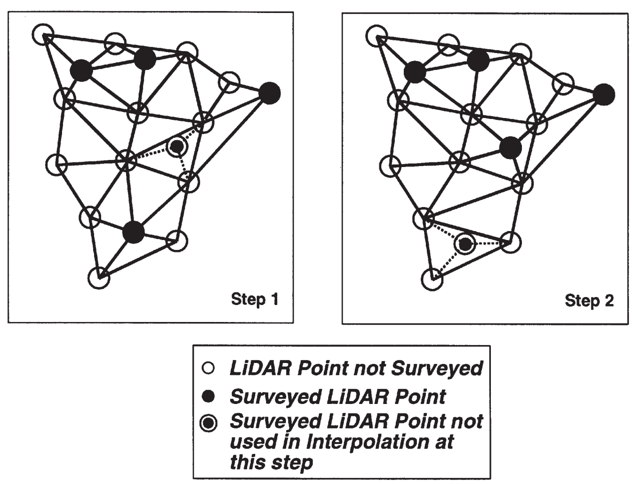

Figure 6. Illustration of two steps in the crossvalidation approach where the TIN is constructed and the interpolation is performed without one of the observations (step 1). In the next step, the removed point is reinserted into the point set and the process is repeated with another observation (step 2).

TABLE 1. DENSITY OF LIDAR "GROUND" RETURNS BY LAND COVER CATEGORY

	Cover Type					
	Pavement	Low Grass	High Grass	Brush/Low Trees	Evergreen	Deciduous
Area per Observation (m^2)	7.67	6.28	5.43	6.28	10.80	8.46
Regularized Post Spacing (m)	2.77	2.51	2.33	2.51	3.28	2.91

TABLE 2. OBSERVED LIDAR ELEVATION ERROR (ELEVATIONS IN CM)

	Cover Type					
	Pavement	Low Grass	High Grass	Brush/Low Trees	Evergreen	Deciduous
N	120	137	98	98	119	82
Signed Mean (cm)	6.0	1.8	−3.8	−6.0	−4.6	1.0
Signif. From 0.0	0.000	0.354	0.047	0.009	0.003	0.723
Absolute Error (cm)	14.9	16.8	15.9	18.9	12.9	20.3
$RMSE_{Observed\ Lidar\ Pts}$	18.9	22.5	18.9	23.3	17.2	25.9

in the DEM production step. The parameters of the TIN construction algorithm in the ESRI ArcView package were adjusted so that all lidar "ground" returns were included in the final TIN. None of the points were eliminated based on weed or proximal tolerance parameters. A script was developed to implement the crossvalidation approach using the TIN interpolation method. A unique TIN was constructed for each of the 654 reference points where the reference point under examination was removed from the TIN construction and the error for that point was computed.

Results and Discussion

Density of Postings by Land-Cover Category

One of the primary goals in specifying the parameters for lidar data collection (i.e., flying height, forward speed, pulse rate, footprint, etc.) is to achieve a high spatial density of lidar pulses. If the subsequent vegetation removal process (i.e., point labeling) works well, then a dense set of "ground" returns will be obtained. Observed posting density was computed by taking all of the study area polygons for the candidate land cover and computing the density of ground returns for that land cover (Table 1). If the ground returns were equally spaced, the equivalent grid of lidar points would have an approximate 2.3- to 3.0-m cell size (Table 1). As expected, the density of observations was lower in the evergreen and deciduous forest categories.

Observed Accuracy

Based on the observed error at lidar points that were surveyed, the error for the overall data set was 21.1 cm ($RMSE_{Observed\ Lidar\ Pts}$). The elevation error ($RMSE_{Observed\ Lidar\ Pts}$) ranged from 17.2 to 25.9 cm among the land-cover categories (Table 2). This error range is relatively low compared to other studies of coarser post-density spacing. The lowest errors were observed for points under evergreen forest, pavement, and high grass land cover. Deciduous forests and brush/low trees exhibited the highest errors (25.9 cm and 23.3 cm, respectively).

The low error in pavement was expected, while the low error in evergreen forests was a surprise. Southern pine forests, particularly managed pines, tend to be even aged without appreciable understory. The needles are high in the overstory and lower branches continually die back as the tree grows higher. Although the needles of Southern pines do not completely drop during the winter, the canopy is surprisingly porous to sunlight (and thus, lidar). It is believed that the high pine canopy and uncluttered forest floor provide two good levels of returns that enable the ground to be easily distinguished in the lidar returns.

