A Multiscale Curvature Algorithm for Classifying Discrete Return LiDAR in Forested Environments

Jeffrey S. Evans and Andrew T. Hudak

***Abstract*—One prerequisite to the use of light detection and ranging (LiDAR) across disciplines is differentiating ground from nonground returns. The objective was to automatically and objectively classify points within unclassified LiDAR point clouds, with few model parameters and minimal postprocessing. Presented is an automated method for classifying LiDAR returns as ground or nonground in forested environments occurring in complex terrains. Multiscale curvature classification (MCC) is an iterative multiscale algorithm for classifying LiDAR returns that exceed positive surface curvature thresholds, resulting in all the LiDAR measurements being classified as ground or nonground. The MCC algorithm yields a solution of classified returns that support bare-earth surface interpolation at a resolution commensurate with the sampling frequency of the LiDAR survey. Errors in classified ground returns were assessed using 204 independent validation points consisting of 165 field plot global positioning system locations and 39 National Oceanic and Atmospheric Administration–National Geodetic Survey monuments. Jackknife validation and Monte Carlo simulation were used to assess the quality and error of a bare-earth digital elevation model interpolated from the classified returns. A local indicator of spatial association statistic was used to test for commission errors in the classified ground returns. Results demonstrate that the MCC model minimizes commission errors while retaining a high proportion of ground returns and provides high confidence in the derived ground surface.**

***Index Terms*—Classification, curvature, digital elevation model (DEM), filtering, forestry, interpolation, light detection and ranging (LiDAR), thin-plate spline, vegetation.**

I. Introduction

LIGHT detection and ranging (LiDAR) vendors have put forth substantial investment into developing ground identification algorithms. Unfortunately, most available algorithms are proprietary and only available in commercial software or through "in-house" vendor postprocessing at considerable cost. A large portion of LiDAR survey cost is incurred during postprocessing to separate ground from nonground returns. This is in large part due to the manual editing required to "clean up" the surface and the need to reclaim software development or purchase costs. Ground returns are often removed from the point cloud and delivered separately in the form of a gridded raster or uniformly spaced point data. Both of these deliverables generalize the data, discarding the original spatial precision and volume of sampled ground returns. The proprietary nature of these filtering methods and the format of provided data make their validation difficult. Furthermore, overfiltering, smoothing, or gridding ground returns compromises the original volume of data, precluding analyses relying on relationships between ground and nonground point volumes.

Retaining a larger volume of ground returns will result in an interpolated ground surface more closely matching the scale and precision of the original LiDAR point cloud [1] and will greatly benefit any applications concerned with microtopography. Improved DEMs will in turn increase accuracy and precision in the derived vegetation heights, benefiting vegetation modeling applications. If the scale of an interpolated DEM is commensurate with the LiDAR data, simply subtracting the interpolated ground elevation surface from the LiDAR elevations will provide accurate vegetation heights at each point while preserving density relationships related to vegetation cover, such as the proportion of nonground to total returns [2].

The majority of LiDAR classification approaches have demonstrated utility in ground mapping or feature extraction (e.g., buildings) in urban environments. However, few solutions have specifically addressed identification of nonground returns in forested environments. A few notable approaches have been proposed for classifying LiDAR data in forested environments [1], [3], but perhaps the most room for improvement in the quality of LiDAR-derived DEMs and subsequent vegetation characterizations lies in the initial classification of ground versus nonground returns [3]. Separating ground from nonground returns is challenging due to the convolution of ground and vegetation returns in the unclassified LiDAR point cloud. Classification approaches that overly smooth or reduce the volume of ground returns (e.g., block minimums) are not well suited for many applications, which can benefit greatly from maximizing retention of ground returns while minimizing commission errors (labeling a valid nonground point as ground). An omission error (discarding a valid ground point) is less problematic than a commission error in terms of influence on the interpolated ground surface. An automated and objective LiDAR classification algorithm that appropriately balances commission omission errors will facilitate applications that exploit the 3-D nature of LiDAR, particularly in forestry [4]–[9].

Scale often has a profound effect on the behavior of landscape objects, such as trees. This is due to the topology of objects that do not behave in a uniform linear manner across scales [10]. As an object is assessed across scales, it can split into multiple objects, merge with other objects, persist, disappear, or

Manuscript received December 16, 2005; revised November 6, 2006. This work was supported in part by the Rocky Mountain Research Station and in part by industry partners Potlatch Inc. and Bennett Lumber Products, Inc.

The authors are with the Moscow Forestry Sciences Laboratory, Rocky Mountain Research Station, USDA Forest Service, Moscow, ID 83843 USA (e-mail: jevans02@fs.fed.us).

Digital Object Identifier 10.1109/TGRS.2006.890412

U.S. Government work not protected by U.S. copyright.

in the case of LiDAR, change its amplitude (deviation from surrounding points) [11]. All of these behaviors occur in LiDAR point clouds as a classification/filtering model is applied. Changing object topology across scales argues for an iterative multiscale model applied directly to the LiDAR point cloud.

Algorithms that identify nonground returns based on curvature thresholds within local neighborhoods appear robust for separating ground from nonground in unclassified LiDAR point clouds in forested environments, particularly in complex terrain [3]. Previous attempts to account for slope and scale effects have used variable window sizes or weighting functions to address the interactions of topography and vegetation on LiDAR returns [12]–[15]. However, these methods have been most effective for classifying nonground returns in gentle terrain, open canopy, or urban environments, rather than in, perhaps, the most challenging environment of densely forested complex terrain.

Curvature approaches identify positive local deviations from surrounding points and then iteratively classify them as nonground. Haugerud and Harding [3] presented a virtual deforestation (VDF) algorithm based on identifying positive local curvatures in an interpolated triangulated irregular network (TIN) constructed from LiDAR point data. Once curvatures are identified, TIN nodes that exceed a defined threshold are iteratively deleted, and the model is run again with the remaining data until a convergence threshold is met. Interpolation of TIN models is based on a linear weighting function applied to a vector connecting adjacent points [16]. In complex terrain, this can distort surface morphology, often resulting in large errors [17]. These commission and omission errors can add noise to the surface, particularly in settings with steep slopes and high canopy density.

We present multiscale curvature classification (MCC), an automated approach for classifying LiDAR returns that incorporates the strengths of curvature filtering [3], adds a scale component in the interpolation phase, and a variable curvature tolerance to account for slope interaction with the LiDAR measurements [3]. The model iteratively classifies nonground LiDAR returns that exceed positive curvature thresholds at multiple scales. The three primary improvements of the MCC model on the VDF model are given as follows: 1) thin-plate spline (TPS) interpolation [18], [19]; 2) integration of a multiscale approach where the surface is interpolated at different resolutions; and 3) a progressive curvature tolerance. Unlike TIN, TPS interpolation allows for adjustment of tension between points, providing adherence to input data and controlling the distance at which point samples affect the surface estimates [20]. The spline tension parameter f also reduces large deviations in the surface between sample points, effectively reducing noise [20]. Inclusion of a multiscale approach more comprehensively addresses topological relationships of nonground objects across variable scales. Because the surface is generalized as the scale parameter is changed, it is necessary to change the curvature threshold parameter, effectively addressing changes in slope effect as the data are generalized. In summary, MCC is an automated algorithm that iteratively identifies nonground points that exceed positive curvature thresholds across multiple scales in unclassified LiDAR point clouds.

