Regional distribution of forest height and biomass from multisensor data fusion

Yifan Yu,[1] Sassan Saatchi,[2] Linda S. Heath,[3] Elizabeth LaPoint,[3] Ranga Myneni,[4] and Yuri Knyazikhin[4]

Received 1 March 2009; revised 10 March 2010; accepted 12 April 2010; published 21 August 2010.

[1] Elevation data acquired from radar interferometry at C-band from SRTM are used in data fusion techniques to estimate regional scale forest height and aboveground live biomass (AGLB) over the state of Maine. Two fusion techniques have been developed to perform post-processing and parameter estimations from four data sets: 1 arc sec National Elevation Data (NED), SRTM derived elevation (30 m), Landsat Enhanced Thematic Mapper (ETM) bands (30 m), derived vegetation index (VI) and NLCD2001 land cover map. The first fusion algorithm corrects for missing or erroneous NED data using an iterative interpolation approach and produces distribution of scattering phase centers from SRTM-NED in three dominant forest types of evergreen conifers, deciduous, and mixed stands. The second fusion technique integrates the USDA Forest Service, Forest Inventory and Analysis (FIA) ground-based plot data to develop an algorithm to transform the scattering phase centers into mean forest height and aboveground biomass. Height estimates over evergreen (R^2 = 0.86, $P < 0.001$; RMSE = 1.1 m) and mixed forests (R^2 = 0.93, $P < 0.001$, RMSE = 0.8 m) produced the best results. Estimates over deciduous forests were less accurate because of the winter acquisition of SRTM data and loss of scattering phase center from tree-surface interaction. We used two methods to estimate AGLB; algorithms based on direct estimation from the scattering phase center produced higher precision (R^2 = 0.79, RMSE = 25 Mg/ha) than those estimated from forest height (R^2 = 0.25, RMSE = 66 Mg/ha). We discuss sources of uncertainty and implications of the results in the context of mapping regional and continental scale forest biomass distribution.

Citation: Yu, Y., S. Saatchi, L. S. Heath, E. LaPoint, R. Myneni, and Y. Knyazikhin (2010), Regional distribution of forest height and biomass from multisensor data fusion, *J. Geophys. Res.*, *115*, G00E12, doi:10.1029/2009JG000995.

1. Introduction

[2] It is generally understood that the terrestrial carbon sink required to balance the global carbon budget could be located in the forested land of the Northern Hemisphere [*Keeling et al.*, 1996; *Houghton et al.*, 1999; *Goodale et al.*, 2002]. The uncertainties of the magnitude and distribution of the sink, however, vary geographically and depend on land use history and management, disturbance frequency, and ecological and climate factors [*Schimel et al.*, 2001]. Currently, several national and international scientific research programs are designed to study these uncertainties such as the North American Carbon Program (NACP), Carbo-Europe, etc. The main goal of the NACP is to quantify and assess the role of various components of the carbon sink over North America. Among these components, the woody biomass and amount of carbon stock in forests as well as its spatial distribution play a major role in achieving this goal [*Wofsy and Harriss*, 2002]. Moreover, credible information on vegetation biomass has significant economic value in existing markets for carbon trading and serves as a major indicator to evaluate the ecosystem services [*Foley et al.*, 2005].

[3] However, there is considerable uncertainty in our knowledge of how much carbon is contained in the world's forests. Measurement of the forest carbon is based on the biomass inventories that provide reliable estimates of timber volume, growth, and mortality. These measurements are based on statistical sampling with samples interspersed across forest lands of the United States such that the coverage is not wall-to-wall; costly to perform; and difficult to repeat frequently on national and continental scales. Moreover, forest inventory data cannot be readily integrated in spatial models to study the cycling and dynamics of carbon

[1]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, California, USA.
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.
[3]Forest Inventory and Analysis, Forest Service, U.S. Department of Agriculture, Durham, New Hampshire, USA.
[4]Department of Geography and Environment, Boston University, Boston, Massachusetts, USA.

Copyright 2010 by the American Geophysical Union.
0148-0227/10/2009JG000995

in vegetated ecosystems [*Hurtt et al.*, 2004]. But these ground-based data are crucial for model development and validation while additional information or techniques must be used to apply these data across the landscape.

[4] Measuring and monitoring forest structure and carbon stock and changes from space can help circumvent the limitations associated with inventory data. Active remote sensing techniques using lidar and radar sensors have been used extensively to measure vegetation structure and aboveground biomass. In lidar sensing, backscatter returns from laser altimetry are used to estimate vegetation height and other structural variables that can be used to estimate biomass [*Dubayah and Drake*, 2000; *Drake et al.*, 2002; *Lefsky et al.*, 2002]. Radar backscatter measurements from low frequency synthetic aperture radar (SAR) data at L-band (1.25 GHz) and P-band (0.45 GHz) and at different polarizations are strongly and positively correlated with forest volume and aboveground biomass [*Dobson et al.*, 1992; *Le Toan and Floury*, 1998; *Saatchi et al.*, 2000]. In recent years, a new technique, using interferometric SAR (InSAR) measurements, has been developed to estimate forest height and structure [*Kobayashi et al.*, 2000; *Papathanassiou et al.*, 2001]. The InSAR is a technique that measures the cross-correlation of backscatter measurements from two different locations (two antennas) in space. By knowing the vector between the two antennas (baseline) and their absolute distance to a surface point, the location of the surface point in elevation is determined from the phase difference of InSAR cross-correlation. Over vegetated surfaces, the elevation of the surface point, known as the scattering phase center, is somewhere within the vegetation canopy depending on the vegetation type and its structure. In theory, by knowing the surface elevation, the height of scattering phase center can be determined from one InSAR measurement and consequently related to the vegetation height or biomass [*Hagberg et al.*, 1995; *Askne et al.*, 1997; *Kobayashi et al.*, 2000]. By adding other measurements such as polarimetry or geometry in space (multibaseline), the height of vegetation can be determined unambiguously without the surface elevation [*Papathanassiou et al.*, 2001; *Reigber et al.*, 2000].

[5] In February 2000, the Shuttle Radar Topography Mission (SRTM) provided a near global coverage of land surface elevation from InSAR measurements at C-band (5.3 GHz). The surface elevation derived from SRTM data is biased over the vegetated surfaces proportional with the vegetation height. Several studies have shown that by knowing the surface elevation, SRTM data can be used to estimate the height of the scattering phase center and the vegetation height [*Kellndorfer et al.*, 2004; *Simard et al.*, 2006]. For example, by using the National Elevation Dataset (NED), which is available over the United States, as a reference, it is possible to establish a linear regression model between vegetation canopy height and the difference of SRTM-NED [*Kellndorfer et al.*, 2004]. However, this relationship is highly dependent on forest characteristics such as species, density, and structure and cannot be readily applied on regional and continental scales. In addition, the spatially inconsistent quality of NED, errors in geo-referencing of SRTM and NED, and the effect of SAR scattering mechanism in the vegetation layer on the location of scattering phase center are some of the difficulties in extending the applicability of this approach over large regions. A comprehensive assessment of the quality of the SRTM data and its application for forest structure retrieval has been performed by *Walker et al.* [2007a, 2007b]. Consequently, a series of fusion approaches with field inventory data and ancillary spatial information such as NED and land cover maps have been proposed to extend the application of SRTM data to continental scale mapping of forest canopy height [*Walker et al.*, 2007a, 2007b].

[6] In this study, we present alternative data fusion methodologies to investigate and mitigate problems associated with the SRTM and NED data that would be relevant for regional and continental scale studies. We developed an interpolation algorithm to reduce the amount of area over which SRTM and/or NED data is missing or the quality is below an acceptable level. Optical remote sensing data from Landsat and the National Land Cover Database (NLCD) were used to aid the generation of a final product of SRTM-NED. By using location, biomass, and structure from the USDA Forest Inventory and Analysis plots over the state of Maine, we developed separate algorithms for evergreen, deciduous, and mixed forests, to estimate forest height and biomass from fusion of SRTM-NED and Landsat data. The methodologies used for the state of Maine can also be applied to other regions with slight modifications. These methodologies complement the existing methods and extend the applications to regional mapping of forest carbon in continental US. In addition, the results from this study will provide an assessment of application of SAR interferometry data for forest height and biomass estimation from future planned sensors on-board NASA's DESDynl (Deformation, Ecosystem Structure, and Dynamics of Ice) and European Space Agency's BIOMASS missions.

2. Area of Study

[7] The state of Maine is located between 42° and 47°N latitude, and 67° and 71°W longitude in the northeastern United States. Topography of the region ranges from reasonably flat coastal areas to the mountainous regions further inland. Elevation runs between sea level and 1000 m. Based on the National Land-Cover Database [*Homer et al.*, 2004], approximately 90% of the state's land surface is forested, which includes evergreen, deciduous, mixed, shrub, and woody wetlands. The forested areas are approximately divided in half between softwoods and hardwoods [*Smith et al.*, 2004]. The softwoods mainly consist of firs and pines (mostly evergreen trees) while the hardwoods are mostly made up of maple, birch, and oak (mostly deciduous) [*Little*, 1979].

