

# A Multi-Scale Approach to Mapping Canopy Height

Gordon M. Green, Sean C. Ahearn, and Wenge Ni-Meister

## Abstract

Mapping vegetation height over large areas presents a problem of scale: height varies with the individual tree or stand, but the resolution of available datasets is too low to characterize this variability sufficiently for many applications. We address this problem by fusing 1 km resolution canopy height data derived from satellite-based laser altimetry with higher-resolution land-cover data, resulting in 30 m resolution estimates of canopy height. These are downscaled further to 1 m resolution by simulating individual trees. A web service architecture is used, which allows processing to occur on demand without preprocessing large datasets. We compared the resulting canopy volumes to reference airborne lidar data from 262 randomly located 1 km<sup>2</sup> areas within nine study sites. Results at 30 m resolution show an RMSE of 33 percent of the mean reference volume and an  $R^2$  of 0.77; at 1 m the RMSE is 66 percent and the  $R^2$  is 0.38.

## Introduction

Vegetation height is a key measurement used to estimate a variety of ecological and biophysical variables, including above-ground biomass, surface roughness, and stem volume. Global large-footprint lidar data from the Geoscience Laser Altimeter System (GLAS), gathered as part of the Ice, Cloud and land Elevation Satellite (ICESat) mission, have recently been used to create coarse-grained global canopy height datasets (Lefsky, 2010; Simard *et al.*, 2011). However, these datasets do not capture the fine-grained variability inherent in vegetation height, particularly in disturbed or patchy areas. In contrast, 1 m resolution datasets based on airborne lidar do characterize vegetation height with sufficient granularity for a wide range of applications, but are not available for most areas.

Increasingly, geographic datasets are being made available over the World Wide Web using web mapping services. These applications deliver map images representing regions of interest on demand, so users need not download or store entire datasets to make use of the information they contain. While scientific datasets have long been made available for download over the web, part of what makes web mapping services particularly useful is that requests are fulfilled within seconds, making interactive on-the-fly processing possible.

The availability of higher-resolution land-cover data through web mapping services makes possible a new method of

quantifying canopy height: existing GLAS-based coarse-resolution height data can be fused with higher-resolution land-cover data on demand to generate more detailed canopy height maps. This downscaling process bridges the gap between scales in such a way that coarse-resolution canopy height is translated into a form more descriptive of the phenomenon it measures, and on-the-fly processing means that large volumes of high-resolution data need not be downloaded or preprocessed to obtain information about a specific region of interest.

We implemented a proof-of-concept application that performs this downscaling on demand using a web service architecture. In this architecture, the application runs within a web server, using the request/response mechanism of HTTP (hypertext transfer protocol) to perform the downscaling. It responds to requests that include the bounds of an area of interest, and returns a canopy height surface downscaled to the maximum available resolution that can be efficiently generated to fit the dimensions of the requested area.

The proof-of-concept application performs this downscaling in two parts. First, 30 m maximum height surfaces are estimated by combining canopy density from the 2001 National Land Cover Database (NLCD-2001; Homer *et al.*, 2007), and land-cover from NLCD-2006 (Fry *et al.*, 2011), with 1 km GLAS-based estimates of height from Simard *et al.*, 2011. Meter-scale canopy heights are then simulated using a stochastic table lookup approach, constrained by the estimated 30 m height values and the same NLCD data as was used in the 30 m process. The table lookup data consists of height rasters representing individual trees that were extracted from airborne lidar data and indexed by maximum height. These are selected on demand and placed randomly within each applicable 30 m pixel until the pixel is saturated in proportion to its canopy density according to the NLCD-2001 canopy density product. The result is a process that simulates 1 m resolution canopy height models (CHMs) on demand for any location within the coterminous United States, while storing only the 1 km height and table lookup data.

An overview of the three scales is shown in Figure 1: the background image shows the GLAS-based 1 km resolution canopy height surface from Simard *et al.* (2011); the first inset shows a 30 m resolution height surface that disaggregates the GLAS-based height values using NLCD land-cover and NLCD canopy density data; and the second inset shows a 1 m simulated canopy height model, based on the 30 m results and NLCD data.

Questions raised by the approach include: Do the results provide a reasonable representation of the vertical and horizontal distribution of vegetation within each 1 km

---

Gordon M. Green and Sean C. Ahearn are with the Center for the Advanced Research of Spatial Information, Department of Geography, Hunter College of the City University of New York, 695 Park Avenue, New York, NY 10021 (gordongreen@earthlink.net).

Wenge Ni-Meister is with the Department of Geography, Hunter College of the City University of New York, 695 Park Avenue, New York, NY 10021.

---

Photogrammetric Engineering & Remote Sensing  
Vol. 79, No. 2, February 2013, pp. 185–194.

0099-1112/13/7902-185/\$3.00/0  
© 2013 American Society for Photogrammetry  
and Remote Sensing