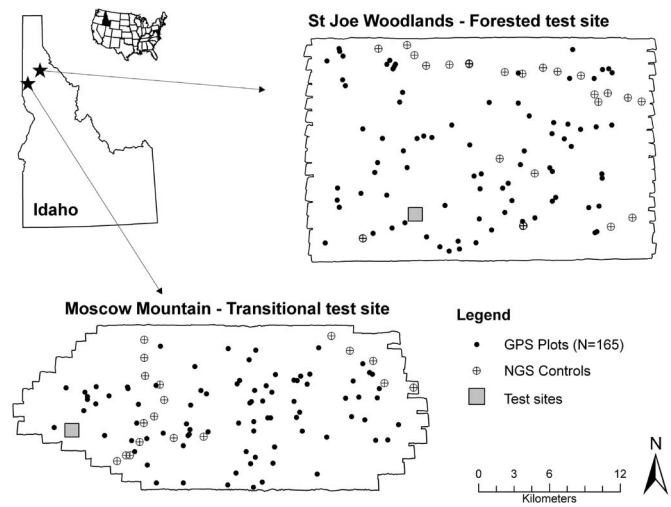

Fig. 1. Location map of the two LiDAR survey areas and test sites.

II. Objectives

The primary goal was to automatically, objectively, and consistently identify ground returns from unclassified LiDAR point clouds. Four criteria were applied in developing the MCC algorithm. The first was flexibility across different LiDAR data formats. The MCC algorithm requires only the X, Y, and Z coordinates as inputs, the three most basic components common to all discrete return LiDAR data sets. The second criterion was simplicity in model parameters. Overly complex or numerous model parameters can introduce subjectivity or result in lack of repeatability. However, the only user-defined parameters in the MCC algorithm are the initial scale and curvature tolerance parameters. The third criterion was that the model be fully automated, i.e., the user need not intervene during the classification filtering process, which can also introduce subjectivity. Finally, the algorithm should preclude or minimize the need for postprocessing, which can be a very time-consuming, costly, and subjective process. The MCC algorithm was designed to objectively, reliably, and repeatability identify ground returns with minimal to no need for manual postprocessing, from any discrete return LiDAR data set acquired over a forested environment.

III. Methods

A. Study Area

The two LiDAR survey areas used in this paper, Moscow Mountain and the Saint Joe Woodlands in north central Idaho (Fig. 1), comprise 88 000 ha of mixed conifer forest with a mixture of industrial forest, private, state, and federal ownerships. Both areas are highly relieved with more topographic complexity in the St. Joe Woodlands. Conifer species vary from *Pinus ponderosa* at xeric sites to *Tsuga heterophylla* in mesic settings. Other common conifer species include *Abies grandis*, *Abies lasiocarpa*, *Larix occidentalis*, *Picea engelmannii*, *Pinus contorta*, *Pinus monticola*, *Pseudotsuga menziesii*, and *Thuja plicata*, which create diverse forest compositions across broad temperature and moisture gradients. The canopy and age

structures are highly variable between stands and often within stands. Both landscapes are actively managed for multiple uses including timber production, wildlife habitat, and watershed protection, representing a large range of anthropogenic and natural influences.

Representative test sites were subsets from each of the larger project areas for a more detailed and intensive accuracy assessment than the entire project area would allow. The first 250-ha site was selected in the St. Joe Woodlands in dense forest on moderate to steep slopes. The second 250-ha site was selected in the Moscow Mt. area across a transitional forest–prairie ecotone with agricultural lands, remnant prairie, open forest canopy, and milder slopes.

B. LiDAR Acquisition

Horizons Inc., a LiDAR data provider located in Rapid City, SD, acquired the LiDAR data in three surveys during the summer of 2003. A three-return OPTECH ALS40 LiDAR sensor was flown with a pulse repetition rate of 20 kHz onboard a Cessna 310 aircraft at an altitude of 2438 m above mean terrain. The nominal post spacing was 1.95 m with a 30-cm footprint. The swath had a nominal width of 904 m and a 20° maximum off-nadir scan angle. LiDAR surveys were flown with an automatic gain control to adjust return sensitivity over variable-intensity terrain at the 1064-nm wavelength. Twelve ground controls were surveyed using a survey grade Trimble 4400 dual-frequency GPS, achieving a horizontal accuracy of < 1 cm through triangulated differential correction with three continuously operating reference stations (CORSs) and three locally established base stations using the Trimble Total Control software. The surveyed controls were used by the vendor to provide a local georectification solution for the LiDAR point cloud, giving the LiDAR data a horizontal and vertical error of < 0.106 m, as reported by Horizons, Inc.

C. LiDAR Processing

Data were delivered from Horizons, Inc., in a generic binary file format containing X and Y coordinates (UTM Zone 11 NAD83), orthometric elevation Z (NADV88), return level (1, 2, or 3), and intensity (0–255). Binary files were converted to ASCII files, utilizing an in-house application provided by Horizons, Inc. Individual flight lines were then converted into point coverages in ArcInfo. Flight lines were trimmed to a maximum scan angle of 15° for the St. Joe Woodlands and 18° for Moscow Mt. to minimize measurement error resulting from high scan angles [21] while preserving complete coverage (i.e., no data gaps between flight lines). The larger 18° threshold for Moscow Mt. was necessary due to less overlap between flight lines compared to St. Joe Woodlands. Scan angles were not included with the delivered data, requiring that flight line centers be digitized. An Arc Macro Language (AML) program was developed to determine the scan angle of each LiDAR measurement based on their calculated distance from the digitized center of the flight line and known aircraft height. Data exceeding 15° or 18° scan angle thresholds were discarded.

D. MCC

Implementation of the MCC algorithm first requires definition of a vector $Z(s)$, which comprised the X coordinate, Y coordinate, and Z (elevation) of all LiDAR returns. The $Z(s)$ vector is used to interpolate a raster surface using a TPS [18]–[20] at a cell resolution defined by scale parameter λ. A TPS is fit using a variable window with the 12 nearest neighbors and an invariant tension parameter f (defined as 1.5 across all scale domains). A 3×3 mean kernel is passed over the interpolated raster defining a new vector $x(s)$, which is coincident with $Z(s)$, consisting of the X coordinate, Y coordinate, and mean surface value. The curvature tolerance t is then added to $x(s)$, and points are classified as nonground by applying the conditional statement "IF $Z(s) > c$ THEN classify as nonground." As the algorithm moves through iterations, $Z(s)$ is redefined as only points not yet classified; these remaining returns are then used at the start of the next iteration. The λ parameter, which initially should approximate the nominal postspacing of the LiDAR data, is calculated in three scale domains: first as 0.5λ, second as λ, and third as 1.5λ. The scale domain l is a model loop where a set of model parameters are run until convergence. A curvature tolerance parameter t in the initial scale domain $(l = 1)$ is defined by the user, and in subsequent scale domains, 0.1 is added.