[8] For this study, the forested areas were separated into deciduous, evergreen, and mixed types. This is due to the fact that the structures of these two types of trees are very different. More importantly, the SRTM was flown in February 2000 during the northern hemisphere winter when all the deciduous forests in Maine would have lost their leaves. Figure 1 shows the spatial distribution of the different types of forests. With this classification at a 90 m pixel scale, the deciduous, evergreen, and mixed types take up approximately 18%, 22%, and 35% respectively, of the total land area. This area is ideal for the study because of the availability of in-situ field measurements as well as the different

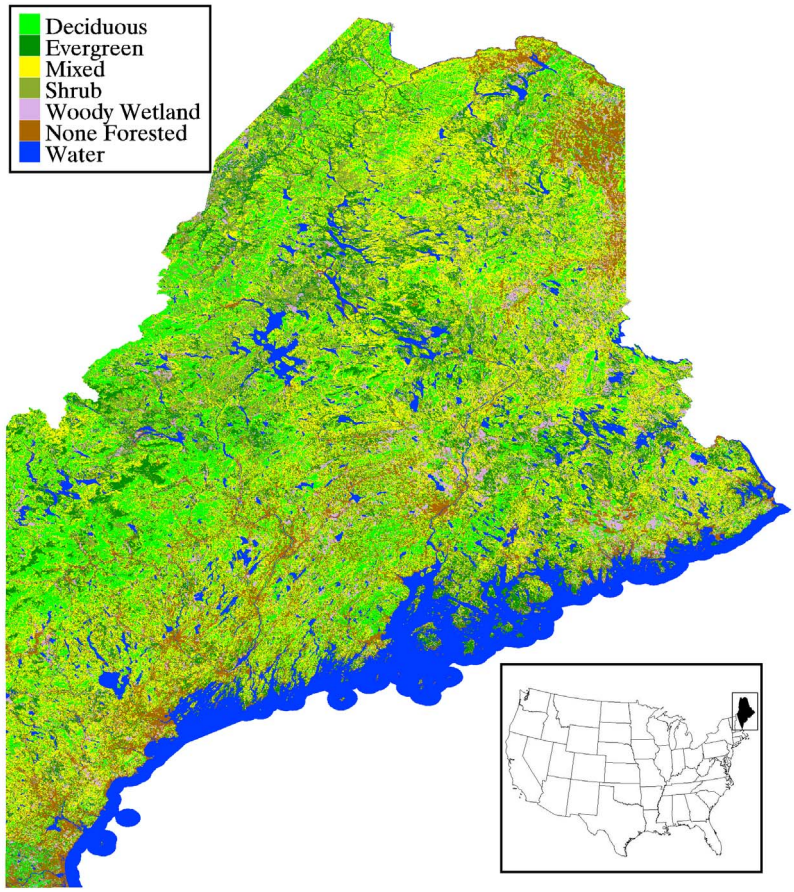

Figure 1. Forest vegetation cover derived from Conterminous United States Land Cover by National Center for Earth Resources Observation and Science, U.S. Geological Survey.

combinations of topography and vegetation types to test the methods.

3. Data

3.1. Forest Inventory Data

[9] For direct field measurements, data were provided by the USDA Forest Service, Forest Inventory and Analysis program (FIA). FIA conducts inventories in three phases. Phase 1 consists of aerial photography and/or remote sensing to reduce variance through stratification. Phase 2 are the traditional FIA ground plot measurements of forest and tree information. In phase 3, a subset of phase 2 plots are selected and more information relating to forest health is collected. Measurements from phase 2 of FIA's inventory program were used for this study because of the larger number of plots available. In the phase 2 measurements, only trees with diameter greater than or equal to 5 inches were measured. In both phase 2 and phase 3, each ground plot sampled is made up of 4 smaller subplots. Each ground data point, therefore, is approximately representative of a circular plot of 37 m radius. Detailed information on the ground data used can be found on the FIA web site (http://fia.fs.fed.us/). Ground inventories for deciduous and evergreen forests from 1999 and 2000 were used to minimize variation of the forest due to time difference from SRTM flight. Inventories from 1999 to 2003 were used for mixed forests because of the small number of plots available. Overall, we chose 613 evergreen, 933 deciduous, and 29 mixed stands in the pool of 2294 plots to perform this study (Table 1). The forest stands covered a wide range of ages and biomass from young secondary to old growth forests. The plot data included a number of structural variables. We chose to examine the forest type, average and dominant heights, basal area (the sum of surface areas of trees with

Table 1. Summary of Statistics From FIA Ground Data

			Biomass			
	Percent of Land (From NLCD2001)	Number of Plots	Mean (Mg/ha)	Median (Mg/ha)	Skew	Percent < 200 Mg/ha
Deciduous	18	933	116	112	0.394	92.4
Evergreen	22	613	108	105	0.412	94.5
Mixed	35	29	131	113	0.184	86.2

diameters at breast height and in the unit of m^2/ha), and aboveground biomass in our analysis.

3.2. Remote Sensing Data

[10] Several sets of remote sensing data that included all of Maine were used to develop relationships with the ground data. With the exception of the National Elevation Dataset (NED), similar time period of coverage were used for all the remote sensing data. We selected the remote sensing data and the ground data to be as close to the year 2000 as possible to minimize errors due to variations in the forest structure over time from the acquisition of SRTM data.

3.2.1. Shuttle Radar Topography Mission

[11] The Shuttle Radar Topography Mission was conducted in February 11–22 of 2000. It uses interferometric synthetic aperture radar (InSAR) techniques and covers global landmasses between ±60° latitude. The main instrument on the shuttle is a C-band (5.6 cm wavelength) ScanSAR interferometer. The C-band radar has 4 sub-swaths using horizontal (H) and vertical (V) polarization (2 H sub-swaths and 2 V sub-swaths) with 2 ScanSAR beams illuminating 2 sub-swaths at a time. Together, the swaths cover about 225 km. There is also a X-band (3 cm wavelength) interferometer on-board, but since it did not have a ScanSAR mode, it had a much smaller swath and did not provide continuous coverage (see http://www.jpl.nasa.gov/srtm). The SRTM data is available to the public in 3 arc-second resolution (≈90 m) outside of the United States, and full resolution of 1 arc-second (≈30 m) in the U.S. The consistency and near-global coverage of SRTM make it a great choice for studying terrestrial biomass distribution. The data was re-sampled into 3 arc-second resolution for this study.

[12] The National Geospatial-Intelligence Agency (NGA) and NASA conducted an extensive ground campaign to collect ground truth and produced global validation of the SRTM data. For North America, SRTM was shown to have absolute geo-location error of 12.6 m, absolute height error of 9.0 m, and relative height error of 7.0 m, with the biggest component of the error coming from random errors caused by instrument thermal noise and residual geometric decorrelation effects [*Rodriguez et al.*, 2006]. It was shown that the random error is correlated with topography and radar brightness: typically <5 m, but at higher latitudes or for flat regions with good coverage, <2 m. This means that a large portion of the area of interest will have much better absolute SRTM height error than 9.0 m. But this also means that it is important to produce a map of confidence levels to accompany any large scale forest canopy and biomass maps generated using SRTM.

[13] It is well known that SRTM radar signal does not always penetrate all the way through the forested canopy to the ground. In a dense canopy forest, the scattering phase center height generally falls somewhere between the canopy top and the ground surface. *Kellndorfer et al.* [2004], *Simard et al.* [2006], and *Heo et al.* [2006] have all shown that it is possible to retrieve forest height information from SRTM. *Sarabandi and Lin* [2000] showed that the equivalent scattering phase center height for a vegetated forest stand depends on forest characteristics such as soil moisture, tree density, and tree types as well as the InSAR parameters such as frequency, polarization, and incidence angle. These factors control the relative significance of scattering mechanism within the forest and consequently impact the location of the scattering phase center. For example, if the double-bounce or volume-surface interaction scattering dominates, the scattering phase center shifts towards the ground. This may happen in a tall forest with a sparse canopy or when forest floor is smooth and very wet [*Saatchi and McDonald*, 1997; *Sarabandi and Lin*, 2000].

[14] In the case of the SRTM product used for this study, the InSAR parameters for individual pixels are not known since most pixels at these latitudes are averaged from multiple passes (part of the post processing to reduce the error.) Therefore, we can only assume that the SRTM data gives a scattering phase center height with a certain amount of vertical error as well as horizontal geo-location error, and focus on the physical characteristics of the forest stands which also affect the scattering phase center.

[15] In a forest with closed canopy and dense leaf coverage, the scattering phase center height will fall closer to the top of the canopy. As the forest becomes less dense, or in the case of deciduous trees which have lost all their leaves during the time of SRTM flight, the scattering phase center height becomes lower and forest density, tree types, and soil moisture may play a larger role in impacting the location of scattering phase center [*Saatchi and McDonald*, 1997]. The biomass value of a forest is correlated with the density and height of the forest, which are two of the physical parameters that affect the scattering phase center height. It is reasonable that there is also a relation between scattering phase center height and forest biomass. In the case of deciduous forests, density is expected to play a bigger role in the amount of penetration, due to the loss of leaves to allow more radar penetration. These variations will need to be taken into consideration when retrieving forest parameters from SRTM for different forest types.

3.2.2. National Elevation Dataset

[16] The National Elevation Dataset (NED) is a seamless digital elevation map (DEM) produced by the U.S. Geological Survey (USGS). It is assembled from the best available source data over the region of interest. The source data consist of 10-m DEM, 30-m Level 2 DEM, 30-m Level 1 DEM, 2-arc-second DEM, and 3-arc-second DEM from the USGS National Mapping Program's Sales Database. The different source data are reprojected into the same projection and then processed and matched to remove artifacts and fill in missing data at the boundaries through interpolation. The dataset is also updated as better DEMs become available over various regions [*Gesch et al.*, 2002]. However, because of the piecewise production of the NED data, there are areas with missing data, and the quality of the DEM may vary from region to region.