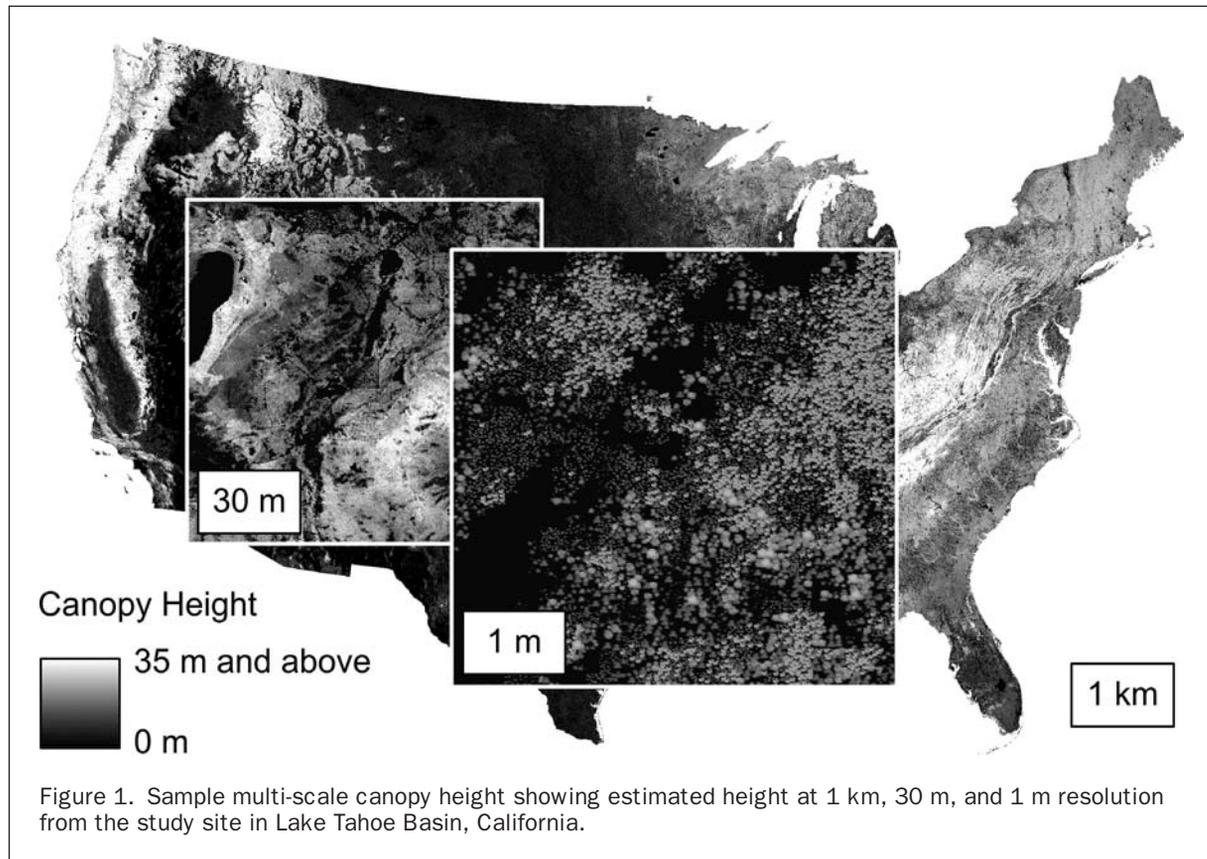


Figure 1. Sample multi-scale canopy height showing estimated height at 1 km, 30 m, and 1 m resolution from the study site in Lake Tahoe Basin, California.

pixel? Which method of estimating the vertical distribution of canopy heights results in the most accurate downscaling? Do temporal differences in the source data affect the results? We address these questions by applying the downscaling process to the Simard *et al.* (2011) 1 km height dataset, and comparing the results to CHMs derived from airborne lidar. We test four types of height distributions for their effects on accuracy. To assess the effects of changes over time, we also test the downscaling process using land-cover data from both NLCD-2006 and NLCD-2001, and use Landsat-based change data to compare results with and without areas identified as having changed between 1990 and 2005, or between 2000 and 2005.

## Source Data

### Web Mapping Services

The Web Map Service (WMS) interface standard (Beaujardiere, 2006) published by the Open Geospatial Consortium (OGC) defines an interface by which map data can be shared over the World Wide Web. It specifies the data elements required in each conforming HTTP request and response, such as the bounding coordinates and their spatial reference, the image format, and the pixel data type. One of the requirements of the standard is that WMS-compliant services must respond to GetMap requests, whereby map images are generated for given spatial extents. These requests are typically made by web-based mapping software that displays imagery for a given area when a user navigates to it using an interactive map. Requests may also be made by any software that supports HTTP. The prototype application uses the WMS protocol over HTTP to retrieve NLCD information on demand for any requested location within the coterminous United States, using services made available by the United States Geological Survey (USGS).

It then performs the downscaling operation, and returns the results to the original requestor in the form of an HTTP response that includes a link to a downscaled height raster, which can be used for display or further processing.

We used two 30 m resolution datasets to constrain the downscaling process. For land-cover type, we used NLCD-2006, which is based on more recent imagery than NLCD-2001 (Fry *et al.*, 2011). Because NLCD-2006 did not include a revision of the 2001 canopy density layer, we used NLCD-2001 for canopy density information. NLCD-2001 canopy density includes values for pixels that may be classified as non-forested, such as those classified as impervious surface or grassland, while eliminating areas that are unambiguously un-forested, such as water bodies (Homer *et al.*, 2007). We chose NLCD-2006 because the underlying data is temporally closer to the GLAS data underlying the Simard, *et al.* (2011) dataset. We also evaluated the results using land-cover data from NLCD-2001 to check whether the methodological and temporal differences between the two NLCD versions affected the results. Landsat-based data representing temporal changes in normalized difference vegetation index (NDVI), compiled from Global Land Survey datasets by the USGS and the National Aeronautics and Space Administration (NASA), and made available through web services by Esri and partners as described in Green (2011), were used to further assess the impact of temporal differences between datasets on the results.

### GLAS-based Canopy Height

Laser altimetry data (Zwally *et al.*, 2002) from the GLAS large-footprint lidar sensor has been shown in recent studies to provide moderately accurate measurements of canopy height with, for example, an RMSE (root mean squared error) of 2.2 m compared to airborne lidar in Lee *et al.* (2011).

Each returned waveform, or shot, consists of a record indicating the timing and amplitude of energy returned from the land surface within a ~65 m diameter footprint ellipse, binned at sub-meter vertical intervals (Harding and Carabajal, 2005). Because the large footprint causes slope and surface roughness to strongly affect the shape and extent of the returned waveform (see Yang *et al.*, 2011), the effects of topography on the shape of the waveform need to be considered. Further, because GLAS shots fall along discontinuous ICESat tracks, with as much as 15 km between tracks at the equator and ~170 m between shots along satellite tracks (Schutz *et al.*, 2005), a method must be developed to estimate canopy height between samples for most mapping applications.

Lefsky (2010) and Simard *et al.* (2011) created continuous global canopy height maps from GLAS data, each using slightly different approaches. Lefsky estimated mean and 90<sup>th</sup> percentile height on a per-patch basis, establishing patch boundaries using 500 m resolution data from the Moderate Resolution Imaging Spectroradiometer (MODIS), addressing the slope problem using an empirical approach based on the shape of the waveform. Simard *et al.* (2011) used a per-pixel regression-tree approach, based on MODIS and gridded climatological data, to develop a continuous 1 km height surface from GLAS data acquired in 2005, eliminating shots using slope-based and other criteria. In Simard *et al.* (2011) height is defined as RH100, the distance from signal start to the ground peak of the waveform, which corresponds closely to maximum canopy height. The dataset provides height information for land-cover types classified as non-forested, resulting in a height attribution for a greater proportion of the landscape.

For this reason, the Simard *et al.* (2011) dataset is used here as the basis of the downscaling process. Its overall vertical accuracy was assessed at RMSE = 6.1 m, R<sup>2</sup> = 0.5 (Simard *et al.*, 2011).

#### Airborne Discrete-Return Lidar

We used airborne lidar data gathered from nine study sites to calibrate and validate the downscaled estimates of canopy height. These were downloaded primarily from *OpenTopography.org* and the National Center for Airborne Laser Mapping (NCALM). Study site locations are shown in Table 1 and Figure 2. The sites were chosen based on the availability of datasets with sufficient sampling density (>1 post per m<sup>2</sup>) to characterize canopy height at the sub-tree level. Lidar-based 1 m bare earth digital elevation models (DEMs) were subtracted from their corresponding first-return digital surface models (DSMs) to create canopy height models (CHMs), mosaicked to form a single reference 1 m CHM for each study site.

Reference 30 m CHMs were created by resampling the reference 1 m CHMs to the same resolution as the CHMs generated by the 30 m downscaling process, so the two can be compared. Each 30 m reference pixel was assigned the 98<sup>th</sup> percentile value of the 1 m pixels falling within it. The 98<sup>th</sup> percentile was chosen to avoid anomalously high pixels, while approximating the maximum height measurement used in the Simard *et al.* (2011) map. The 1 m and 30 m CHMs derived from airborne lidar were used for calibration and validation only, and are not part of the downscaling process.