The classification steps are given as follows.

1) A surface is interpolated using $Z(s)$ and a TPS. Two model parameters for scale domain l are applied, namely: 1) the scale parameter λ and 2) the curvature tolerance t (initial λ and t parameters are user defined).
2) A 3×3 mean kernel is passed over the interpolated surface, and a new vector $x(s)$ is declared that is coincident with $Z(s)$.
3) Curvature in scale domain l is then calculated by

$$c = x(s) + t \tag{1}$$

where $x(s)$ is the mean elevation vector coincident with $Z(s)$ in scale domain l and t is the curvature tolerance parameter in scale domain l.

4) LiDAR points are then classified as nonground and removed if they meet the following condition:

$$\text{IF } Z(s) > c \text{ THEN classify as nonground} \tag{2}$$

where c is the curvature in scale domain l and $Z(s)$ is the measured LiDAR elevation.

5) The convergence threshold j is then assessed, and the model either iterates or starts with the next scale domain.

MCC has two levels where iteration occurs, namely: 1) each scale domain $(l = 1{-}3)$ where parameters are defined and changed when the convergence criteria are met and 2) a nested loop defining model convergence j within l (Fig. 2), where a single set of model parameters is used. Once the model converges within a scale domain, the λ and t parameters are changed, and the algorithm proceeds to the next scale domain until the convergence threshold j is reached in the final scale

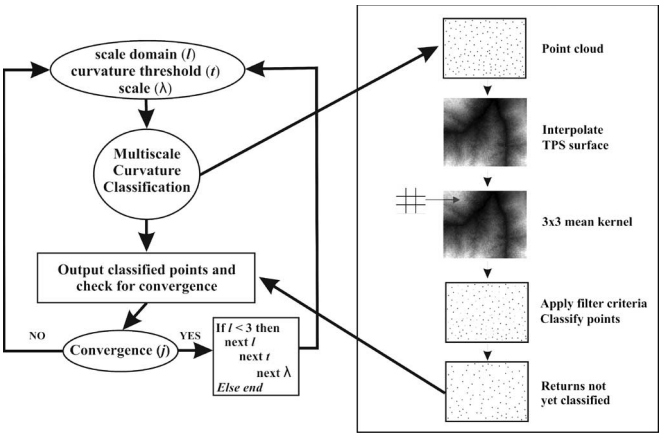

Fig. 2. Flow diagram of the MCC model, where $l = 1-3$ scale domains, and t and λ are parameters defining curvature tolerance and scale, respectively.

domain, upon which MCC terminates. A convergence threshold of 0.1 was used for this paper, meaning the model terminated when less than 0.10% of nonground returns were classified from the returns remaining from the last iteration.

Negative blunders are a common occurrence in LiDAR data, which may be caused by the scattering of the photons in a returned laser pulse [3]. Scattering lengthens the time for an emitted laser pulse to return to the aircraft sensor, inflating the calculation of distance traveled, hence causing a measurement error where the surface elevation is erroneously recorded as being below the surrounding measurements. It should be noted that curvature classification approaches can potentially remove valid returns surrounding negative blunders, which can expand the edge artifact around a negative blunder to create a distinct "bomb crater" effect [3]. To address negative blunders, Haugerud and Harding [3] suggested setting the curvature tolerance parameter to four times the interpolated cell size and selecting returns exceeding this negative curvature threshold. However, it should be noted that under certain circumstances, returns that appear to be negative blunders can be in fact valid returns (e.g., sinkholes). Therefore, the preceding suggestion to remove potential negative blunders can be implemented as an optional last model loop to employ at the discretion of the user if needed.

E. Assessment of Interpolators Within MCC

Three alternative interpolation methods to TPS were tested in the MCC model, namely: 1) Ordinary Kriging; 2) Inverse Distance Weighting; and 3) TIN. The fit of each all-return surface derived from the unclassified LiDAR point data was evaluated using the root-mean-square error (RMSE) statistic (3) calculated across ten jackknife replicates, withholding 10% of the data in each replicate. The critical consideration in assessment of interpolators within MCC was the fit of the data to the derived surface throughout the classification process. Since interpolation error will be most apparent in rougher surfaces fit before reaching the final model solution, an assessment of the unclassified LiDAR data surface fit is the best indicator of interpolator performance within the model.

F. Validation

The classified ground returns were interpolated into 1-m bare-earth DEMs for the two LiDAR survey areas (Fig. 1) using an iterative finite-difference (IFD) interpolation method [22]. Note that the IFD was not an option for incorporation into the MCC model because of the inability of the model to fit surfaces generated by all of the LiDAR returns. Smith *et al.* [23] explored the effect of different interpolators on DEM accuracy and showed that the IFD model provided the most accurate results for ground estimates derived from LiDAR. This model is readily available using the Topogrid command in ArcInfo and is specifically designed for creating DEMs [22].

The two DEMs were used as proxies for assessing the MCC ground return classification results. They were validated at 204 locations (Fig. 1) distributed across the entirety of the two project areas, using two independent data sets, namely: 1) field GPS locations $(N = 165)$ collected during June–August of 2003 and 2004, with a minimum of 200 static measurements logged with a Trimble ProXT and differentially corrected with online base station files to a reported horizontal accuracy of < 0.6 m and a vertical accuracy of < 2.0 m (Trimble Pathfinder Office software), and 2) geodetic control monument locations $(N = 39)$ obtained from the National Geodetic Survey (NGS) (http://www.ngs.noaa.gov/cgi-bin/ds_radius.prl).

Validation points were distributed across the entirety of the two project areas, making these independent validation data appropriate for assessing any global elevation bias in the DEMs. Residual error was calculated by subtracting the independent validation elevations at the 204 coincident ground locations from the DEM surface elevations. RMSE (3) was calculated and used in conjunction with the residual error distribution to assess global bias.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{S}_i - z_i)^2}{n-1}} \tag{3}$$

where

$\hat{S}_i$ predicted;
z_i observed;
n number of observations.