To determine if the differences in errors between pairs of land-cover categories was statistically significant, an independent samples *t*-test was conducted using the mean absolute error (Table 2). Levene's test for equality of variances was first used to test if the individual variances were equal. The appropriate *t*-test was used depending on whether the variances were assumed equal or not. Errors in the pavement and evergreen categories were not significantly different from one another at the 0.05 level (Table 3). Mean errors for the low- and high-grass land covers were significantly greater than for pavement and evergreen. Although the grasses are either cut or in senescence during the winter months, enough stubble may remain to cause confusion when classifying lidar returns. The highest errors were in the brush/low trees (23.3 cm) and deciduous forest (25.9 cm) land cover (Table 2). The brush/low trees are obviously a mutistory environment, causing problems for both the automated and manual lidar labeling processes. The deciduous category would not contain leaves during the early spring lidar collection. However, the forest floor typically contains a somewhat random collection of low vegetation. This low understory may obscure the ground from the laser pulses.

Tendency to Underpredict/Overpredict

An independent samples *t*-test was used to test if the *mean signed* error for each land-cover category was significantly different from zero. Observations in the pavement, high grass, brush/low trees, and evergreen were significantly different from zero (Table 2). Pavement elevations were overpredicted (+6.0 cm) while high-grass, brush low/trees, and evergreen observations were underpredicted (−3.8 to −6.0 cm), on average.

TABLE 3. SIGNIFICANCE LEVELS FOR DIFFERENCE BETWEEN MEAN ABSOLUTE ERROR BY LAND COVER

	Pavement	Low Grass	High Grass	Brush/ Low Trees	Evergreen
Low Grass	0.250				
High Grass	0.495	0.609			
Brush/Low Trees	0.020	0.083	0.019		
Evergreen	0.190	0.022	0.042	0.035	
Deciduous	0.006	0.106	0.028	0.531	0.000

Total Observed Error after Interpolation

In the typical application with lidar points, a DEM would be created using an interpolation algorithm. Otherwise, a TIN would be constructed and contour lines or points-of-interest would be interpolated. The interpolation process was expected to introduce additional error in the DEM. The separate analysis using the crossvalidation approach to estimate the observed error after a spatial interpolation process (i.e., linear interpolation within a TIN) is shown in Table 4. As expected, total errors for each land-cover class increase, although not by a substantial amount. The increase in total error ranges from 0.4 cm to 3.3 cm. A paired *t*-test was also used to test for a significant difference between the mean absolute interpolated error and the mean absolute error at the lidar reference points (Table 4). For most land-cover categories, the mean absolute error observed in the interpolated elevations was significantly larger than the mean absolute error observed at the reference points (at the 0.05 probability level or better). Surprisingly, the interpolation process actually resulted in a significant improvement in the elevation values in deciduous forests (decreasing from 25.9 cm to 23.5 cm RMSE). It is believed that small neighborhood smoothing introduced during the TIN interpolation process is the reason for improvement. The lidar returns are somewhat more variable under the deciduous canopy, and this smoothing reduces overall error.

Isolating Error from Lidar System

The errors discussed thus far, such as $RMSE_{Observed\ Lidar\ Pt}$, include errors caused by horizontal displacement and the error from the surveying of reference data. If we assume that the surveyor-induced elevation error in the accuracy assessment process was equal across land-cover categories, the average error from the lidar collection process ($RMSE_{LIDAR\ System}$) can be computed by subtracting the surveyor-introduced error ($RMSE_{Survey}$) and elevation error from the X-Y position error and terrain slope ($RMSE_{Horiz,Slope}$). The RMSE for surveying was 3.1 cm and the RMSE for horizontal error was assumed to be 120 cm. For example, using Equation 6, the resulting $RMSE_{Horiz,Slope}$ for the pavement category would be 3.01 cm.