[17] In the case of the state of Maine, we found several small patches with missing data or various imaging artifacts (e.g. smearing). These are mainly due to lack of high precision data or a mismatch in geo-referencing between adjacent patches of DEM. Sometimes, these artifacts are hard to locate when observing only the NED image by itself. However, when taking the difference of SRTM-NED, the artifact becomes pronounced. Figure 2 shows the SRTM-NED image for a small area of Maine as well as a histogram of the pixel values. The histogram shows most points fall into the expected range of values, but there are also small

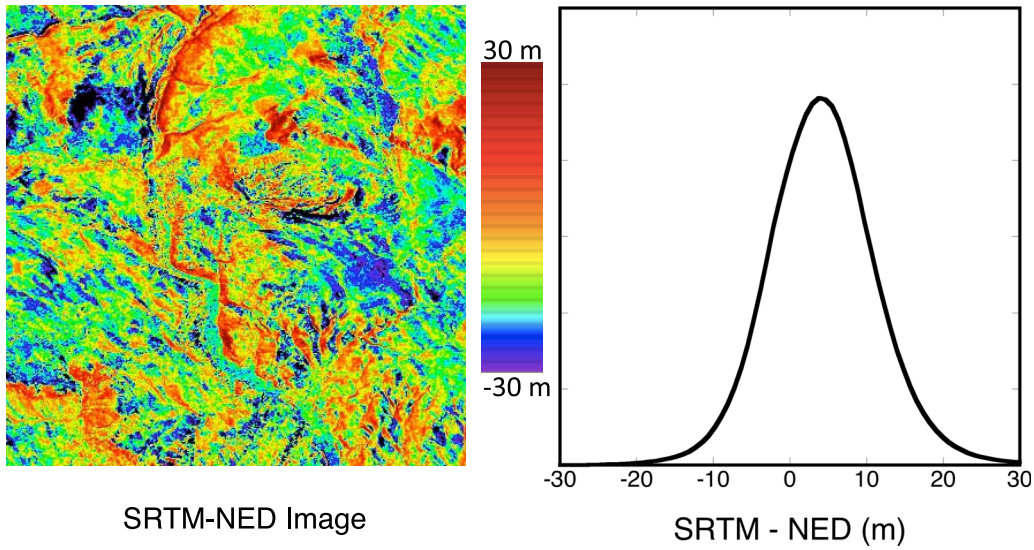

Figure 2. (left) Image of SRTM-NED over a small area of Maine (400 × 400 pixels) and (right) histogram of pixel values in the image. Image shown is prior to interpolation with mean height of 8.91 m and range between −31 to 58 m.

numbers of unrealistically large values as well as unphysical negative values.

3.2.3. Landsat

[18] The Landsat data for this study were from the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) used in GeoCover data set (Landsat GeoCover. 2003). ETM+ provides data from eight spectral bands with 30 m resolution for the visible and near-infrared bands, 15 m for the panchromatic band, and 60 m for the thermal infrared band. The spectral bands 2, 4, and 7 (two reflected IR and one green) were obtained from USGS in 30 m resolution. The wavelength range of bands 2, 4, and 7 are .525–.605 μm, .75–.90 μm, and 2.09–2.35 μm respectively. The ETM+ scenes were mosaicked and cut into the desired area and re-sampled into 90 m resolution to match the SRTM and NED images. More information about the Landsat program can be obtained at the USGS Landsat Program web site (http://landsat.usgs.gov/).

[19] The Landsat image was also used to generate images of various vegetation indexes. The spectral bands 4 and 2 from ETM+ were used to develop the Green Normalized Difference Vegetation Index (GNDVI) and Specific Leaf Area Vegetation Index (SLAVI):

$$GNDVI = \frac{NIR - G}{NIR + G}, \qquad SLAVI = \frac{NIR}{MIR + R} \tag{1}$$

Both indexes are different from the commonly used Normalized Difference Vegetation Index (NDVI). The ETM+ green (band 2) and mid-infrared (band 7) have a strong sensitivity to forest height and the timber volume [*Butera*, 1986; *Horler and Ahern*, 1986; *Puhr and Donoghue*, 2000]. In general, GNDVI is sensitive to canopy cover and disturbance severity and is used in monitoring crops, whereas SLAVI is primarily sensitive to volume or height of vegetation [*Puhr and Donoghue*, 2000].

3.2.4. Land Cover

[20] The Multi-Resolution Land Characteristics Consortium (MRLC) produced the National Land Cover Database 2001 (NLCD 2001) land cover layer. One of the products of this database is a land cover classification map derived from Landsat imageries and ancillary data. The NLCD 2001 was produced by using a decision tree classification [*Homer et al.*, 2004].

[21] The land cover map was downloaded at 30 m resolution in an equal area projection. It is then reprojected into the same geographic latitude/longitude projection as the other images but with 1/3 the pixel width and height as the other images, which is roughly the same pixel size as the original 30 m image. The image is then re-sampled into the same size pixels as the others and then registered with the other images via linear shift only. The re-sampling algorithm divides the higher resolution image into windows of 3 by 3 pixels with no overlap. A simple majority is first used for the pixels in each window to determine if the result pixel will be land or water. For land pixels, the NLCD 2001 classification of over 20 classes is simplified down to 5 classes for this study, consisting of: evergreen forest, deciduous forest, mixed forest, woody wetlands, and non-forest. Each window that will be classified as land is classified as one of the 5 classes mentioned above if and only if that class accounts for over 50% of the land pixels in the window. If none of the 5 classes account for over 50% of the land pixels, then the window is classified as mixed forest. More detailed information about the NLCD 2001 can be found at the MRLC Consortium web site (http://www.mrlc.gov/mrlc2k_nlcd.asp).

4. Methodology

[22] Several studies have shown that linear relationships can be established between the difference of SRTM-NED and forest canopy height [*Kellndorfer et al.*, 2004; *Heo et al.*, 2006; *Walker et al.*, 2007a, 2007b]. This works especially well in areas where the forest is more homogeneous

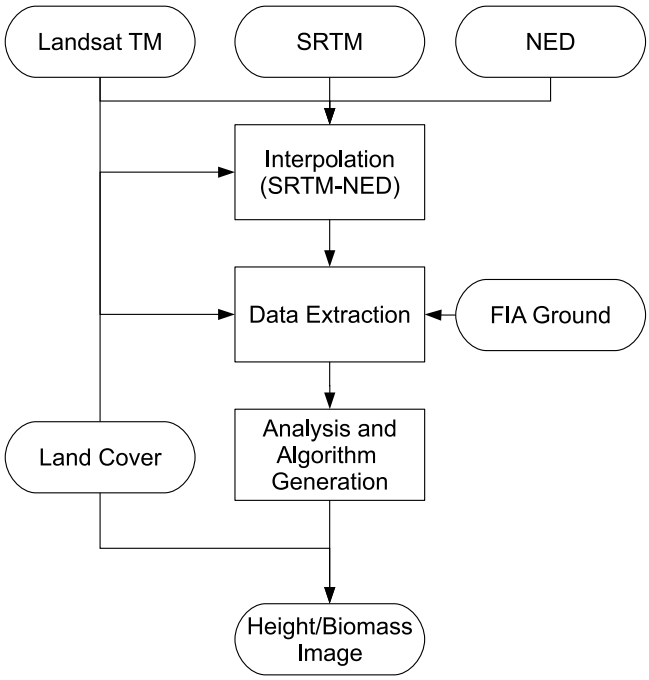

Figure 3. Flowchart of the overall methodology of the production of height and biomass maps using remote sensing and ground data.

and with closed canopy, which reduces the errors introduced due to varying forest characteristics. However, when deriving canopy height on a larger regional scale, variations in forest characteristics are bound to increase significantly as a bigger range of stand age and tree species comes into play.

[23] We introduced several methods to minimize errors in remote sensing data and ground observations. Ancillary data were also brought in to help determine forest characteristics. We also looked at two different ways of obtaining above ground live biomass (AGLB) information from the remote sensing data. An overall procedure is presented for the derivation of height and biomass maps on a large regional scale, in this case, the state of Maine. The order of the methods used and the types of remote sensing data used for the different steps are shown in Figure 3. Details on what is performed in each step are presented below.

4.1. Correction of SRTM-NED

[24] In the ideal case, the value of the difference between SRTM and NED over a forest stand would be somewhere between zero and the actual height of the forest stand with the location of the scattering phase center height dependent on forest characteristics. However, due to the errors inherent in both data sets, there will be erroneous points that fall outside of the physically reasonable range. We created an algorithm to interpolate over as many of these erroneous points as possible. Figure 4 shows the flow of our interpolation algorithm.

[25] The interpolation program first finds all the pixels it deems erroneous by marking any pixels where SRTM-NED either falls below 0 or goes above a prescribed value based on distribution of SRTM-NED values over field plot locations as well as the distribution of mean canopy height in the FIA field measurements (Figure 6). For the set of pixels identified as erroneous, NLCD 2001 land cover map is used to determine if the pixel is forested or not. If it is not forested, the SRTM-NED value is set to zero. If it is forested, we look at a window of 3 by 3 pixels centered at the "bad pixel". Within this window of nine pixels, we look at the other forested pixels (with reasonable values of SRTM-NED); and if there are enough (threshold of 4 used in this study) surrounding "good pixels", the value of the bad pixel is interpolated from the mean of the surrounding forested pixels. This assumes that the forest in question is similar to the surrounding forests, and by filling the gap in the data, we preserve the mean value of forest height or density in the 3 × 3 box.