G. Evaluation

More rigorous accuracy assessment was applied to the two 250-ha test sites than was feasible across the entire LiDAR survey areas (Fig. 1). A jackknife validation was conducted, in each subset test site, to evaluate the quality and error of the DEMs derived from the classified LiDAR ground returns. The jackknife validation was run with 100 iterations, randomly withholding 10% (forested $n = 13\,544$; transitional $n = 16\,003$) of the classified ground returns in each iteration and interpolating a new DEM from the remaining data, using an IFD interpolation model. So as not to ignore multiple sources of error [24], a total RMSE (4) was calculated, which included both the vertical measurement error reported by the LiDAR vendor $(\mathrm{RMSE}_m = 0.106\ \mathrm{m})$ and the interpolation error across

jackknife replicates (RMSE_i). Accuracy was assessed using the cumulative $\mathrm{RMSE}_{\mathrm{total}}$ across all jackknife replicates and the residual error calculated by subtracting the withheld ground return heights from the corresponding height values in the interpolated ground surface, as follows:

$$\mathrm{RMSE}_{\mathrm{total}} = \sqrt{(\mathrm{RMSE}_m)^2 + (\mathrm{RMSE}_i)^2} \tag{4}$$

where RMSE_m is the RMSE of measured LiDAR elevations and RMSE_i is the RMSE of interpolated LiDAR elevations.

Chrisman [25] refers to a raster surface as "a distribution of possible realizations in which the true value lies." This is referred to as a spatial random field or a stochastic image. Uncertainty in the DEM can be modeled by a set of realizations of a spatial random field to provide a range of error that bounds the "true" value [26]. To assess DEM quality, a Bayesian Monte Carlo approach was implemented by setting the distributional priors in each simulation to the range of the original DEM, with the addition of the vendor reported measurement error (0.106 m) defining the allowable standard deviation in the random field [27]. One thousand Monte Carlo simulations were conducted with replacement on the interpolated DEM. Residual error was calculated in each simulation by generating a random field following the preceding criteria and applying a mean filter to account for autocorrelated errors [26]. The original DEM was then subtracted from the random field. RMSE (3) was calculated for each simulation condition, and the cumulative RMSE was calculated across all 1000 simulations [28]. Mapping cumulative RMSE allowed assessment of the spatial distribution of uncertainty [26], [29].

A local indicator of spatial association (LISA) statistic was used on the classified ground returns within the two test sites to identify commission errors based on discontinuities in the local autocorrelation [30]. A spatial weighting matrix was calculated using the six nearest neighboring ground returns and the associated local autocorrelation value calculated at each point. Spatial outliers were identified, and the neighborhood deviation from the mean using these six nearest neighbors was calculated to determine the magnitude of the potential error.

Finally, visual inspection was used in conjunction with the results of the LISA statistic to assess overall surface quality and examine potential commission errors, which are apparent visually due to their obvious discontinuity from the surrounding surface when viewed in shaded relief.

IV. Results

A. Model Performance

The curvature tolerance parameter t was sensitive to slope effects and λ and therefore was tested on a subset of the data to determine the best starting value. In this paper, initial parameters of $\lambda = 1.5$ and $t = 0.3$ consistently identified nonground returns. The majority of nonground returns were identified in the first scale domain; the volume of additional nonground returns identified in subsequent scale domains asymptotically declined (Fig. 3, Table I). Processing the returns across three scale domains (Fig. 4) was sufficient to classify nonground returns while minimizing omission errors.

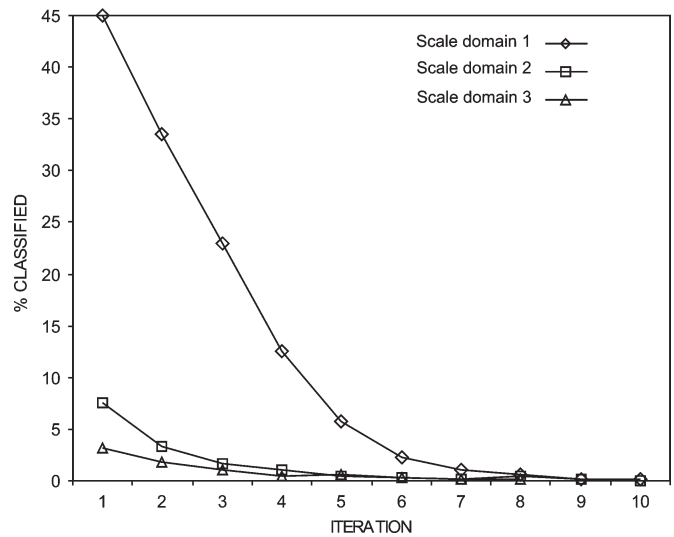

Fig. 3. Percent of returns classified in each of the three scale domains of MCC operation in the transitional forest test site.

TABLE I
Number of Total LiDAR Returns and Nonground and Ground Returns Classified at the Two Selected Test Sites

Test Site	Total Returns	Non-Ground Returns (%)	Ground Returns (%)
Forested	995,075	859,632 (86.4%)	135,443 (13.6%)
Transitional	302,292	124,039 (41.0%)	178,253 (59.0%)

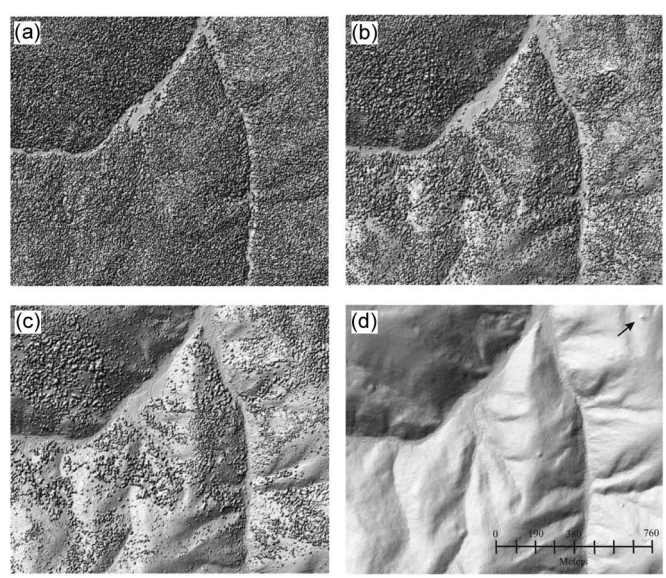

Fig. 4. Shaded relief of interpolated classification results at four stages of MCC operation in each scale domain at the dense forest test site. (a) Preclassified. (b) Convergence in scale domain 1. (c) Convergence in scale domain 2. (d) Convergence in scale domain 3 (final result). The largest potential commission error is indicated by the arrow.