TABLE 1. STUDY SITES WITH DOMINANT LAND COVER TYPES (E = EVERGREEN NEEDLELEAF; D = DECIDUOUS BROADLEAF; M = MIXED FOREST; G = GRASSLANDS; U = URBAN/MIXED FOREST); THE AREA VALUE REPRESENTS THE PORTION OF THE SITE USED IN THIS STUDY

Site	Location	Land Cover Type	Location (Lon, Lat)	Area (km <sup>2</sup> )	Date (m/yyyy)
1	Boston, MA (suburban)	U	-71.21, 42.26	145	6/2002
2	Flathead Lake, MT	E	-113.81, 48.46	90	5,9/2005
3	Independence Lake, CA	E	-120.33, 39.43	43	7/2007
4	Lake Tahoe Basin, CA	E	-120.00, 38.88	160	8/2010
5	Pleasant, ME	D	-69.33, 45.55	87	10-11/2007
6	Tenderfoot Creek, MT	E	-110.88, 46.93	121	9/2005
7	Tuscaloosa, AL	M	-87.78, 33.24	2	12/2010
8	Yakima, WA	G	-120.53, 46.93	25	4/2008
9	Yosemite National Park, CA	E	-119.59, 37.74	81	7/2007

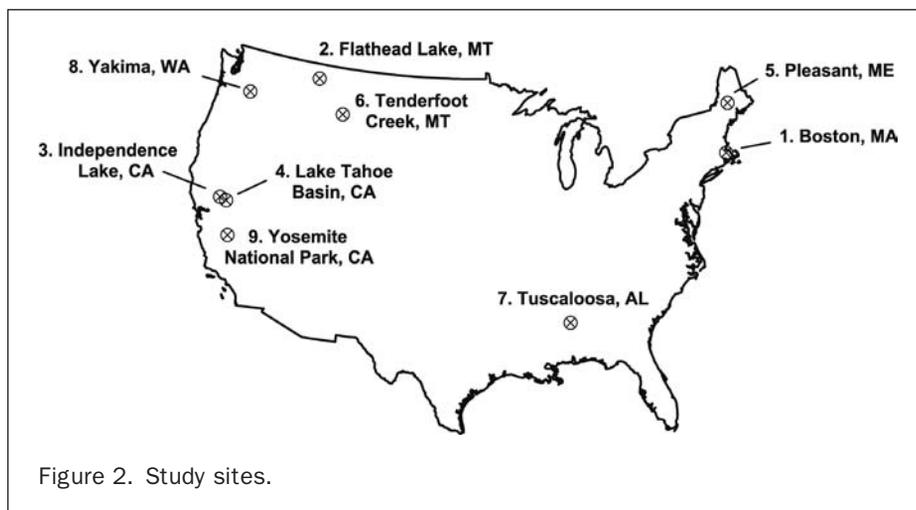


Figure 2. Study sites.

## Methodology

### Thirty Meter Downscaling Process

The process used to downscale from the 1 km Simard *et al.* (2011) dataset to 30 m resolution for a given region of interest (ROI) is illustrated in Figure 3, and consists of the following steps:

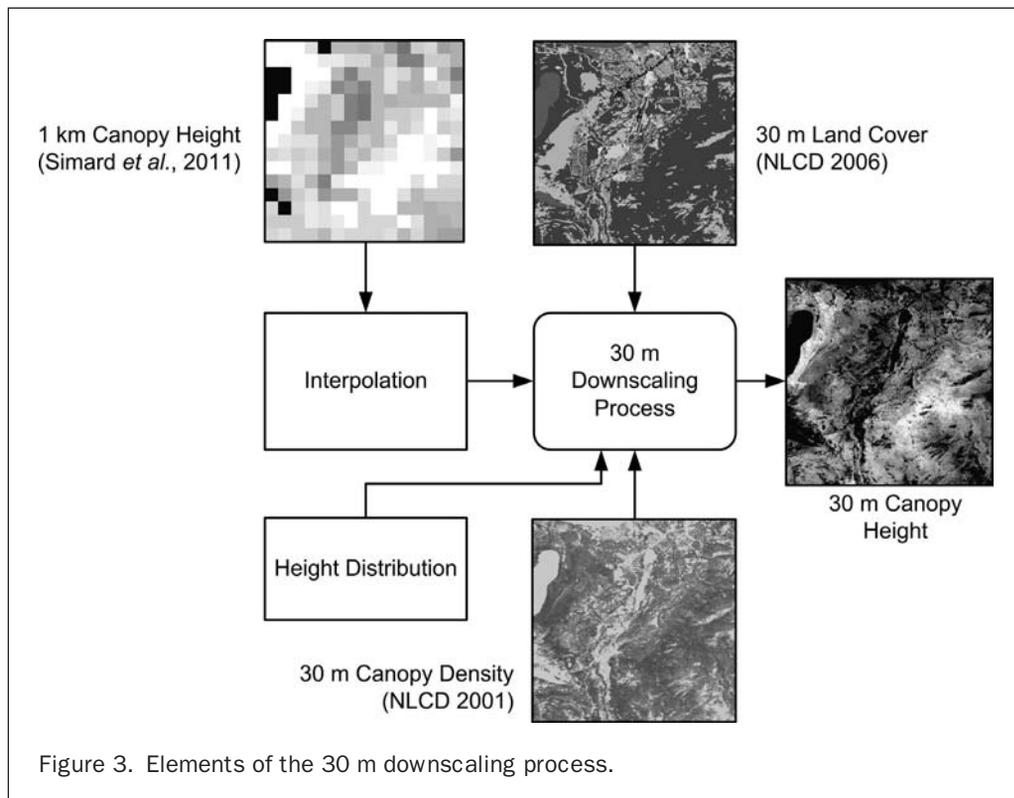
1. Acquire thematic land cover data for the ROI using the NLCD-2006 web mapping service (see USGS EROS Web Map Services reference for details).
2. Reclassify the land-cover data into four simplified classes: evergreen needleleaf, deciduous broadleaf (includes woody wetlands and developed), mixed forest, and low/partial vegetation (shrub, grassland, and emergent herbaceous).
3. Acquire canopy density data for the ROI using the NLCD-2001 web mapping service (see USGS EROS Web Map Services reference for details).
4. Scale the canopy density pixel value to a range from zero to one.
5. Create a 30 m resolution output raster with the same bounds as the ROI and set its pixel values to zero.
6. For each 30 m pixel in the new raster:
  - a. If the value in the corresponding canopy density pixel is below a threshold, assign a canopy height of zero.
  - b. Otherwise, locate the four closest 1 km pixels in the 1 km maximum height raster based on the distance of each 1 km centroid to the 30 m pixel centroid. Weight each of the four height values based on the inverse of its distance, and sum them to arrive at an interpolated height value for the 30 m pixel.
  - c. Sample a height scaling factor (between zero and one) from a height distribution, and multiply this factor by the interpolated maximum height value to determine the estimated height of the 30 m pixel.
  - d. If the land-cover type is low/partial, reduce the estimated height by a constant scaling factor.
  - e. Assign the estimated height value to the pixel in the output raster.
7. Output the resulting 30 m CHM.