[26] If the pixel is marked erroneous because it has a SRTM-NED value over the prescribed threshold, the "bad pixel" is interpolated by taking the mean of the surrounding "good pixels" and adding one standard deviation to it. This was done for two reasons: (1) to reflect the increased possibility of a higher-than-average forest because of the larger value of SRTM-NED; (2) to approximately preserve the variance of the forest within the 3 × 3 box. A larger variability in the local region of forest would give an interpolated forest height that is farther from the mean, while no variability in the surrounding forest would give an interpolated forest height. The standard deviation is calculated from all forested "good pixels" within the interpolation window (as indicated by NLCD 2001 land cover map). This method can be repeated with as many iterations as desired. The pixels that were interpolated in the current iteration will not be used again in the current iteration. The interpolated pixels will only be included for interpolation in the next iteration.

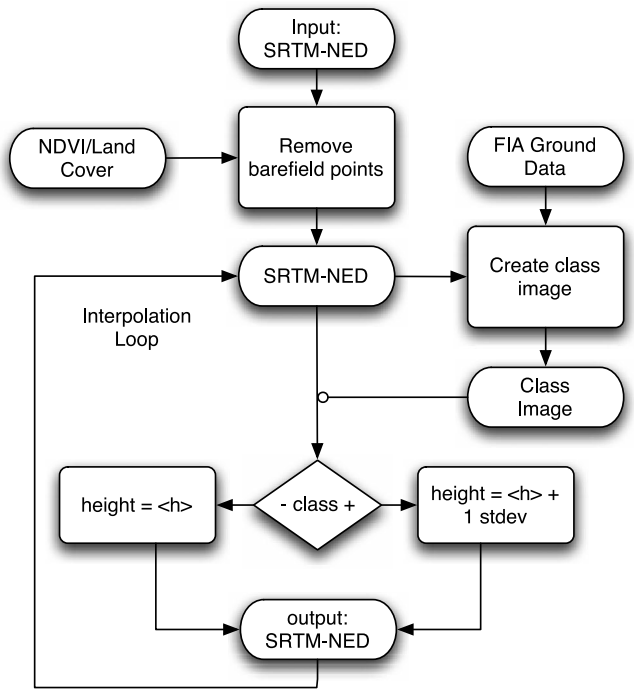

Figure 4. Flow of the algorithm for interpolating bad points in SRTM-NED. Multiple iterations can be run to interpolate more points.

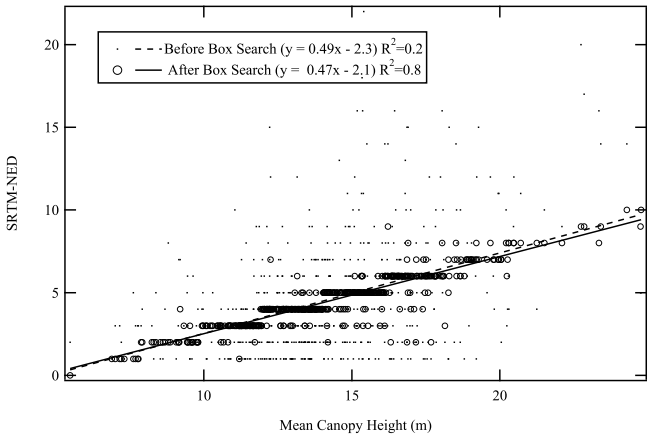

Figure 5. Fitting between SRTM-NED and mean canopy height for evergreen forests before and after the box search algorithm.

The number of pixels interpolated in each iteration drops exponentially with each iteration.

4.2. Integration of FIA Data and Remote Sensing Data

[27] Due to the inherent errors in the geo-location of the data, especially of the field measurements, the pixel with the same location coordinates as the ground plot is not necessarily the actual pixel where the field measurements were taken. One example of this is when the ground plot is an isolated forest stand that is surrounded by crop fields. Using the latitude and longitude coordinates given for the ground measurement, the plot appears to fall in a crop field on the remote sensing image instead. Another obvious example of geo-location error is when a ground forest plot appears to fall into a body of water which is next to the forested areas. To help correct this random error in geo-location, we utilized a box search program written in IDL, similar to an approach used by *Anderson et al.* [2006].

[28] First, SRTM-NED values are extracted from the locations of the ground data. A simple linear regression fit is then applied between the SRTM-NED values and each of the field measurements of mean canopy height and above ground live biomass (AGLB). We then return to the SRTM-NED data and scan within a window of 3 pixels by 3 pixels centered around the location given by the ground data. The pixel within the window that fits best with the linear relationship found above is selected as the final remote sensing data. The new set of remote sensing data is then fitted again to the field measurements using linear regression. This gives the final corrected regressions between SRTM-NED and canopy height and AGLB. The final result from the box search method has minimal change in slope and intercept of the fitting functions. It is mostly a reduction in the spread of the data points. This can be seen in Figure 5. Tests were also performed using random regression equations. The improvements to the random regressions were negligible while the actual results showed significant improvements.

[29] Histograms in Figure 6 show the distribution of forest canopy height of the ground samples as well as SRTM-NED values of the corresponding corrected locations. All the data follow a roughly normal distribution. Averages calculated from ground plots and extracted SRTM-NED values show that SRTM on average penetrates 10.4 m, 9.9 m, and 10.0 m into the canopy for deciduous, evergreen, and mixed types respectively. A few outliers also exist in the SRTM-NED values that are most likely erroneous. These happen to be in the deciduous group. (See Figure 6. Outliers fall to the right of the vertical line in Figure 6 (bottom).) These points were removed from subsequent data analysis.

4.3. Algorithm Development

[30] The February time frame for SRTM means that deciduous trees in this region were leafless. This would greatly affect the amount of penetration of radar signal over deciduous forest stands. It is then best to separate the different types of forests when building a relationship between remote sensing data and in-situ field measurements. The FIA field data were separated into three groups consisting of evergreen, deciduous, and mixed forests. Using the SRTM-NED data and in-situ field measurements, simple linear relationships were developed between SRTM-NED and forest canopy height and forest biomass using least-squares fit. See Figures 7–9 for the functional fits of these relationships.

[31] An exponential relationship between scattering phase center height, forest canopy height, SRTM-NED, and VI

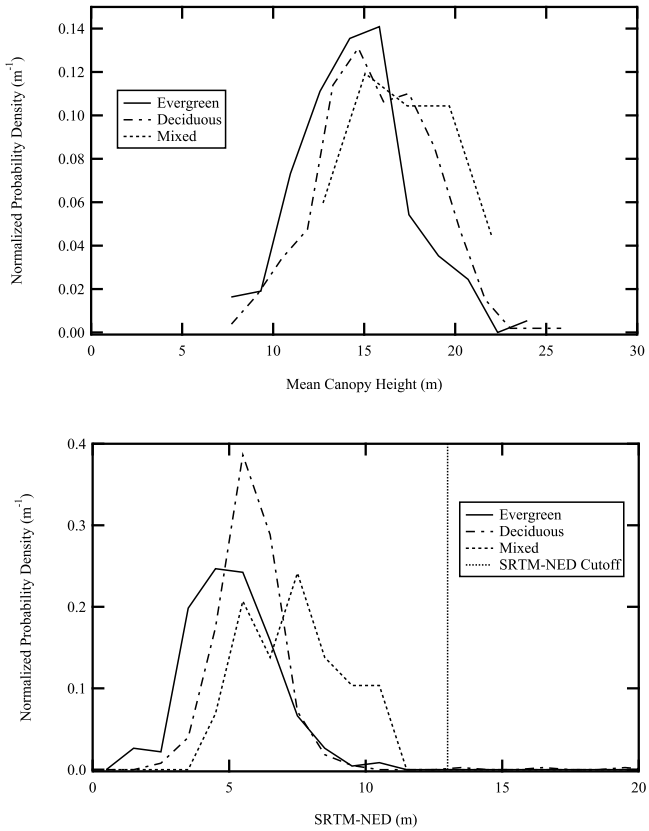

Figure 6. Histograms of (top) ground measurements of mean canopy height and (bottom) SRTM-NED values of pixels corresponding to the ground data locations. Note that the widths of the bins are not shown here (1 m). Lines are drawn between the middle of the bin widths. Vertical line in Figure 6 (bottom) represents cut off point for "too large" values of SRTM-NED.

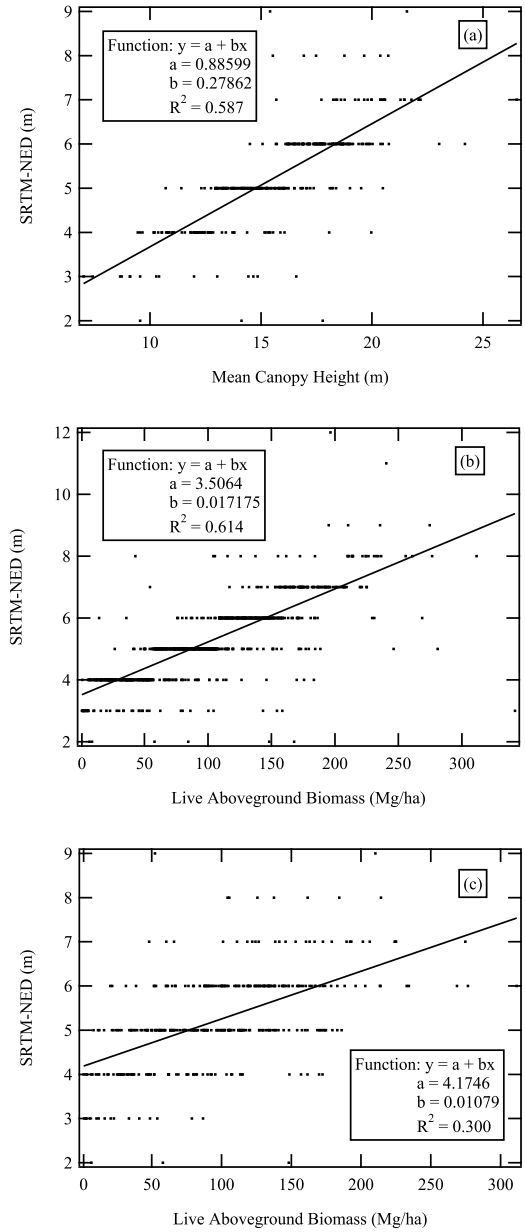

Figure 7. Linear fits for deciduous forests of SRTM-NED versus (a) mean canopy height, (b) AGLB with SRTM-NED pixels extracted using box search based on biomass values, and (c) AGLB with SRTM-NED pixels selected by box search algorithm based on height. RMSE is 1.9 m for height and 42 Mg/ha for biomass.