B. Assessment of Interpolators Within MCC

In the testing of different interpolation methods within the MCC model, TPS was the best interpolator of the four alternatives considered (Table II; Fig. 6). Ordinary Kriging produced the poorest fit to the unfiltered point clouds and caused some surface erosion [Fig. 6(a)], and proved to be overly processor intensive and unrealistic to implement due to difficulty

TABLE II
FORESTED TEST SITE JACKKNIFE RMSE AND RESIDUAL ERROR RANGES (IN METERS) OF FOUR INTERPOLATION METHODS TESTED IN THE MCC MODEL FOR INTERPOLATING UNCLASSIFIED LIDAR RETURNS

Interpolation Method	RMSE	Minimum residual	Maximum residual
Ordinary Kriging	4.8881	-28.016	32.892
Inverse Distance Weighting	4.8210	-26.883	26.679
Triangulated Irregular Network	4.5082	-29.695	27.008
Thin Plate Spline	1.8071	-8.094	10.475

TABLE III
RESIDUAL ERROR STATISTICS AND RMSE (IN METERS) BETWEEN DEMS INTERPOLATED FROM CLASSIFIED LIDAR GROUND ELEVATIONS AND TWO INDEPENDENT VALIDATION DATA SETS

STATISTIC	GPS FIELD PLOTS			NGS MONUMENTS		
	St. Joe	Moscow Mt.	All	St. Joe	Moscow Mt.	All
Minimum	-15.6186	-53.8300	-53.8300	-6.6304	-5.7145	-6.6304
Maximum	12.3020	33.7933	33.7933	4.7700	8.8110	8.8110
Mean	0.4100	-1.7868	-0.6951	0.2390	1.1048	0.6014
Median	0.0205	-0.5899	0.0000	0.0000	0.4555	0.0000
St. Dev	4.4618	10.7677	8.3077	2.3931	4.6679	3.5041
RMSE	0.4100	1.7655	0.6951	0.2390	0.9470	0.6157
Count (N)	81	84	165	21	18	39

in automatically and consistently fitting semivariogram functions throughout model iterations. Inverse distance weighting introduced curvature artifacts that resulted in dramatic errors, in some cases misclassifying entire hill slopes as nonground [Fig. 6(b)]. TIN maintained exact values at the sample locations but produced large deviations between LiDAR sample points, and significantly eroded some hillslopes [Fig. 6(c)]. The TPS produced the best fit to the unclassified LiDAR data (Table II) and was the only interpolator that did not erode the ground surface [Fig. 6(d)].

C. Validation

The range and standard deviation of the field GPS residuals were higher than those from the NGS monument residuals. This is most likely due to NGS monument locations predominately occurring in valley bottoms or on ridge tops with little canopy cover, while the field GPS plots mainly occurred beneath dense canopy, which diminishes field GPS accuracy. The NGS monument elevations were expected to be much more accurate than the nonsurvey grade field GPS elevations, although only 5 of 39 NGS monuments had a vertical accuracy of < 1 cm (i.e., level B or better vertical control). Across both study areas, the mean residual error was -0.695 for the GPS locations and 0.601 for the NGS monuments (Table III). These are small biases considering the large magnitude of some outliers (especially in the field GPS elevations) that inflated the residual error and RMSE [24], [31].

RMSE alone failed to provide information about the residual error distributions, such as the balance between small and large deviations, and skewness [32]. The frequency distribution of the residuals can be used to indicate the direction and magnitude of a bias [16], uncovering possible systematic or conditional sources of error. In this paper, the median values of zero in both the field GPS plot and NGS monument residual distributions (Table III) suggested little bias in the classified ground returns. Limiting the NGS monument validation data to just the five locations with level B or better vertical control produced a RMSE of only 0.04 m across the two LiDAR survey areas.

TABLE IV
RESIDUAL ERROR DISTRIBUTION STATISTICS (IN METERS) ACROSS 100 JACKKNIFE REPLICATES IN EACH TEST SITE

Residual Error Statistics	Forested	Transitional
Minimum	-1.108	-0.3715
Maximum	1.288	0.5716
Mean	-0.008	0.0407
Median	-0.009	0.0357
Standard Deviation	0.287	0.121
Skew	0.139	0.445
Kurtosis	0.954	1.795

D. Evaluation

Results from the jackknife validation indicated that the $\mathrm{RMSE}_{\mathrm{total}}$ in the dense forest test site was 0.306 and 0.166 in the transitional forest test site. The slightly higher $\mathrm{RMSE}_{\mathrm{total}}$ in the dense forest test site can be attributed to steeper slopes and denser canopy, excluding ground returns. Analysis of the residual error across all jackknife replicates demonstrated normally distributed errors (Table IV) in both test sites, indicating minimal directional bias in the distribution. The mean residual errors in the dense forest and transitional forest test sites were -0.0081 and 0.0407, respectively, indicating very small biases. The mean and median nearly match, indicating a symmetric error distribution (Table IV). The standard deviation (Table IV) is consistently quite small, showing a constrained range of error. A bivariant q–q plot of interpolated versus measured values (not shown) showed the same distributional shape and variance in both test sites, all supporting a very good model fit.

Monte Carlo simulation identified areas with higher uncertainty that represent strong slope curvatures, commission errors, or areas with comparatively high sample density. Subtle horizontal bands in the RMSE surface of forested areas [more obvious at the dense forest test site, Fig. 6(a)] coincided with overlapping flight lines, where the sampling density was effectively doubled. Although it seems counterintuitive that a higher sample density should increase the RMSE, this could be related to areas of higher sample density causing greater fine scale variability than surrounding areas.

Two landscape features produced unusually high RMSE ($\sim$3 m) in the transitional forest test site. The first is a building [circle 1 in Fig. 6(b)]. The second is a small forested knoll with a drainage ditch directly adjacent [circle 2 in Fig. 6(b)]. Although the nonground returns were successfully removed from the knoll, the juxtaposition of the knoll and ditch created a local area of higher uncertainty. With these two exceptions, the simulation results indicated high confidence in the DEMs and identified no areas exhibiting extreme uncertainty.

The LISA statistic effectively identified potential commission errors and allowed the evaluation of the magnitude of each error. The largest potential commission error identified at the dense forest test site consisted of two points that deviated from the mean of the six nearest neighbors by 5.96 m and is apparent upon close visual inspection of the interpolated DEM surface (Fig. 4). This potential commission error and others less

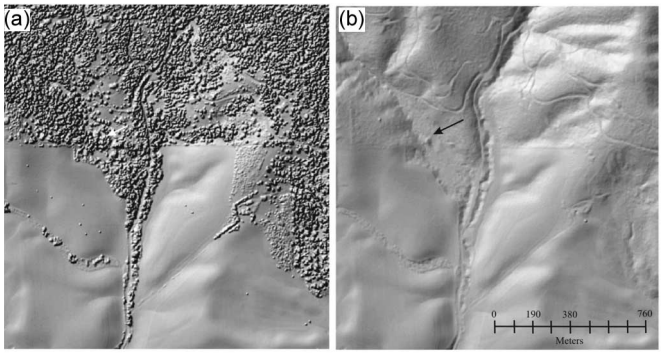

Fig. 5. Before and after shaded relief of the interpolated surface of all returns and ground returns from the transitional forest test site. A commission error caused by a building is indicated by the arrow.

pronounced could be real nonground features of the landscape (e.g., stumps or woody debris) that the MCC model failed to eliminate. In the transitional forest test site, the same building already mentioned caused the largest commission error, consisting of six points that deviated from the mean of the six nearest neighbors by 8.06 m (Fig. 5).