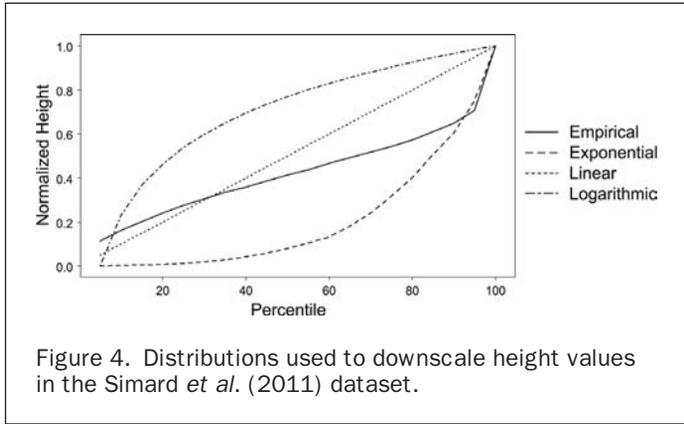
We used the canopy density layer to identify forested pixels in Step 6a because we found, based on comparisons with aerial photographs, that pixels classified as non-forested, such as impervious surface or grassland, often contain trees, whereas the canopy density layer indicates percent tree cover independent of land-cover type. The interpolation in Step 6b is required because, without smoothing, the coarse pixel values in the 1 km height data cause artificial differences at pixel boundaries in the generated 30 m CHM. Simple weighting was chosen over kriging or other more complex interpolators because the goal of the interpolation is to smooth pixel boundaries efficiently, rather than to account for spatial autocorrelation within the height data. The threshold and scaling factor set in Steps 6a and 6d, and the sampling process described in Step 6c, were selected as part of the calibration process described below.

Because the actual vertical distribution of canopy heights in any location is unknown, we configured the system to sample from each of several distributions, to determine which one results in the most accurate downscaling overall. These are linear, exponential, logarithmic, and an empirical distribution derived by averaging the vertical distributions of randomly-selected 30 m reference CHM pixels within the Flathead Lake, Montana, Lake Tahoe Basin, California, and Pleasant, Maine sites. These distributions are shown in Figure 4.

### One Meter Downscaling Process

The results of the 30 m downscaling process are used as the basis of an additional downscaling procedure that builds 1 m CHMs stochastically. We use the canopy density and land-cover data retrieved in Steps 1 and 3 of the 30 m process to constrain the placement of individual tree CHMs drawn from table lookup data. These CHMs were extracted from the lidar-based reference 1 m CHMs and assembled into a table indexed by maximum height. The CHMs represent 20 deciduous trees with heights ranging from 5.4 m to 24.7 m, and 20 evergreen needleleaf trees with heights ranging from 4.3 m to 46.2 m.





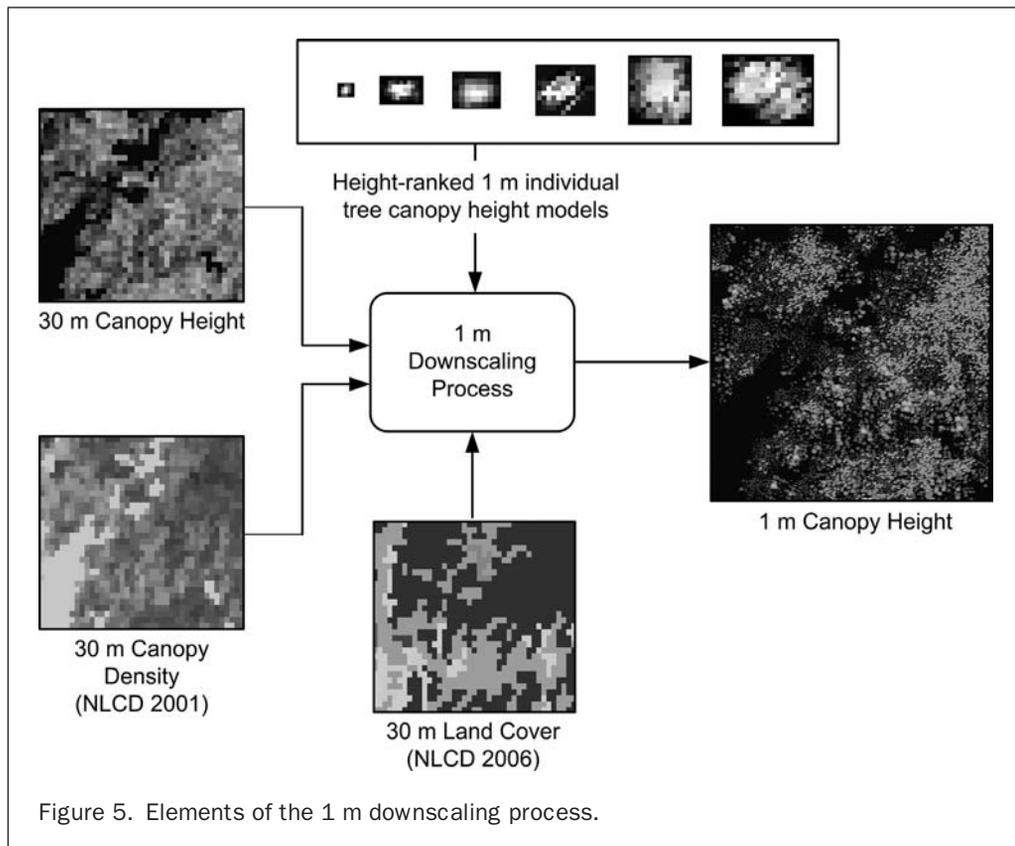
Evergreen needleleaf trees were selected from the lidar-based CHMs from the Flathead Lake, Montana study area, and deciduous trees from the Boston, Massachusetts data. Tree type was determined visually from coincident aerial photography. Trees in open-canopy conditions, where the boundaries of the tree are visually identifiable, were manually selected to represent a range of heights. The 1 m CHMs were cropped to the apparent boundary of each tree, and each resulting small surface was treated as representative of all trees of similar heights, for either the deciduous broadleaf or evergreen needleleaf categories. Sample individual tree CHMs are shown as part of Figure 5, and the processing steps for downscaling a region of interest to 1 m resolution are listed below.

1. Create a 1 m resolution output raster with the same bounds as the ROI and set its pixel values to zero.
2. For each 30 m pixel resulting from the 30 m process described above:

- a. Set the available area to the area of the current 30 m pixel.
- b. Scale the available area by the canopy density associated with the 30 m pixel, which ranges from zero to one.
- c. While the available area is greater than zero:
  - i. Choose a random location within the 30 m pixel boundary in the output raster where no height value has yet been assigned.
  - ii. Calculate a height for the sample tree by selecting a random value from a Gaussian distribution ( $\mu = 0$ ,  $\sigma = 2$  m) and adding it to the height value of the current 30 m pixel as output by the 30 m process, to simulate random variation among individual trees.
  - iii. Select the CHM representing either a deciduous or coniferous tree from the lookup table with the highest maximum height that is less than the calculated height, using the simplified land-cover type from Step 2 of the 30 m process to determine the type of tree.
  - iv. For each 1 m pixel in the selected individual tree CHM:
    01. Add the difference between the calculated height from step ii and the maximum height of the CHM to the pixel value.
    02. Using the selected random location as the center point, write the adjusted value to the corresponding location in the output raster, if its value is greater than the value already written.
    03. If the original value in the destination pixel was zero, decrease the available area by the area represented by the written pixel.
    04. If the tree pixel falls outside of the bounds of the 30 m pixel, write the pixel value to the adjacent 30 m cell similarly.

3. Output the resulting 1 m CHM.

If a height value is found in the 1 km data that is larger than that of the tallest CHM in the lookup data, the difference between the estimated height and the tallest model is added to the pixels of the tallest model, to extrapolate beyond the



range of the available samples. This is particularly relevant in the western United States, where evergreen needleleaf tree heights may be above the 46.2 m maximum found in the lookup table.