was also developed. The equation used for fitting to this relationship is

$$h_{sp} = h\left(1 - e^{-\alpha VI}\right) \tag{2}$$

where h_{sp} is the scattering phase center height, h is the mean canopy height, and VI is the vegetation index. The reasoning behind this is that the VI values should have some degree of correlation with forest density and thus can be used as a surrogate for forest density [*Huete et al.*, 2002]. Then we can assume that VI will affect the location of the scattering phase center height. Forest stands with denser canopy coverage (higher VI) would give a higher h_{sp} while those with less dense canopy would give a lower h_{sp} with the asymptotes approaching h and 0 respectively. For the deciduous forest stands, the data points had two distinct coefficients for regions of lower VI and regions of higher VI. This is likely due to the difference in characteristics of young and mature forests. The deciduous data set was then further divided into two separated groups using VI as a threshold. See Figure 10 for the results of equation (2).

[32] A second way of deriving a biomass map is through the use of allometric equations between height and biomass.

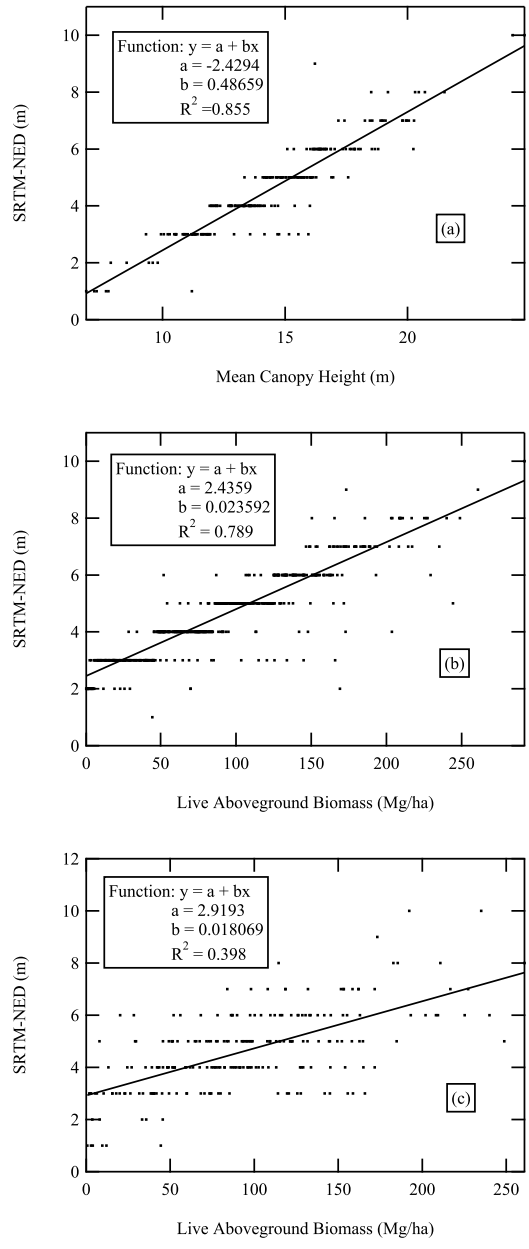

Figure 8. Linear fits for evergreen forests of SRTM-NED versus (a) mean canopy height, (b) AGLB with SRTM-NED pixels extracted using box search with biomass values, and (c) AGLB with SRTM-NED pixels selected by box search algorithm based on height. RMSE is 1.1 m for height and 25 Mg/ha for biomass.

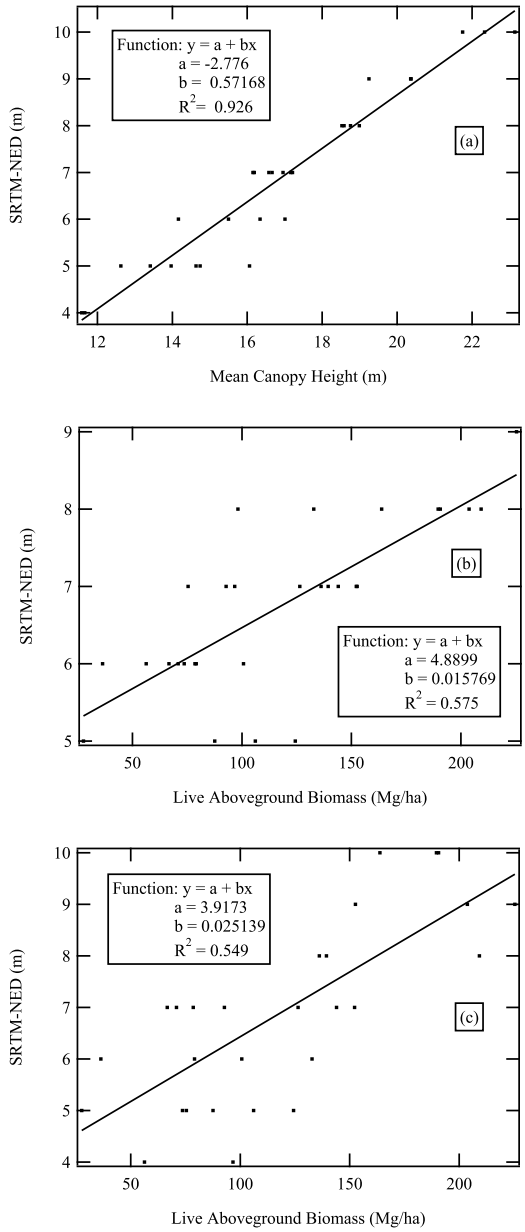

Figure 9. Linear fits for mixed forests of SRTM-NED versus (a) mean canopy height, (b) AGLB with SRTM-NED pixels extracted using box search with biomass values, and (c) AGLB with SRTM-NED pixels selected by box search algorithm based on height. RMSE is 0.8 m for height and 34 Mg/ha for biomass.

Using the FIA ground data, three separate allometric equations were developed for the three types of forest stands (see Figure 11). These allometric equations can then be applied over the height map that was generated using equation (2) to produce a biomass map. The exponential relationship of canopy height with ancillary data was chosen because it handles the extreme values of SRTM-NED better than the linear fits.

[33] Once separate relationships for canopy height and AGLB were developed for each of the three forest groups, height and biomass maps were generated with the aid of the NLCD 2001 land cover classification. First, functions for each forest type were applied over the entire region on a 90 m scale using NLCD 2001 land cover map to separate the different forest types. This 90 m image is then re-sampled to 270 m. This was done by dividing the image into 3 pixel by 3 pixel blocks. For each block where there are more forested pixels than non-forested pixels, a mean value of all the forested pixels were calculated and used as the value for the resulting 270 m pixel. If there were more non-forested pixels than forested, then the resulting 270 m pixel is set to a value of 0. This represents the 270 m pixel being classified into non-forested class based on area of coverage. A final check was performed on the resulting height and biomass maps using threshold values to cut off any unrealistic values of height and biomass. All the above operations were performed using the IDL software.

4.4. Validation

[34] Validation was performed on the linear fits of height and biomass with SRTM-NED as well as the exponential fits of canopy height with SRTM-NED. First, a bootstrapping method was used to test the fitted functions. Second, the spatial patterns of forest biomass generated in this study were compared with biomass map provided by FIA. Mean and standard deviation values for each of the forest types were also compared with those calculated from FIA field measurements (Table 5).

[35] For the bootstrapping method, each of the three types of forest plots were randomly placed into two groups. Functional relationships were developed between field measurements and remote sensing data from plots in group 1, and R^2 values were calculated. These functions from group 1 were then applied to plots in group 2, and R^2 and root mean squared error (RMSE) values are calculated. The two groups were then compared with each other.

[36] R^2 values matched well between group 1 and group 2 for both types of height relationship and the biomass relationship. The R^2 for deciduous and evergreen forest AGLB were slightly higher in group 2 than group 1. Table 2 summarizes the R^2 and RMSE values calculated from group 2 of the bootstrapping validation.