The computed error in elevation for the pavement category from the lidar system process alone would be derived as

$$\begin{aligned} RMSE_{LIDAR\ System} & \\ &= \sqrt{18.9^2_{observed\ LIDAR\ Pts} - 3.07^2_{Survey} - 3.01^2_{Horiz,Slope}} \\ &= \sqrt{18.9^2_{observed\ LIDAR\ Pts} - 18.48} = \sqrt{338.73} = 18.41_{cm}. \end{aligned} \quad (7)$$

The $RMSE_{LIDAR\ System}$ was calculated for each land-cover category (Table 5). Removing the average error from the surveying process and horizontal error on these low slopes only slightly decreases the error observed in the lidar system process. The observed RMSE at the reference points decreased by only 0.3 cm to 1.1 cm. Thus, the actual error from the sensor measurement and point labeling process in the lidar elevations is clearly the largest sources of error provided to the user.

Estimates of observed elevation error for other terrain slopes, such as mountainous areas, can be made from the different error sources. Assume a 25° mean slope with a standard deviation of 5°. Using Equation 6, the RMSE of elevation error caused solely by the horizontal displacement on 25° slopes would be substantial: 42.3 cm. If we assume that other factors in the lidar remote sensing process have the same effect on 25° slopes as they do on very low slopes (as in this study), then the estimated observable RMSE for each land-cover category is shown in Table 5. Observable elevation errors at lidar points (i.e., no interpolation) would be almost twice as large, ranging from 46 cm to 49 cm.

Because practical applications of lidar data utilize some type of spatial interpolation (e.g., TIN, Kriging, etc.), the final elevation surface will include the addition of such interpolation error. Depending on the nature of the lidar “ground” observations, the interpolation may decrease observable error, as with the deciduous category in this study, or improve the elevation surface. The interpolation process may or may not be influenced by the actual terrain slope. Particularly for a dense set of elevation points on homogeneous terrain (i.e., an inclined planar surface), the final interpolation on an inclined

TABLE 4. OBSERVED LIDAR ELEVATION ERROR AT POINTS AND INTERPOLATION ERROR AT POINTS (ELEVATIONS IN CM)

	Cover Type					
	Pavement	Low Grass	High Grass	Brush/Low Trees	Evergreen	Deciduous
$RMSE_{Observed\ Lidar\ Pts}$	18.9	22.5	18.9	23.3	17.2	25.9
$RMSE_{Observed\ in\ TIN}$	**22.1**	**25.8**	**22.2**	**26.6**	**17.6**	**23.5**

TABLE 5. OBSERVED AND ADJUSTED LIDAR ELEVATION ERROR (CM)

	Cover Type					
	Pavement	Low Grass	High Grass	Brush/Low Trees	Evergreen	Deciduous
Mean Slope	1.67°	2.27°	1.08°	4.15°	2.30°	2.50°
No Interpolation						
$RMSE_{Observed\ Lidar\ Pts}$	18.9	22.5	18.9	23.3	17.2	25.9
$RMSE_{Lidar\ System}$	18.4	21.4	18.5	21.3	16.2	25.2
$RMSE_{Observed\ Lidar\ Pts}$ on 25° slope (5° stddev)	46.2	47.5	46.3	47.4	45.4	49.3
TIN Interpolation						
$RMSE_{Observed\ in\ TIN}$	22.1	25.8	22.2	26.6	17.6	23.5
Interpolation Error ($RMSE_{Interp}$)	11.5	12.6	11.6	12.8	3.7	−12.9
$RMSE_{Observed\ in\ TIN}$ on 25° slope (5° stddev)	47.6	49.1	47.7	49.1	45.5	47.6

planar surface may be as accurate as an interpolation on a planar surface. Based on Equation 3, estimates of the "additional" interpolation error can also be made by using the observed RMSE from the TIN interpolation (from Table 4) as the dependent variable and solving for the interpolation contribution: i.e.,

$$RMSE_{Interp} = \sqrt{RMSE^2_{observed\ in\ TIN} - RMSE^2_{LIDAR\ System} - RMSE^2_{Survey} - RMSE^2_{Horiz,Slope}}. \tag{8}$$

The results indicate that the RMSE introduced by interpolation ($RMSE_{Interp}$) ranges from −12.9 cm (an improvement) to +12.8 cm (an increase in overall error). Compared to survey error or horizontal displacement error in low slopes, these interpolation errors are much larger. However, because the interpolation error is also not a simple additive term, the final observed error in the surface model ($RMSE_{Observed\ in\ TIN}$) only increases from 0.4 cm to 3.3 cm (Table 5). As noted above with the deciduous category, the interpolation can even decrease the overall error in the final terrain model.