There are several limitations and assumptions in the process that affect the accuracy of the results. First, because they are static samples drawn from areas of low canopy density, the individual-tree CHMs do not account for the effects of canopy density or species on crown shape. Second, only deciduous broadleaf and evergreen needleleaf types are included, and are assigned according to land-cover type without attempting to differentiate between species. Third, mixed forest is assumed to be distributed equally between the two types, and low vegetation is treated as deciduous broadleaf. Fourth, the small number of individual-tree CHMs in the lookup table unrealistically limits the variability of crown shapes in the simulated CHMs. Finally, the downscaling is limited in its accuracy to that of the Simard *et al.* (2011) data. Within the coterminous United States, the Simard *et al.* (2011) 1 km data has a maximum height of 54 m, which is less than the height of the tallest trees. We speculate that this may be due to the incomplete coverage of the underlying GLAS data, because ICESat tracks may not intersect the tallest forests within the coterminous United States.

### Calibration

We used the reference data derived from airborne lidar from three of the study sites to verify and calibrate the downscaling process: Flathead, Montana, Lake Tahoe, California, and Pleasant, Maine. A spatially random sample of 1 km<sup>2</sup> areas ( $n = 90$ ) spread across the three sites were identified, and the 30 m process was run for each in stages. Figure 6 shows the main processing steps. We found that, although there is only a weak relationship between canopy density and height ( $R^2 = 0.28$  in the calibration sites), using canopy density as the basis of the sampling improves the apparent accuracy slightly over sampling randomly for each pixel. This corresponds to Step 6c of the 30 m downscaling process. The effect on the output is shown in Figure 6e and 6f. We also found that CHM pixels in the low/sparse category had a lower average height value than the forested classes, and, when estimating the height of a 30 m pixel in the low/sparse category, we scaled the calculated height by 0.6 to approximate this relationship. A canopy density threshold of 0.1 was selected based on the assumption that coverage of less than 10 percent indicates only part of a single tree, or parts of a small number of trees, are likely to fall within the 30 m pixel. The 1 m process is constrained by the results of the 30 m process, and the NLCD data as described above.

### Results and Discussion

The prototype application was run at 30 m and 1 m resolution for 262 randomly selected 1 km<sup>2</sup> areas located across the nine study sites. Downscaled CHMs at each resolution and location were generated using linear, exponential, logarithmic, and empirical height distributions. Figure 7 shows sample 30 m resolution outputs, created using the different height distributions, and representative output of the 1 m process is shown in Figure 8. These results suggest that the process does provide an improved, if approximate, representation of canopy vertical and horizontal variability within 1 km pixels.

Evaluating the results more quantitatively presents a challenge: the underlying relationships between height and NLCD data are weak, and the height value ascribed to any single pixel is likely to be inaccurate, even if the vertical distribution selected for the site is correct. A review of the 30 m downscaled CHMs bears this out. We compared the 30 m resolution

reference CHMs to the 30 m CHMs output from the downscaling process using the logarithmic distribution. The per-pixel agreement is low, with an  $R^2$  of 0.20 and RMSE of 9.2 m, based on a random selection of 2,680 pixels. The problem is exacerbated at 1 m resolution, where the stochastic placement of trees ensures little correlation between simulated and reference CHMs on a per-pixel basis.

To address this limitation, we compared the downscaling results to the reference 1 m and 30 m CHMs using canopy volume, calculated as per-pixel height multiplied by area, summed for each 1 km sample area, correcting for variations in the area represented by each pixel. This approach allows the accuracy of the results to be assessed without relying on

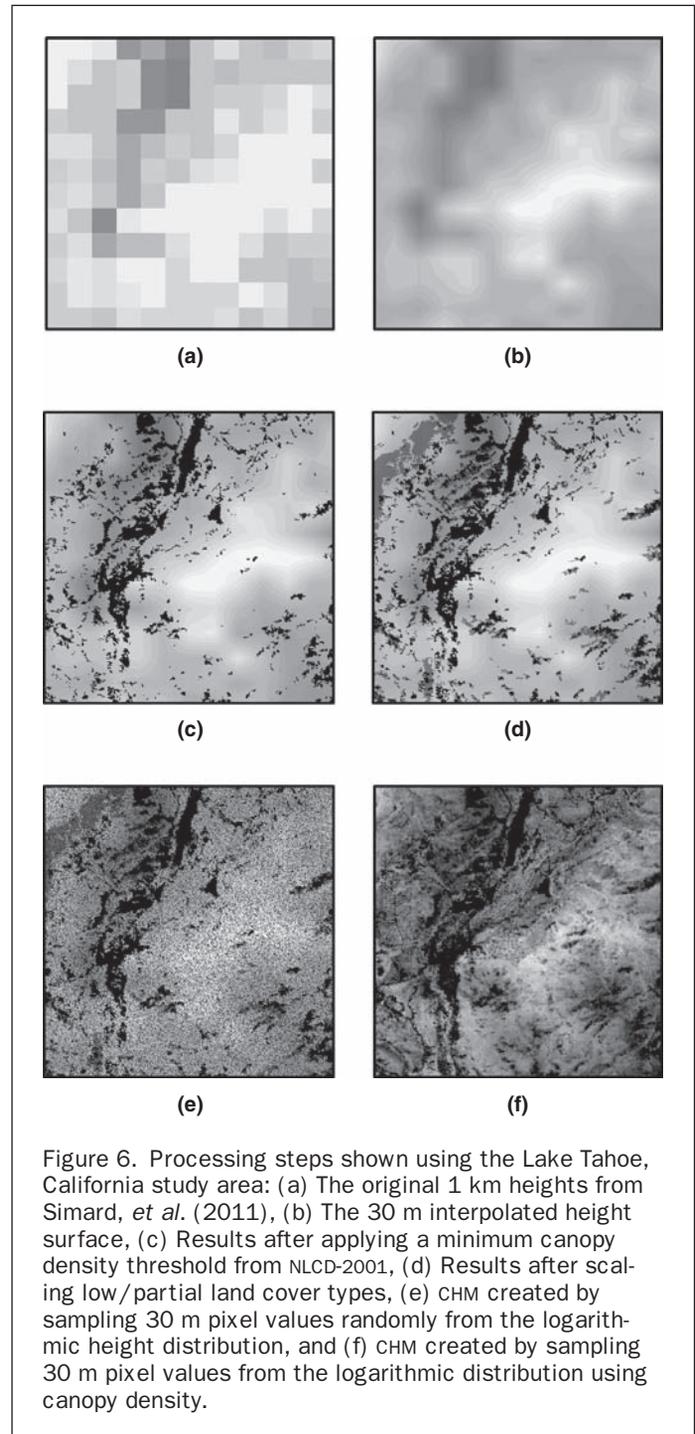
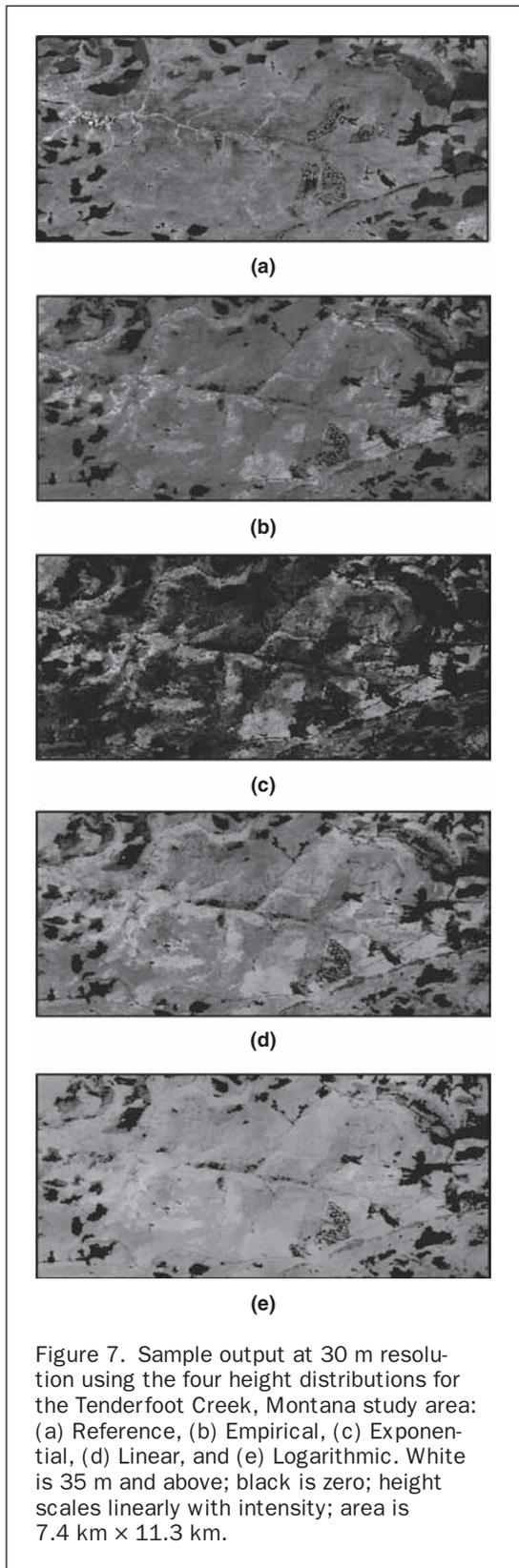