5. Results

5.1. Structure and Biomass Relations

[37] The structural variables measured in FIA plot data and the plot level biomass values derived from allometric equations provided us with the necessary data to examine the relationships between average forest height, biomass, and basal area. First, we separated the plots associated using three forest types: evergreen, deciduous, and mixed. Although the original plot data had species level information, the regrouping of the plots into three forest types was primarily due to similar forest classes in NLCD vegetation map that can eventually allow the extension of the results over the entire country. However, depending on regions, the algorithms for estimating forest biomass from remote sensing data or field estimation may be different depending on the forest structure and growth mechanism or environmental and edaphic conditions. We examined the relations between aboveground biomass, basal area, and average height (Figure 12). Basal area was the best indicator of biomass, representing almost 90% of the variations in all three forest

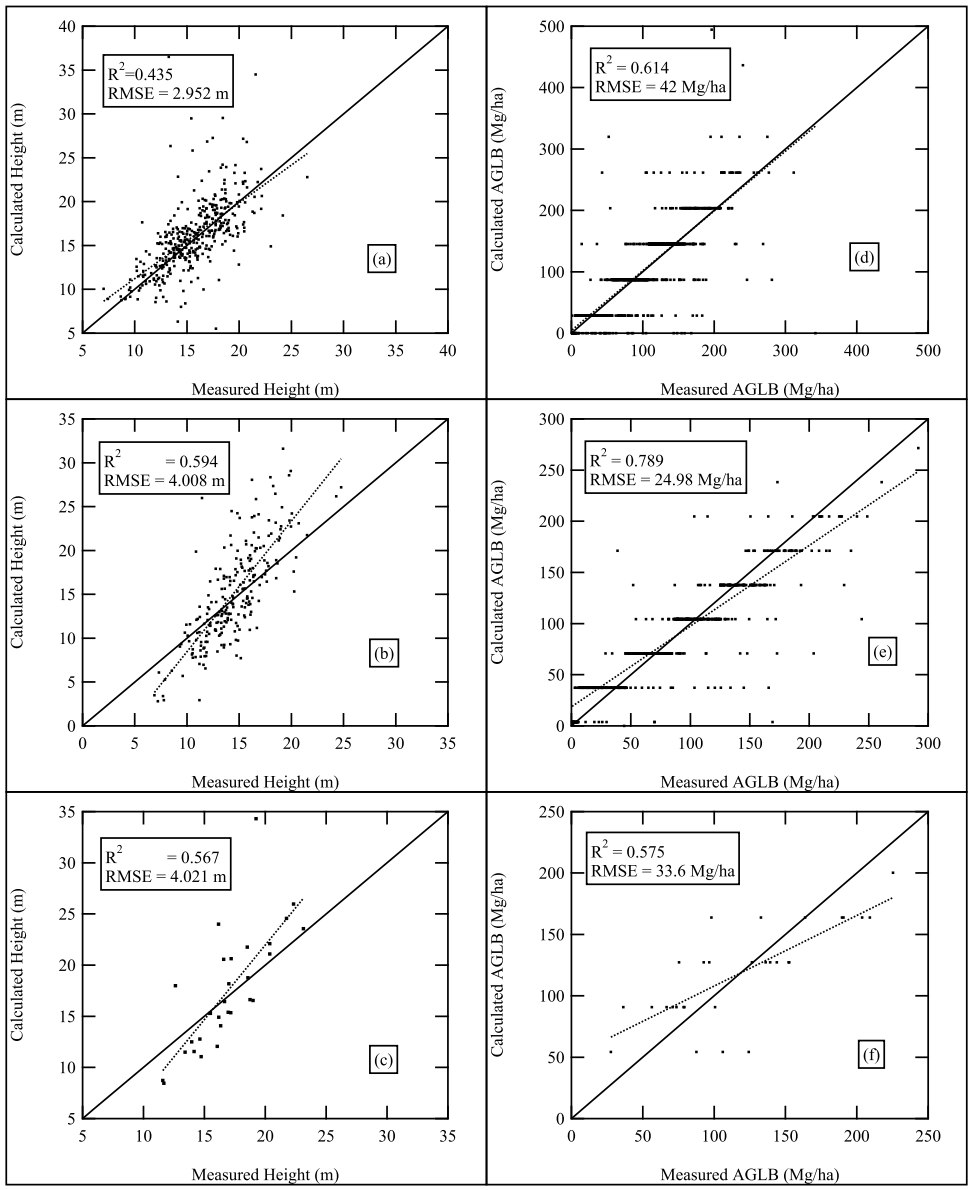

Figure 10. Exponential fit of height and linear fit of biomass plotted against measured values. Solid lines are one to one lines ($y = x$). Dotted lines are best fit lines. All data points were included in these fits. Graphs shown are:mean canopy heights for (a) deciduous, (b) evergreen, (c) mixed; and above ground live biomass for (d) deciduous, (e) evergreen, (f) mixed.

types. This could be easily predicted as the allometric equations are derived from DBH (Diameter at Breast Height). Average canopy height showed an approximately linear relationship with the aboveground biomass and represented only 40–55% of the variations [*Lefsky et al.*, 2005]. A linear relationship derived from regression provided a positive correlation between biomass and height for evergreen (R^2 = 0.39), deciduous (R^2 = 0.39), and mixed forests (R^2 = 0.56). Average height and basal area, however, showed a weak nonlinear relationship in all three forest types. Average height at the plot scale represented only 20–35% of the basal area variations, suggesting forest density and tree distributions within the plot playing a major role in determining the biomass density of the plots. Basal area captures the tree density within the plots, whereas average height is almost independent of tree density. This confirms a long accepted theory in forestry: that the height growth of dominant and co-dominant trees are generally independent of stand density and instead relate strongly to site quality [*Wenger*, 1984]. Another potential structural parameter is the basal area weighted height or the Lorey's height, which is often used in lidar remote sensing [*Naesset*, 1997]. However, in the absence of tree level field data for this study, we were unable to examine the relationship between Lorey's mean height and forest biomass.

5.2. Interpolation of SRTM-NED

[38] There were a total of around 15 million land pixels in the state of Maine (roughly 90 m by 90 m in area per pixel.) Approximately 17% of these pixels had a negative

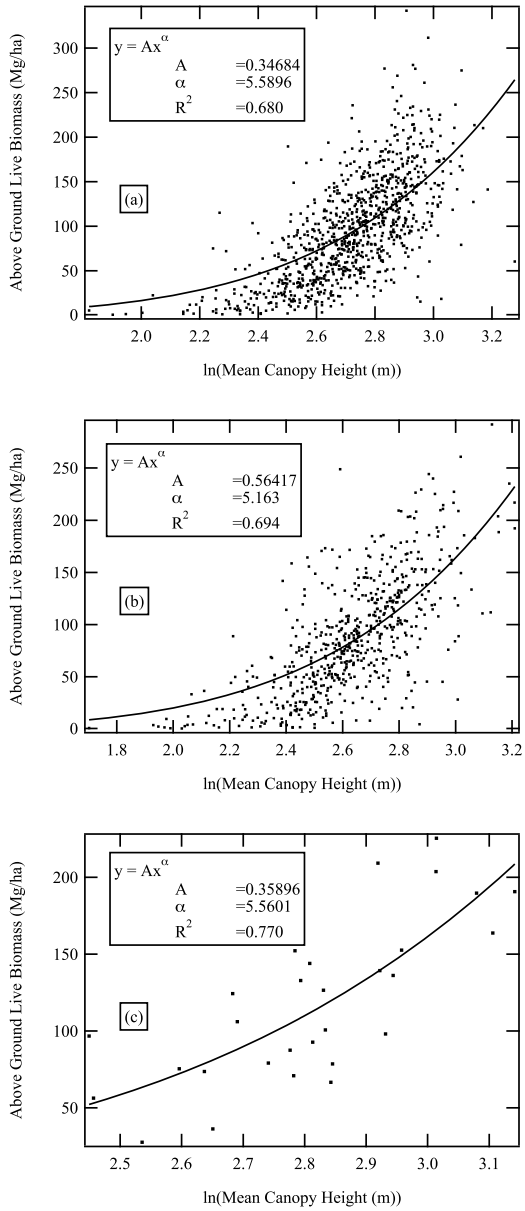

Figure 11. Allometric equations developed from FIA ground data for (a) deciduous, (b) evergreen, and (c) mixed forests between mean canopy height and AGLB.

SRTM-NED difference, and approximately 0.5% had SRTM-NED values that were deemed too high. After running the interpolation algorithm for 50 iterations, about 10% of the land pixels remain with negative SRTM-NED values and most of the pixels with SRTM-NED values that were too high were interpolated. The number of points interpolated each iteration diminishes exponentially with the number of iterations. Overall, after 50 iterations, approximately 44.5% of the pixels that were erroneous were interpolated. After the interpolation, a number of pixels in the SRTM-NED image remained unchanged. Majority of these pixels fall within blocks of troublesome regions where finding enough neighboring "good pixels" for interpolation is difficult. Other regions tend to be areas of rugged terrain or in areas of deciduous trees where the scattering phase centers were low because of the impact of the tree-surface interactions.

[39] The remaining pixels that were not interpolated were marked and had their values in the final biomass and height products set to 0. Not much else can be done for these areas as it is limited by the quality of the remote sensing data available. The USGS is continuously updating the NED data set and results will improve as better NED data becomes available.

5.3. Height and Biomass Algorithms

[40] The linear relationships between SRTM-NED and mean forest canopy height showed the highest correlation values, ranging from $R^2 = 0.587$ for deciduous forests to $R^2 = 0.926$ for mixed forests (Figures 7–9). P-values are less than 0.001 for all three forest types. As expected, the deciduous forests showed the worst R^2 value out of the three groups of forests (Figure 7). This is mostly due to the lack of leaves in the deciduous forests. This would lower the scattering phase center height, as well as introduce more noisy random scatterers of branches. Surprisingly, mixed forest showed the best R^2 value of the three. Root mean squared errors (RMSE) were also calculated for the different relationships. Linear relationship for height had RMSE of approximately 2 m for deciduous forests and 1 m for evergreen forests. Exponential relationship of height had much higher RMSE across the board. Biomass estimations from direct linear relationships had RMSE of 42 tons per hectare for deciduous forests and RMSE of 25 tons per hectare for evergreen forests. Table 3 summarizes the R^2 and RMSE values for the different function types and forest types.

[41] Figures 7c, 8c, and 9c show the results for biomass relationship using corrected plot locations obtained from box search algorithm applied on height. While the R^2 values are lower than those for biomass relationships based on box search results using biomass (Figures 7b, 8b, and 9b), they are still improved from the case where no correction for geo-location errors were performed.