V. Discussion

A. Model Performance

Vegetation density largely determines the ratio of ground-to-vegetation returns. The contrast between the vegetation structures and the selected test sites is evident in the percentages of ground and nonground returns identified (Table I). The vast majority of nonground returns are removed within the first scale domain (Figs. 3 and 4). By starting the first scale domain with a cell size of approximately half the postspacing, the MCC model first removes those returns producing the highest curvatures at a finer resolution. As nonground points are removed, the remaining data become more generalized. In the next scale domain, the scale parameter changes to a larger cell size, effectively compensating for this smoothing of the data and thus changing the resulting surface topology. Since the curvature tolerance parameter compensates for slope effect (which increases as resolution coarsens), the sensitivity of the curvature tolerance parameter changes as well. After the first few iterations of the model within a given scale domain, most extreme curvatures are classified, and each subsequent iteration further refines the classification results. After the first scale domain converges, the majority of large surface curvatures have been classified, and subsequent scale domains, in effect, smooth small surface modulations created by low understory vegetation and ground debris. In summary, the scale and curvature tolerance parameters work in concert to account for changes in the topology of objects in the point cloud as nonground returns are removed from consideration. Varying the scale and curvature tolerance parameters in the MCC algorithm robustly addresses variable canopy configurations interacting with slope, as manifested in the LiDAR point cloud.

Abrupt edge features (e.g., buildings) typically will not exceed the positive curvature thresholds used by the MCC algorithm to identify nonground returns. Reconfiguring the algorithm to identify and remove such edge features would erode or remove road and stream bank cuts. Since this algorithm is designed for forested landscapes where road cuts typically far outnumber buildings, retention of these features is imperative. Morphological filters have been shown to be effective for removing buildings [15], and future research should explore the possibility of combining multiple classification approaches to exploit the strengths of each.

A few other limitations of the MCC model need to be addressed. Processing time is a current drawback to implementation of the MCC model. Fitting splines on high volumes of data is very time and processor intensive, and there is much room for improvement. Mitasova *et al.* [20] propose a spline model that uses quadtree segmentation for handling large data volumes and appears to be quite promising. Implementation of MCC in a more efficient language than AML and integration of the quadtree approach [20] would dramatically improve the speed and efficiency of the model. In the LiDAR data used in this paper, an attribute indicating "last return" was not provided so it was not considered in the implementation of the model. Running the model on last returns would increase speed by reducing the overall volume of the data.

B. Interpolators Within MCC

While TPS clearly outperformed the other interpolators tested within MCC (Table II), further investigation of the influences of tension and smoothing in TPSs may provide more flexibility across a large range of LiDAR data sets [20]. Three scale domains were adequate for good model performance in this study, but different forest structures or LiDAR acquisition parameters (i.e., very dense post spacing) may require changes in the curvature threshold parameter and/or the number of scale domains required.

C. Validation

A variety of validation methods were chosen in this paper to more rigorously evaluate the performance of the MCC algorithm and quality of the derived DEM. Many factors influence the creation of a DEM, including the interpolation method, resolution, and sampling error. Further consideration of these factors was beyond the scope of this paper but have been addressed by Smith *et al.* [23], [33]. The idea of using a derived DEM surface as a proxy for classification quality has been introduced in previous research [34] as a means of validating the large volume of data that LiDAR surveys provide. Lewis and Hutchinson [35] empirically demonstrated that the details of the probability distribution of spatially correlated errors are far less important in determining data quality than the inherent quality of the derived surface.

The intention of including the nonsurvey grade field GPS and NGS control validation points was not to assess local accuracy but to identify any global bias in the classified ground elevations across a number of independent validation points. The vertical error residuals from the GPS and NGS validation may be primarily random, given that the median errors were zero (Table III), which supports the idea of assessing global

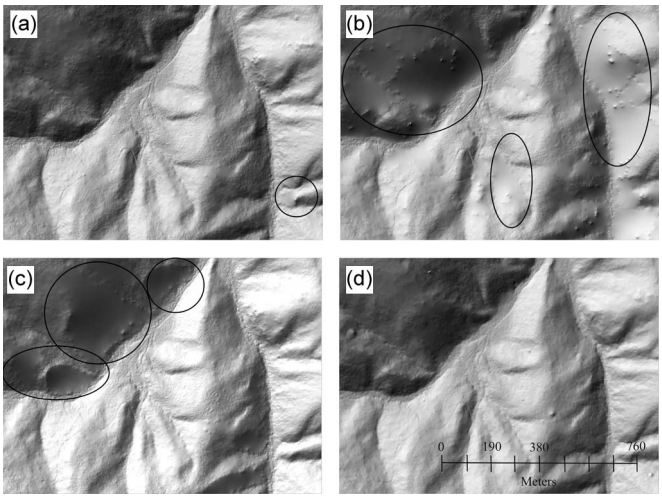

Fig. 6. Shaded relief's comparing different interpolators within MCC. (a) Ordinary Kriging. (b) Inverse distance weighting. (c) Triangular irregular network. (d) TPS.

bias using validation data less accurate than the LiDAR data [36]. If a representative validation sample were to be calculated based on the number of observations in the LiDAR, the number of field samples required for 95% confidence intervals would be virtually impossible to acquire [37], precluding many traditional validation techniques.

D. Evaluation

Less understory vegetation and surface clutter will increase certainty in the ground estimate, as demonstrated by the Monte Carlo simulations (Fig. 6). Including the LiDAR measurement error in the total RMSE reported for the jackknife validation and Monte Carlo simulation results more honestly reflects the cumulative error in the estimate of ground elevation and more conservatively assesses the accuracy of the derived DEM. The LISA point pattern statistic provides an unbiased empirical method for validating the classified ground points and quantifying the magnitude of identified errors. The results illustrate that potential commission errors are kept to a minimum (Figs. 4 and 7). Potential commission errors can vary from low amplitude errors representing understory vegetation to entire nonground objects that the model fails to identify [3], [15]. Some level of surface modulation is to be expected in forested areas due to low understory vegetation, logs, stumps, and other ground debris [3].

Commission and omission errors play an important role in LiDAR point classification. Omission errors are more difficult to observe and quantify than commission errors. However, if occurring in moderation omission errors has less influence on the quality of the resulting DEM. Generally, the results indicated no severe biases in classified ground returns, either globally (Table III) or locally (Table IV, Figs. 4 and 7), suggesting that MCC appropriately balanced commission and omission errors.