Figure 6. Processing steps shown using the Lake Tahoe, California study area: (a) The original 1 km heights from Simard, *et al.* (2011), (b) The 30 m interpolated height surface, (c) Results after applying a minimum canopy density threshold from NLCD-2001, (d) Results after scaling low/partial land cover types, (e) CHM created by sampling 30 m pixel values randomly from the logarithmic height distribution, and (f) CHM created by sampling 30 m pixel values from the logarithmic distribution using canopy density.



per-pixel agreement. Canopy volume, however, is highly variable between locations, and difficult to compare as a result. Therefore, we used the unit-less ratio of RMSE in  $m^3$  to mean reference volume in  $m^3$  to characterize the error. As shown in Table 2, the logarithmic distribution results in the most

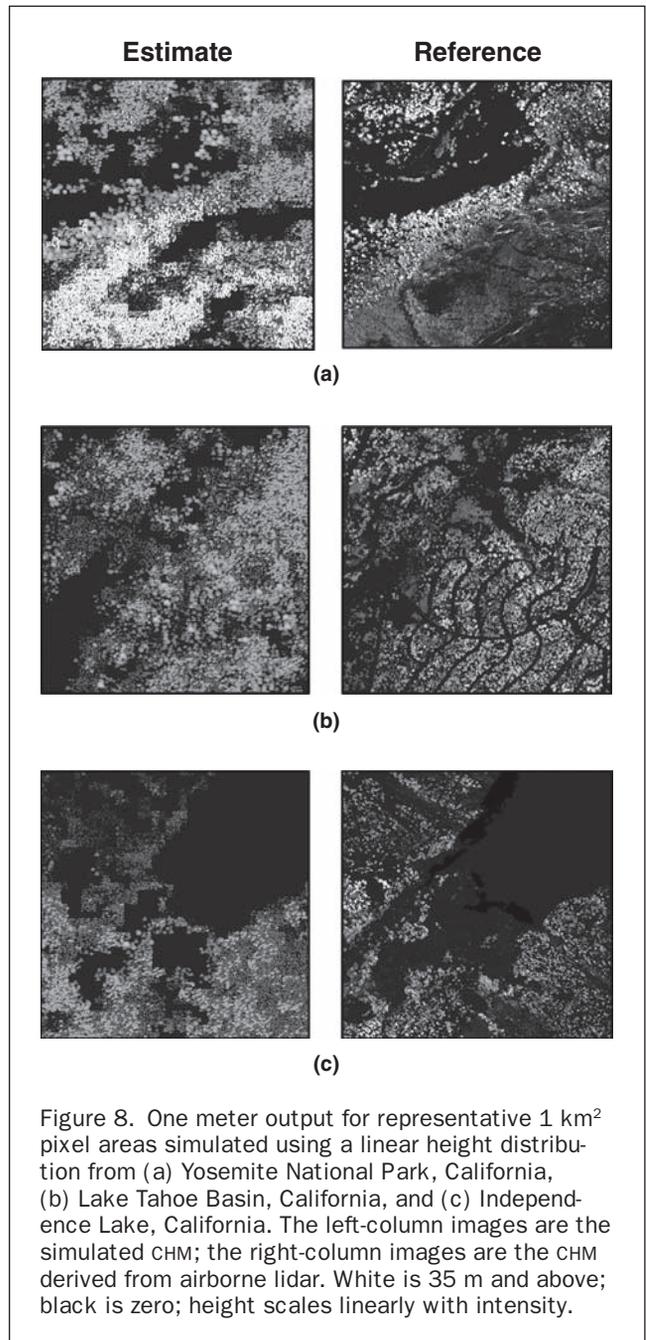


TABLE 2. OVERALL ACCURACY OF DOWNSCALING USING DIFFERENT HEIGHT DISTRIBUTIONS IN RANDOMLY-PLACED 1 km<sup>2</sup> AREAS. RMSE IS EXPRESSED AS M<sup>3</sup> DIVIDED BY MEAN REFERENCE CANOPY VOLUME IN M<sup>3</sup>. THE TWO LEFT COLUMNS SHOW THE STATISTICS CALCULATED FOR THE ESTIMATED VERSUS REFERENCE 1 M CHMS, AND THE TWO RIGHT COLUMNS SHOW THE RESULTS FOR THE ESTIMATED VERSUS REFERENCE 30 M CHMS, WITH  $n = 262$  FOR BOTH