[42] Figure 10 shows the exponential functional fits of mean canopy height as a function of SRTM-NED and Landsat derived VI (equation (2)), as well as linear fit of AGLB as a function of SRTM-NED. The fits shown are the one to one scatter plot between calculated values and field measurements. The R^2 values in the exponential fit of mean canopy height are lower than those obtained through the linear functions. Lowest is still deciduous forests with R^2 of 0.4 while evergreen has a R^2 of 0.6. The simple linear relationship for mean canopy height had higher R^2 values

Table 2. Validation of Results Through Bootstrapping[a]

	Height Linear		Height Exponential		Biomass Linear	
	RMSE (m)	R^2	RMSE (m)	R^2	RMSE (Mg/ha)	R^2
Deciduous	2.1	0.613	3.7	0.331	45.9	0.405
Evergreen	1.3	0.805	4.5	0.600	41.1	0.406
Mixed	0.9	0.912	3.3	0.580	33.5	0.530

[a]Data points were randomly split into two groups: G1 and G2. Functional fits were developed using points in G1, and then applied to G2 to calculate R^2 values and RMSE. All values shown are calculated from G2.

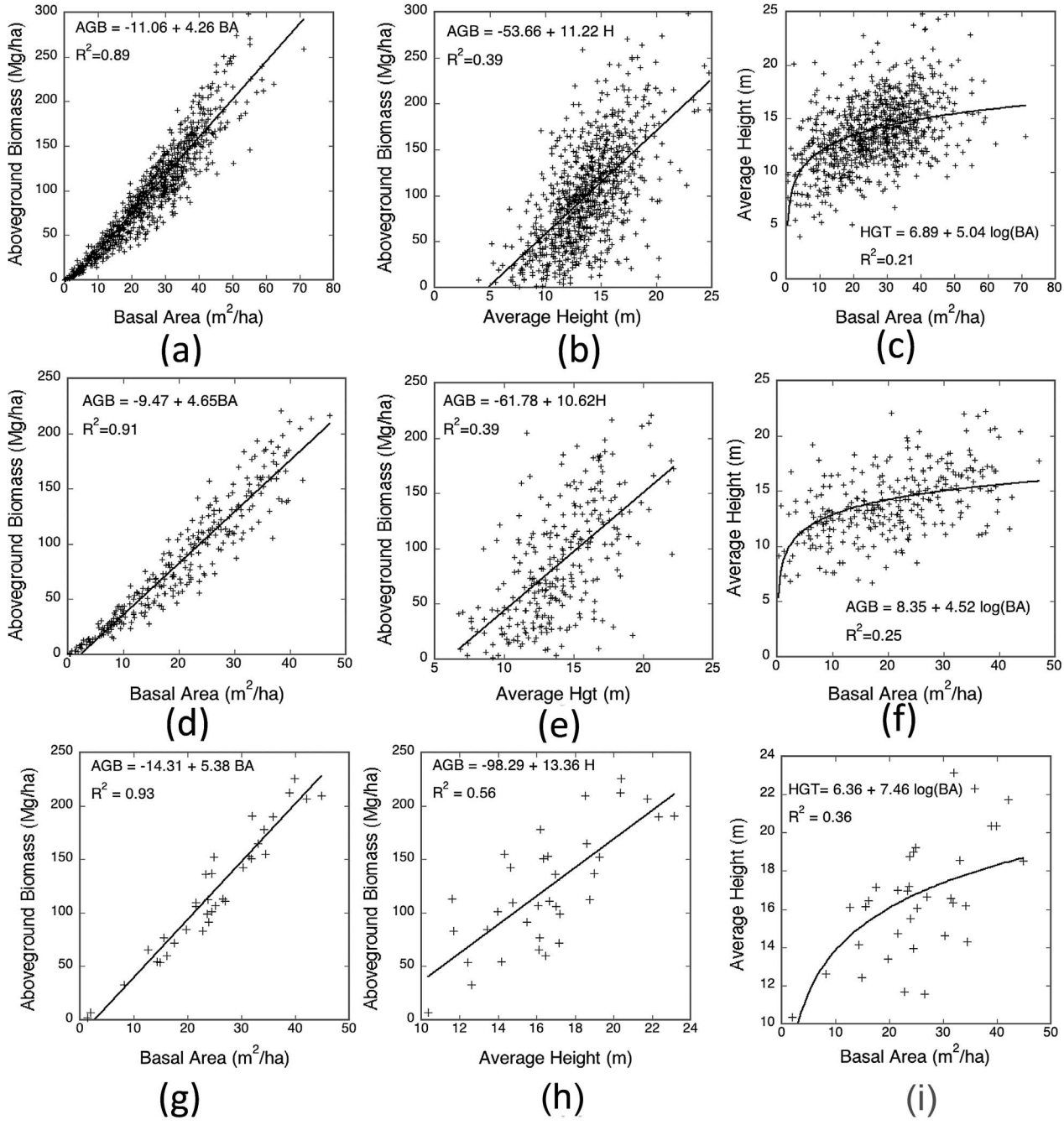

Figure 12. Relationships between basal area, height, and biomass from FIA ground data. (a, b, and c) Deciduous plots; (d, e, and f) evergreen plots; (g, h, and i) mixed plots.

across all three forest types than the exponential relationship with VI.

[43] Table 3 shows a summary of the goodness of fit and error estimation for various forms of functions developed in this study. There are two functions of mean canopy height and two functions of AGLB. The linear function of height and biomass are direct linear correlations between the respective quantities and SRTM-NED. The SRTM-NED values used here were extracted from separate runs of the box search algorithm: one for height and one for biomass. The exponential height values were obtained through the

Table 3. Calculations of Goodness of Fit and Root Mean Square Error[a]

	Height (m)				Biomass (Mg/ha)			
	Linear		Exponential		Direct		From Height	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Deciduous	0.587	1.925	0.435	2.952	0.614	42.13	0.179	57.80
Evergreen	0.855	1.136	0.594	4.008	0.789	24.98	0.246	66.36
Mixed	0.926	0.802	0.567	4.021	0.575	33.60	0.283	70.54

[a]Values are calculated using the entire data set.

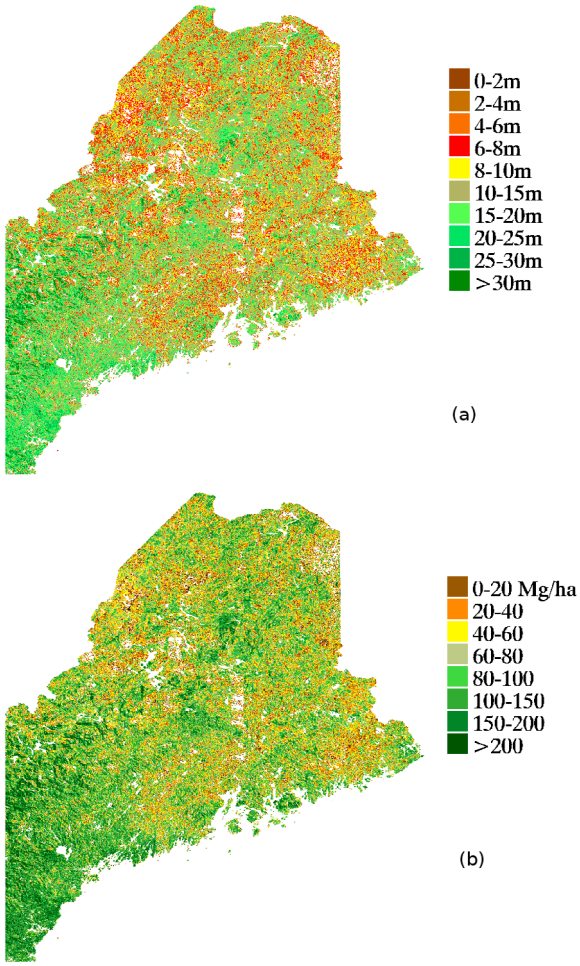

Figure 13. (a) Final mean canopy height and (b) AGLB results for the state of Maine. Products are generated at 270 m resolution.

application of equation (2) with VI as the functional variable. The "From Height" values of biomass are derived by applying the allometric equations shown in Figure 11 to height values derived through the use of equation (2). The root mean squared errors (RMSE) are also shown for all of the functions and all of the forest types. RMSE of height ranged from approximately 0.8 m (mixed) to 2 m (deciduous) with the linear relationship. RMSE of AGLB ranged from 25 tons per hectare (evergreen) to 42 tons per hectare (deciduous).

5.4. Final Height and Biomass Maps

[44] Final images of mean canopy height and AGLB were generated for the state of Maine at 270 m pixel resolution. Figure 13 shows the final biomass and height results. The results were grouped into classes to produce the final images. Mean canopy height is grouped into classes of 2 m intervals in the 0–10 m height range, and 5 m intervals for heights above 10 m. For AGLB, we used 20 Mg/ha intervals for values less than 100 Mg/ha range, and 50 Mg/ha intervals for biomass values greater than 100 Mg/ha to approximately represent the errors in biomass estimation. Large spatial variation in the height and AGLB of the forests are also visible in the images. Areas with negative SRTM-NED or missing data that could not be corrected by the interpolation algorithm were blocked from height and biomass maps by setting their values to zero.