Summary

The results of this study clearly show the variation in elevation error by land-cover category. Surprisingly, the evergreen forest category exhibited low errors similar to those of the pavement category. The RMSE for all categories are very low compared to other sources of digital elevation data used for county-wide mapping projects or national programs (e.g., photogrammetry, IFSAR, etc.). These data collected at the 2-m nominal posting are within the standards approved by FEMA for North Carolina's state flood plain mapping program (http://www.ncfloodmaps.com, last accessed 22 October 2003). For this fairly dense set of lidar data (i.e., a 2-m nominal posting), the additional error introduced by interpolation is very low, adding up to 3.3 cm to any land-cover class. In fact, interpolation may even improve the final representation of terrain elevation in surface models. The interpolation error (introduced by TIN linear interpolation) is larger than the surveyor error but does not have a major impact on the total error budget. The horizontal error of lidar points is typically large (roughly 120 cm RMSE) and is expected to have a significant impact on the observed elevations in steeper terrain. Because the horizontal error is a function of flying height, and the aircraft would likely be flying at higher altitudes above the valley floors, the horizontal error may be even greater. The error budget model developed here helps one to understand the relative amounts of error introduced in the final terrain model (either a DEM or TIN). Our model predicts observable errors on 25° slopes as twice those on relatively low slopes (i.e., less than 4°). The four categories of error investigated in this study could be further refined to identify dominant parameters. For example, in steeper terrain, it is hypothesized that an interactive effect of terrain slope with other factors (e.g., point-labeling process, lidar footprint, scan angle, etc.) may influence the overall observed error in such a way as to further decrease accuracy. Future research could investigate these relationships and others, such as other nominal posting densities, point-labeling algorithms, and the accuracy of surface form (e.g., slope, aspect, curvature).

References

Adams, J.C., and J.H. Chandler, 2002. Evaluation of lidar and medium scale photogrammetry for detecting soft-cliff coastal change, *Photogrammetric Record*, 17(99):405–418.

Bowen, Z.H., and R.G. Waltermire, 2002. Evaluation of light detection and ranging (lidar) for measuring river corridor topography, *Journal of the American Water Resources Association*, 38(1):33–41.

Cobby, D.M., D.C. Mason, and I.J. Davenport, 2001. Image processing of airborne laser altimetry data for improved river modeling, *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(2):121–138.

Cowen, D.J., J.R. Jensen, C. Hendrix, M.E. Hodgson, and S.R. Schill, 2000. A GIS-assisted rail construction econometric model that incorporates lidar data, *Photogrammetric Engineering & Remote Sensing*, 66(11):1323–1326.

Dowman, I., and P. Fischer, 2001. A comparison of airborne IfSAR and LIDAR data over the Vaihingen test site, *Proceedings of the OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Elevation Models* (K. Torlegard and J. Nelson, editors), 01–03 March, Stockholm, Sweden (OEEPE Publication No. 40, available from the Bundesamt für Kartographie und Geodaesie, Frankfurt am Main, Germany, or at http://www.oeepe.org/publications/pdf/no_40.pdf, last accessed 23 October 2003), pp. 160–169.

Hodgson, M.E., J.R. Jensen, L. Schmidt, S. Schill, and B. Davis, 2003. An evaluation of lidar- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs, *Remote Sensing of Environment*, 84:295–308.

Lloyd, C.D., and P.M. Atkinson, 2002. Deriving DSMs from lidar data with kriging, *International Journal of Remote Sensing*, 23(12): 2519–2524.

MacEachren, A.M., and J.V. Davidson, 1987. Sampling and isometric mapping of continuous geographic surfaces, *The American Cartographer*, 14(4):299–320.

Maling, D.H., 1989. *Measurements from Maps*, Pergamon Press, New York, N.Y., 577 p.

Raber, G.T., J.R. Jensen, S.R. Schill, and K. Schuckman, 2002. Creation of digital terrain models using an adaptive lidar vegetation point removal process, *Photogrammetric Engineering & Remote Sensing*, 68(12):1307–1315.

(Received 20 September 2002; accepted 05 November 2002; revised 21 February 2003)