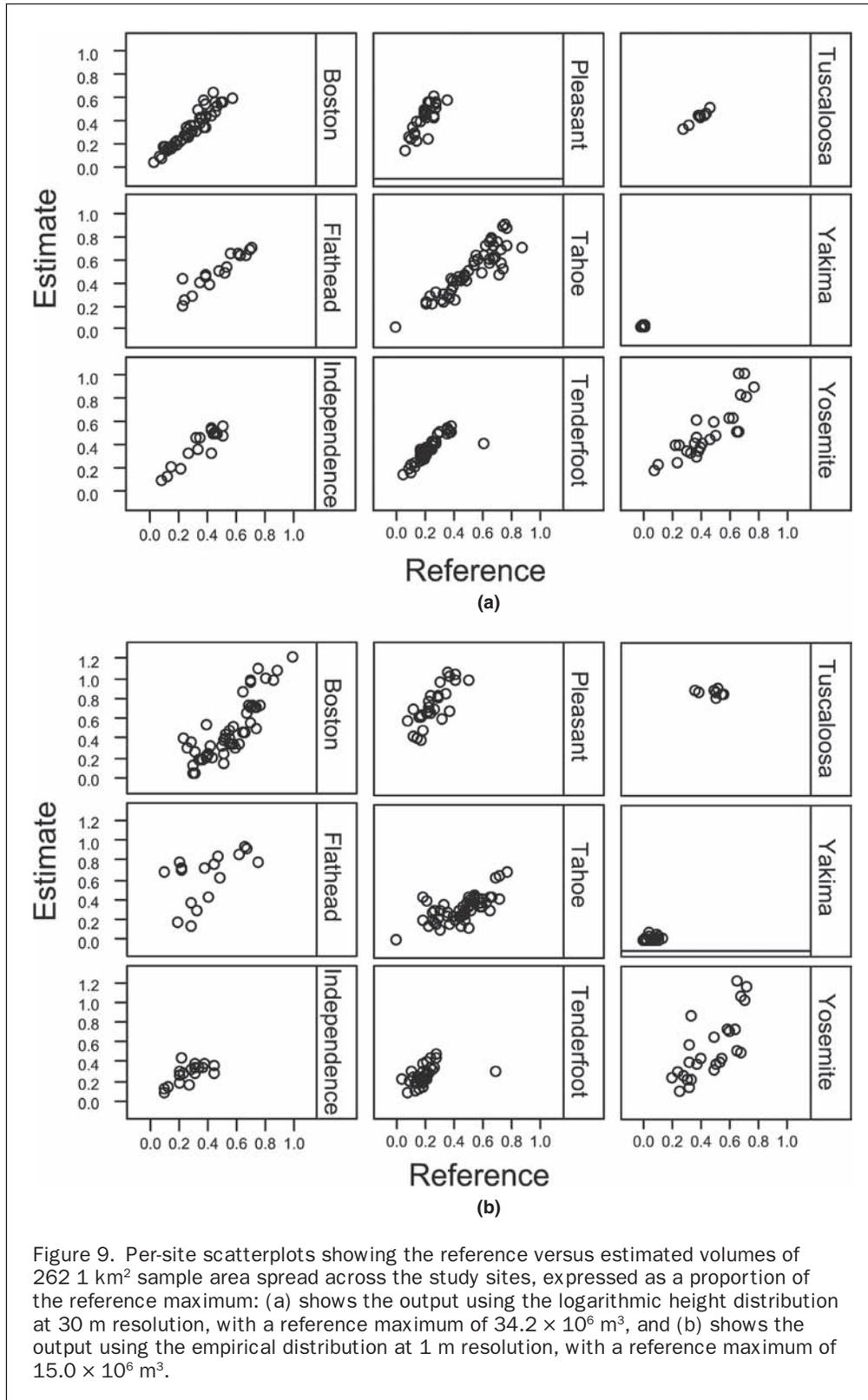
	<i>Random 1 km<sup>2</sup> locations</i>			
	<i>1 m resolution</i>		<i>30 m resolution</i>	
	<i>RMSE</i>	<i>R<sup>2</sup></i>	<i>RMSE</i>	<i>R<sup>2</sup></i>
Empirical	0.66	0.38	0.45	0.61
Exponential	0.69	0.33	0.70	0.27
Linear	0.84	0.37	0.36	0.67
Logarithmic	0.95	0.46	0.33	0.77
	<i>n = 262; mean = 5.55 × 10<sup>6</sup> m<sup>3</sup></i>		<i>n = 262; mean = 12.91 × 10<sup>6</sup> m<sup>3</sup></i>	

accurate downscaling at 30 m, and the empirical distribution results in the most accurate downscaling at 1 m resolution.

Figures 9a and 9b show the reference versus estimated canopy volumes for the 1 km<sup>2</sup> sites, at 30 m and 1 m resolution, respectively. The 30 m output uses the logarithmic distribution, and the 1 m output uses the empirical distribution,

selected based on the RMSE shown in Table 2. The 30 m resolution output is more accurate than the 1 m output, which shows a greater divergence between reference and estimated volumes among the study sites.

These results suggest that the choice of height distribution does affect the results, and identifies the distributions



that produce more accurate results for the given study areas. However, the reliability of these results are limited by the small number of available sample sites, as well as the accuracy of the underlying height data and the underlying weak correlations between height and canopy density. Changes in 1 m heights were unnaturally demarcated by the 30 m pixel boundaries in many cases, suggesting that smoothing when generating the 1 m CHM may also improve results. The limited number of sample tree models in the table lookup data, and their unrepresentative crown shapes, may also contribute to the inaccuracy of the 1 m results.

To assess the effects of temporal differences on the results we compared the random sample points to the Landsat NDVI change products from 1990 to 2005 and 2000 to 2005 (Green, 2011), and removed any sample within 1 km of an area with a negative or positive NDVI change. Using the logarithmic height distribution, the squared correlation among the 126 remaining points at 30 m was similar, with an  $R^2$  of 0.77. The RMSE reflects a slightly higher accuracy at 28 percent of the reference average ( $12.91 \times 10^6 \text{ m}^3$ ) for the  $1 \text{ km}^2$  areas around the selected points.

We also ran the 30 m process using the NLCD-2001 land-cover dataset. NLCD-2001 land-cover data should be more similar to the 2001 canopy density product than NLCD-2006, but is likely to differ more from the 2005 GLAS-based data. The difference in the normalized canopy volume RMSE was negligible (unchanged to two decimal places), although the  $R^2$  increased slightly to 0.79, using the same sample locations as the other tests. Figure 10 shows the overall relative volume by site, normalized to the maximum, as estimated using land-cover data from NLCD-2001, from NLCD-2006, and as calculated using reference data, indicating that the differences between the two are small.

Computationally, the memory footprint for the proof-of-concept application within the coterminous US is  $\sim 1 \text{ GB}$ . A  $100 \text{ km}^2$  area at 30 m requires  $\sim 8$  seconds of processing time, and a  $1 \text{ km}^2$  area at 1 m requires  $\sim 6$  seconds, with most of the time spent querying the NLCD web services in both cases. This indicates that the overall approach is