The definition of "ground" is application and scale dependent, making it easy to disconnect the resolution of an interpolated DEM surface estimate from the resolution of the LiDAR sample data. Block minimums are a clear example of a

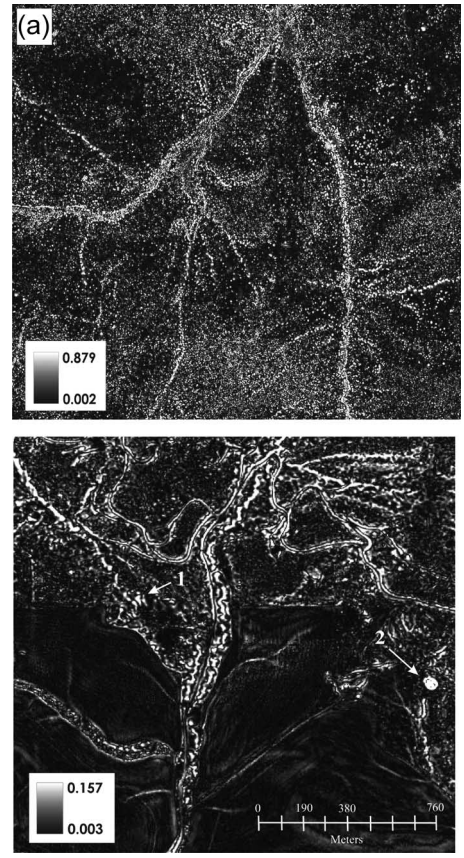

Fig. 7. Cumulative RMSE from 1000 Monte Carlo simulations in (a) the forested test area and (b) transitional forest test area. The RMSE range portrayed in the legend in (b) excludes two features (indicated by white arrows) that produced the RMSEs of ~3 m.

LiDAR filtering strategy that can potentially disconnect a surface estimate from the resolution of the LiDAR sample data. Increasing the block size to identify the minimum ground height effectively smoothes the data to a resolution that is coarser than the LiDAR sampling frequency. This produces a contiguous grid of ground estimates that is consistently derived across the landscape, which may be convenient for some applications, but is decoupled from the highly variable nonground samples that are of primary interest for many applications. The more sophisticated MCC classification strategy is most suited for forested landscapes with or without complex terrain. The MCC algorithm is efficient in that it maximizes the number of classified ground returns, making it an attractive algorithm for a variety of applications.

VI. Conclusion

The complex vertical arrangement of LiDAR data requires an iterative processing approach to converge on an accurate solution of identified nonground returns. The multiscale facet of the MCC model is perhaps its most useful innovation. By addressing topologic differences across scales [10], MCC effectively identifies nonground returns in the vertical distribution of elevations in the LiDAR point cloud. The dynamic scale and curvature tolerance parameters work in concert to account for changes in the topology of objects in the point cloud as

nonground returns are removed from consideration. Exploiting multiscale topological relationships removes much of the low-level surface modulation created by understory vegetation that other filtering approaches often miss while simultaneously maintaining a high volume of ground returns. The quality of the LiDAR-derived DEM directly affects the quality of the LiDAR-derived canopy heights used in subsequent vegetation modeling. The MCC model employs minimal parameters and requires only the most basic data inputs (X, Y, and Z), providing flexibility and repeatability across LiDAR data sets. Further research is needed to evaluate how well the MCC algorithm handles different vegetation canopy geometries (e.g., broadleaf deciduous) and topographic settings (e.g., canyonlands). These differences may affect the user specification of the number of scale domains (model loops), scale parameter (resolution within a scale domain), curvature tolerances, or the spline parameters.

The MCC model is currently implemented in AML running under Workstation ArcInfo. The development of MCC in ArcInfo was undertaken to provide a classification tool that is readily available to the LiDAR user community. Current and subsequent version(s) of the code are available for download at http://forest.moscowfsl.wsu.edu/gems/lidar/.

ACKNOWLEDGMENT

This research is a product of the Sustainable Forestry component of Agenda 2020, a joint effort of the USDA Forest Service Research & Development and the American Forest and Paper Association. Research funds were provided by the Rocky Mountain Research Station and industry partners Potlatch, Inc. and Bennett Lumber Products, Inc. The authors would like to thank A. Davidson for her assistance in postprocessing of the LiDAR data and T. Rice and M. Murphy for manuscript reviews. They would also like to thank two anonymous reviewers who provided thorough and thoughtful reviews, making this manuscript much stronger.

REFERENCES

[1] K. Kraus and N. Pfeifer, "Determination of terrain models in wooded areas with airborne laser scanner data," *ISPRS J. Photogramm. Remote Sens.*, vol. 53, no. 4, pp. 193–203, Aug. 1998.

[2] J. E. Means, S. A. Acker, J. Brandon, B. J. Fritt, M. Renslow, L. Emerson, and C. Hendrix, "Predicting forest stand characteristics with airborne scanning LiDAR," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 11, pp. 1367–1371, 2000.

[3] R. A. Haugerud and D. J. Harding, "Some algorithms for virtual deforestation (VDF) of LiDAR topographic survey data," *Int. Arch. Photogramm. Remote Sens.*, vol. XXXIV-3/W4, pp. 211–217, 2001. [Online]. Available: pugetsoundlidar.ess.washington.edu/vdf4.pdf

[4] H.-E. Andersen, R. J. McGaughey, and S. E. Reutebuch, "Estimating forest canopy fuel parameters using LiDAR data," *Remote Sens. Environ.*, vol. 94, no. 4, pp. 441–449, Feb. 2005.

[5] R. O. Dubayah and J. B. Drake, "LiDAR remote sensing for forestry applications," *J. For.*, vol. 98, no. 6, pp. 44–46, Jun. 2000.

[6] M. A. Lefsky, W. B. Cohen, S. A. Acker, G. G. Parker, T. A. Spies, and D. Harding, "LiDAR remote sensing of the canopy structure and biophysical properties of douglas-fir western hemlock forests," *Remote Sens. Environ.*, vol. 70, no. 3, pp. 339–361, Dec. 1999.

[7] M. A. Lefsky, W. B. Cohen, G. G. Parker, and D. J. Harding, "LiDAR remote sensing for ecosystem studies," *BioScience*, vol. 52, no. 1, pp. 19–30, Jan. 2002.

[8] K. Lim, P. Treitz, K. Baldwin, I. Morrison, and J. Green, "LiDAR remote sensing of biophysical properties of tolerant northern hardwood forests," *Can. J. Remote Sens.*, vol. 29, no. 5, pp. 658–678, Oct. 2003.

[9] M. Maltamo, K. Eerikainen, J. Pitkanen, J. Hyyppa, and M. Vehmas, "Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 319–330, Apr. 2004.

[10] L. Florack and A. Kuijper, "The topological structure of scale-space images," *J. Math. Imaging Vis.*, vol. 12, no. 1, pp. 65–79, Feb. 2000.

[11] G. J. Hay, P. Dube, A. Bouchard, and D. J. Marceau, "A scale-space primer for exploring and quantifying complex landscapes," *Ecol. Model.*, vol. 153, no. 1, pp. 27–49, Jul. 2002.