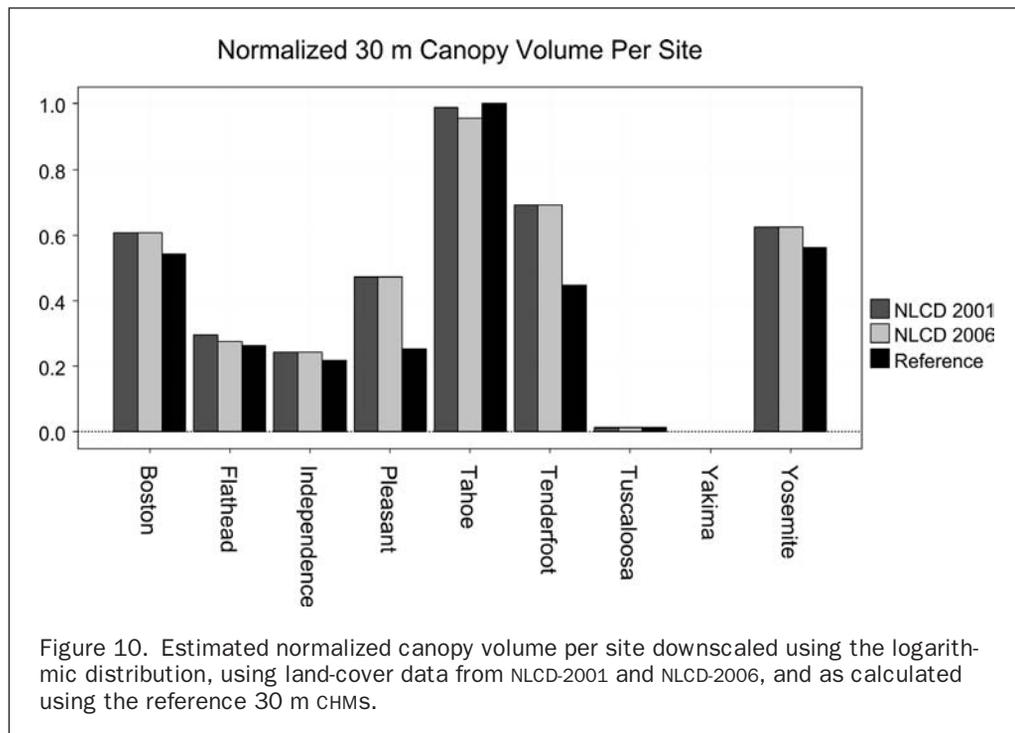
computationally feasible, but that further optimization is required before it can be used in web-based maps, where users expect faster response times.

## Conclusions

In this paper we demonstrated an approach to estimating canopy height at 30 m and 1 m resolution using on-demand fusion of existing datasets. Our initial accuracy assessment shows an RMSE of 33 percent of the mean reference volume and an  $R^2$  of 0.77 at 30 m resolution, an RMSE of 66 percent of the mean reference volume, and an  $R^2$  of 0.38 at 1 m resolution, using the height distributions that resulted in the most accurate results, compared to data derived from airborne lidar. The process offers a more complete, but approximate, characterization of horizontal and vertical variability of canopy height within each  $1 \text{ km}^2$  area than is available using scalar statistics like percent tree cover or maximum height derived from coarse-grained height data. Until national high-resolution canopy height models are available based on continuous lidar coverage with a sufficiently high point density, such an approximation may be useful for applications that need to bridge the gap in scale between coarse-grained height data and tree height as observed in the local landscape.

## Acknowledgments

The authors would like to thank the anonymous reviewers for their many insights and suggestions, which greatly improved the manuscript. This material is based on data services provided by the OpenTopography Facility with support from the National Science Foundation under NSF Award Numbers 0930731 and 0930643. Yosemite National Park, California, Tuscaloosa, Alabama, Pleasant, Maine, Flathead Lake, Montana, Independence Lake, California, and Tenderfoot Creek, Montana lidar data acquisition and processing was completed by the National Center for Airborne Laser Mapping (NCALM: <http://www.ncalm.org>). NCALM funding was



provided by NSF's Division of Earth Sciences, Instrumentation and Facilities Program. EAR-1043051. This material is based on services provided to the Plate Boundary Observatory by NCALM (<http://www.ncalm.org>). PBO is operated by UNAVCO for EarthScope (<http://www.earthscope.org>) and supported by the National Science Foundation (No. EAR-0350028 and EAR-0732947), from the EarthScope Yakima Lidar project, Washington. Lake Tahoe Basin Lidar data was provided by Tahoe Regional Planning Agency. Boston-area airborne lidar data was made available by the Office of Geographic Information (MassGIS), Commonwealth of Massachusetts, Information Technology Division. Data was available from the U.S. Geological Survey; the USGS home page is <http://www.usgs.gov>.

## References

- Beaujardiere, J. (editor), 2006. OpenGIS Web Map Service (WMS) Implementation Specification OGC 06-942, URL: <http://www.opengeospatial.org/standards/wms>, Open Geospatial Consortium, Wayland, Massachusetts (last date accessed: 14 November 2012).
- Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham, 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States, *Photogrammetric Engineering & Remote Sensing*, 77(9): 858–864.
- Harding, D., and C. Carabajal, 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure, *Geophysical Research Letters*, 32(L21S10):1–4.
- Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J.N. VanDriel, and J. Wickham, 2007. Completion of the 2001 National Land Cover Database for the conterminous United States, *Photogrammetric Engineering & Remote Sensing*, 73(4):337–341.
- Green, K., 2011. Change Matters, *Photogrammetric Engineering & Remote Sensing*, 77(4):305–309.
- Lee, S., W. Ni-Meister, W. Yang, and Q. Chen, 2011. Physically based vertical vegetation structure retrieval from ICESat data: Validation using LVIS in White Mountain National Forest, New Hampshire, USA, *Remote Sensing of Environment*, 115(11):2776–2785.
- Lefsky, M.A., 2010. A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System, *Geophysical Research Letters*, 37(L15401):1–5.
- Schutz, B.E., H.J. Zwally, C.A. Shuman, D. Hancock, and J.P. DiMarzio, 2005. Overview of the ICESat Mission, *Geophysical Research Letters*, 32(L21S0):1–4.
- Simard, M., M. Pinto, J. Fisher, and A. Baccini, 2011. Mapping forest canopy height globally with spaceborne lidar, *Geophysical Research Letters*, 116(G04021):1–12.
- USGS EROS Web Map Services: NLCD 2001 Land Cover, 2012. URL: [http://imsref.cr.usgs.gov/WMS\\_Capabilities/USGS\\_EDC\\_LandCover\\_NLCD2001/capabilities\\_1\\_3\\_0.xml](http://imsref.cr.usgs.gov/WMS_Capabilities/USGS_EDC_LandCover_NLCD2001/capabilities_1_3_0.xml), US Geological Survey, Sioux Falls, South Dakota (last date accessed: 14 November 2012).
- USGS EROS Web Map Services: USGS EDC Land Cover NLCD2006, 2012. URL: [http://imsref.cr.usgs.gov/WMS\\_Capabilities/USGS\\_EDC\\_LandCover\\_NLCD2006/capabilities\\_1\\_3\\_0.xml](http://imsref.cr.usgs.gov/WMS_Capabilities/USGS_EDC_LandCover_NLCD2006/capabilities_1_3_0.xml), US Geological Survey, Sioux Falls, South Dakota (last date accessed: 14 November 2012).
- Yang, W., W. Ni-Meister, and S. Lee, 2011. Assessment the impacts of surface topography, off-nadir pointing and vegetation structure on vegetation lidar waveforms using an extended geometric optical and radiative transfer model, *Remote Sensing of Environment*, 115(11):2810–2822.
- Zwally, H.J., B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, J. Bufton, J. Dezio, D. Hancock, D. Harding, T. Herring, B. Minster, K. Quinn, S. Palm, J. Spinhirne, R. Thomas, 2002. ICESat's laser measurements of polar ice, atmosphere, ocean, and land, *Journal of Geodynamics*, 34(11):405–445.