[12] S. L. Hyun and H. N. Younan, "DTM extraction of LiDAR returns via adaptive processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2063–2069, Sep. 2003.

[13] J. Kilian, N. Haala, and M. Englich, "Capture and evaluation of airborne laser scanner data," *Int. Arch. Photogramm. Remote Sens.*, vol. XXXI, pp. 383–388, 1996. [Online]. Available: www.ifp.uni-stuttgart.de/publications/1996/laser_wien.pdf

[14] G. Vosselman, "Slope based filtering of laser altimetry data," *Int. Arch. Photogramm. Remote Sens.*, vol. XXXIII, pp. 935–942, 2000. [Online]. Available: www.itc.nl/personal/vosselman/papers/vosselman2000.adam.pdf

[15] K. Zhang, S.-C. Chen, D. Whitman, M.-L. Shyu, J. Yan, and C. Zhang, "A progressive morphological filter for removing nonground measurements from airborne LiDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 872–882, Apr. 2003.

[16] E. H. Isaaks and R. M. Srivastava, *An Introduction to Applied Geostatistics*. New York: Oxford Univ. Press, 1989.

[17] M. B. Gousie and W. R. Franklin, "Augmenting grid-based contours to improve thin-plate DEM generation," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 1, pp. 69–79, Jan. 2005.

[18] I. Briggs, "Machine contouring using minimum curvature," *Geophysics*, vol. 39, no. 1, pp. 39–48, 1974.

[19] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*. New York: McGraw-Hill, 1995.

[20] H. Mitasova, L. Mitas, and R. S. Harmon, "Simultaneous spline approximation and topographic analysis for LiDAR elevation data in open-source GIS," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 4, pp. 375–379, Oct. 2005.

[21] J. Holmgren and M. Nilsson, "Simulating the effects of LiDAR scanning angle for estimation of mean tree height and canopy closure," *Can. J. Remote Sens.*, vol. 29, no. 5, pp. 623–632, Oct. 2003.

[22] M. F. Hutchinson, "A new procedure for gridding elevation and stream line data with automatic removal of spurious pits," *J. Hydrol.*, vol. 106, no. 2, pp. 211–232, Apr. 198.

[23] S. L. Smith, D. A. Holland, and P. A. Longley, "Investigating the spatial structure of error in digital surface models derived from laser scanning data," *Int. Arch. Photogramm. Remote Sens.*, vol. XXXIV, 2003. [Online]. Available: www.isprs.org/commission3/wg3/workshop_laserscanning/papers/Smith_ALSDD2003.pdf

[24] M. E. Hodgson and P. Bresnahan, "Accuracy of airborne LiDAR-derived elevation: Empirical assessment and error budget," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 3, pp. 331–339, Mar. 2004.

[25] N. Chrisman, "Error in categorical maps: Testing versus simulation," in *Proc. Auto-Carto 9: 9th Int. Symp. Comput.-Assisted Cartography, ASPRS/ACSM*, 1989, pp. 521–529.

[26] G. Hunter and M. Goodchild, "Modeling the uncertainty of slope and aspect estimates derived from spatial databases," *Geogr. Anal.*, vol. 29, no. 1, pp. 35–49, 1997.

[27] A. Journel, "Modelling uncertainty and spatial dependence: Stochastic imaging," *Int. J. Geogr. Inf. Syst.*, vol. 10, no. 5, pp. 517–522, 1996.

[28] B. Weschsler, Effects of DEM uncertainty on topographic parameters. Syracuse, NY: College Environ. Sci. and Forestry, State Univ. New York, 2000.

[29] D. Theobald, "Accuracy and bias issues in surface representation," in *The Accuracy of Spatial Databases*, M. F. Goodchild and S. Gopal, Eds., Bristol, PA: Taylor & Francis, 1992, ch. 9, pp. 99–106.

[30] L. Anselin, "Local indicators of spatial association," *Geogr. Anal.*, vol. 27, no. 2, pp. 93–115, 1995.

[31] S. E. Reutebuch, R. J. McGaughey, H.-E. Andersen, and W. W. Carson, "Accuracy of a high-resolution LiDAR terrain model under a confer forest canopy," *Can. J. Remote Sens.*, vol. 29, no. 5, pp. 527–535, Oct. 2003.

[32] J. Wood and P. Fisher, "Assessing interpolation accuracy in elevation models," *IEEE Comput. Graph. Appl.*, vol. 13, no. 2, pp. 48–56, Mar. 1993.

[33] S. L. Smith, D. A. Holland, and P. A. Longley, "The effect of changing grid size in the creation of laser scanner digital surface models," presented at the 7th Int. Conf. GeoComputation, Southampton, U.K, 2003. [Online]. Available: geocomputation.org/2003/Papers/Smith_Paper.pdf

[34] T. L. Webster, "LiDAR validation using GIS: A case study comparison between two LiDAR collection methods," *Geocarto Int.*, vol. 20, no. 4, pp. 1–9, 2005.
[35] A. Lewis and M. F. Hutchinson, "From data accuracy to data quality: Using spatial statistics to predict the implications of spatial error in point data," in *Quantifying Spatial Uncertainty in Natural Resources*, H. T. Mowrer and R. C. Congalton, Eds., Chelsea, MI: Sleeping Bear, 2000, pp. 17–35.
[36] R. C. Daniels, "Datum conversion issues with LiDAR spot elevations," *Photogramm. Eng. Remote Sens.*, vol. 67, no. 6, pp. 735–740, 2001.
[37] Z. Li, "Effects of check points on the reliability of DEM accuracy estimates obtained from experimental test," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 10, pp. 1333–1340, 1991.

Jeffrey S. Evans received the B.S. degrees in ecology and statistics from the University of California, San Diego, in 1990 and the M.S. degree in ornithology and biogeography from the University of California, Berkeley, in 1994.

He has been a Landscape Ecologist with the Moscow Forestry Sciences Laboratory, Rocky Mountain Research Station, USDA Forest Service, Moscow, ID, since 1999. His research interests include spatial interactions in ecological systems, spatial statistics, scale, and using remote sensing and GIS to model wildlife and vegetation dynamics. His current research interests include forest structural modeling using LiDAR, scale and sampling relationships between LiDAR and field measurements, multiscale predictive vegetation modeling, spatial variability of fire severity, and landscape genetics.

Andrew T. Hudak received the B.S. degree in ecology, evolution, and behavior from the University of Minnesota, Minneapolis, in 1990 and the Ph.D. degree in environmental, population, and organismic biology from the University of Colorado, Boulder, in 1999.

Since 2001, he has been a Research Forester with the Moscow Forestry Sciences Laboratory, Rocky Mountain Research Station, USDA Forest Service, Moscow, ID. His general research interest is development and application of novel approaches for extracting useful ecological information from remotely sensed data for the benefit of ecologists and managers. His current research interests include remote sensing of canopy structure and composition, wildfire effects, and vegetation change, across forest and rangeland landscapes.