[45] For the creation of the biomass map, we used a combination of the linear and exponential equations (from Figures 7a and 7b, Figures 8a and 8b, and Figures 9a and 9b as well as data from Figures 7b, 8b, and 9b fitted to exponential functions) even though plot data gave the highest R^2 value for the linear fit. The equations were applied at the 90 m level, and then re-sampled to 270 m. Physically, one would not expect the biomass to continue to scale linearly with height after a certain point. Within the ground plots provided, there were very few plots with high biomass values, so the linear equation provided an adequate fit. However, when generating a product image, the linear relationship will break down over areas of high forest height and biomass. To take into account of this effect, we applied the exponential function for areas where SRTM-NED values exceeded those sampled in the field plots.

6. Discussion

6.1. Comparison of Results

[46] Some statistics of biomass values over the state of Maine are calculated and displayed in Tables 4 and 5. The distribution of biomass corresponds well with the division of the three forest categories by area with AGLB fairly evenly distributed across all forest types. The statistics for this study are calculated from the AGLB map at 90 m resolution. However, this resolution would not be well suited for producing a final product. This is due to the fact that at such fine resolution, errors in spatial registration between the SRTM and NED images would create errors in biomass allocation in the final image. To mitigate this problem, the final biomass map was produced at 270 m resolution where an average biomass value was calculated within a 3 × 3 window of the 90 m image using forested pixels of NLCD to represent an average forest biomass instead of an average biomass for the pixel.

[47] Table 5 shows a comparison of the overall statistics of the biomass generated from this study with the biomass map provided by FIA. While the mean values and the range are very similar between the two sets of biomass maps, the standard deviation from our analysis is larger than the FIA map. It is very difficult to interpret this difference. FIA map has been derived from extrapolation of plot data over the region using two MODIS spectral bands at 250 m resolution and decision rule approach. As MODIS signal is not sensitive to forest structure, we expect the FIA approach would have an smoothing effect on the final result and hence reducing the variations of the forest biomass over the landscape. A comparison with the statistics of FIA ground

Table 4. Statistics of Height Calculations Over the Entire State of Maine From the Final Generated Maps at 90 m Resolution[a]

	Height Linear (m)		Height Exponential (m)	
	Mean	Standard Deviation	Mean	Standard Deviation
Deciduous	14.4	12.2	14.9	11.1
Evergreen	15.5	7.5	18.5	12.5
Mixed	13.7	6.6	12.9	10.1

[a]Before re-sampling to 270 m.

Table 5. Comparison of Statistics of AGLB From This Study, the FIA Map Derived From MODIS, and FIA Ground Inventory[a]

	This Study			FIA Map			FIA Inventory		
	Deciduous	Evergreen	Mixed	Deciduous	Evergreen	Mixed	Deciduous	Evergreen	Mixed
Minimum	0.66	0.66	8.93	20.18	20.18	20.18	1.15	0.99	33.1
Maximum	211.5	212.6	212.6	213.0	215.2	215.2	341.81	298.36	225.4
Mean	121.6	115.6	111.2	116.0	103.7	108.8	116.1	107.7	131.2
Standard Deviation	45.8	44.2	36.3	20.8	21.7	20.9	56.4	53.2	50.7

[a]Biomass values are in units of Mg/ha.

inventory shows that the standard deviations of biomass for all three categories of forest types are closer to results obtained by our method than the FIA map derived from MODIS. A visual comparison of the two biomass images also confirms the statistical analysis. The general patterns of biomass distribution are similar between the two sets of images over the landscape. However, there are differences on a pixel-by-pixel basis. This suggests that the existing data and analysis can only provide a rough estimate of the AGLB of the forests because of multiple errors associated with estimating the forest height or biomass from SRTM and NED. The AGLB obtained from this methodology had errors up to 40% of the mean biomass. This error does not allow us to capture small scale variations in forest biomass over the landscape. However, large variations can be readily observed in the map generated from our study. We expect that it would be difficult to map the small scale variations (at about 1 hectare) of the biomass accurately without a more direct remote sensing approach. Active remote sensing, as designed for the future NASA DESDynl mission, with both interferometry and polarimetry at L-band (25 cm wavelength) frequency and multibeam lidar are considered the most promising techniques to directly estimate the above ground biomass or forest structure.

6.2. Caveats and Future Application

[48] In theory, the methodology presented in this paper can be applied to another state or to the entire country. There is nothing in our approach that is specific to the state of Maine, except the FIA inventory plots and the remote sensing data. However, applying the methodology to a larger region requires careful consideration of the following problems:

[49] 1. While previous as well as this study have shown that good regression models can be developed between ground plot measurements and the height of the scattering phase center SRTM-NED, areas of intense topography and the poor quality of NED and possibly SRTM will introduce large errors in forest height and biomass estimation. This effect is noticeable in areas of high elevation in western portion of the Maine image.

[50] Aside from certain tiles of NED image which may have systematic error in elevation, the most significant source of error appears to be geo-location errors. This is evident when comparing the SRTM image with the NED image. The easiest method to see this geo-location error is by generating images of the aspect of the two data sets, and comparing the two aspect images. The geo-location errors would produce artifacts of topography mentioned above. This is because in areas of high slope, even small geo-location errors can create large errors in SRTM-NED. This error can be reduced by producing a coarser resolution image for the final product.

[51] 2. The techniques developed here should work well towards that goal. Even though the interpolation algorithm can correct part of the erroneous pixels, some areas, especially of intense topography, can still prove challenging. The methodologies developed use remote sensing data which are available for the entire US. Calibration for different regions can be accomplished by breaking the entire extent into smaller regions. The same procedure can then be followed which includes obtaining local threshold values. This would produce best results if the regions were chosen based on similar surface and vegetation characteristics. Based on the analysis in this paper, we recommend a stratification of landscape compatible with biome types and elevation gradients.

[52] 3. Generating the final height or biomass map at a coarser resolution to avoid geo-location or missing data points can introduce errors associated with the scaling up of the algorithm or the products. The scaling errors are related to the surface heterogeneity. By aggregating the remote sensing data before applying the estimation algorithm, we may encounter errors due to mixing vegetation types and characteristics. We performed some statistical analysis on the heterogeneity of the forests in Maine and found that by aggregating pixels from 1 ha to 100 ha (1 km resolution), more than 60% of forested pixels changed from hardwood or softwood to mixed forests (assuming more than half of the aggregated pixel is mixed). Applying the mixed forest biomass algorithm to a large pixel comprising of smaller pixels of homogeneous stands is not the same as aggregating the biomass of smaller pixels. This is especially important when coarse resolution remote sensing data are used with algorithms developed at smaller scales. In general, one expects the accuracy of estimation to be optimum when the resolution of the remote sensing data, the plot size used in algorithm development, and the pixel size for algorithm implementation are the same. An alternative approach would be to use a multiscale probabilistic technique with ground data at one scale and as many remote sensing data layers as possible at different resolutions. This would utilize as much information as possible from different remote sensing data sets to complement each other. This approach would also be prune to errors. However, it provides a rigorous upscaling mechanism and improved estimation accuracy at some optimum scale [*Irving and Willsky*, 2001; *Slatton et al.*, 2005; *Saatchi et al.*, 2010].

[53] We did not include an error analysis of biomass estimation that can include measurement errors associated with SRTM or NED data sets. This is mainly due to the fact that there was no reasonable model to quantify and include errors associated with both data sets in the estimation model.

Errors associated with the phase noise in SRTM data can be included as an additive noise term in the estimation model [*Kellndorfer et al.*, 2004]. The phase noise error can be reduced by multilooking (averaging) the radar imagery and hence coarsening the spatial resolution. However, this approach may not work in this study for the following reasons: (1) There is no plot data at coarse resolution to compare with the averaged product. (2) Multilooking or averaging will reduce the noise error but does not reduce the overall error if the biomass estimation is biased. (3) The SRTM data is produced by multiple polarizations and different looks. There is no analytical or numerical model to quantify the radar performance error for each pixel [*Rodriguez et al.*, 2006].

7. Conclusion

[54] The work was produced on a spatial resolution of roughly 90 m, with final product images generated at 270 m resolution. While good correlation can be developed between scattering phase center height of SRTM and field measurements of mean canopy height with appropriate filtering of plots, caveats exist when attempting to apply this method for generating products over a large region. An interpolation algorithm was developed to correct for erroneous pixels and other ancillary remote sensing data were used to help constrain the final output. Even after applying correction techniques, some regions such as areas of intense topography can still be problematic. It is most likely that the main source of error is the geo-location between the SRTM and reference ground height from NED. Although caveats exist, overall statistical results showed good agreement with published ground inventory data from FIA, with our method estimating a total AGLB of 700 Gg for the state of Maine for the year 2000 (FIA data shows total AGLB estimation of 759 Gg for the year 2002.) This study also provided a tested methodology which can be applied, with a degree of automation, towards the generation of height and biomass maps for the entire United States.

References

Anderson, J., M. E. Martin, M.-L. Smith, R. O. Dubayah, M. A. Hofton, P. Hyde, B. E. Peterson, J. B. Blair, and R. G. Knox (2006), The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire, *Remote Sens. Environ.*, *105*(3), 248–261, doi:10.1016/j.rse.2006.07.001.

Askne, J. I. H., P. B. G. Dammert, P. B. G. Dammert, L. M. H. Ulander, L. M. H. Ulander, and G. Smith (1997), C-band repeat-pass interferometric SAR observations of the forest, *IEEE Trans. Geosci. Remote Sens.*, *35*(1), 25–35.

Butera, M. K. (1986), A correlation and regression analysis of percent canopy closure versus tms spectral response for selected forest sites in the San Juan National Forest, Colorado, *IEEE Trans. Geosci. Remote Sens.*, *24*(1), 122–129.

Dobson, M. C., F. T. Ulaby, F. T. Ulaby, T. LeToan, T. LeToan, A. Beaudoin, E. S. Kasischke, and N. Christensen (1992), Dependence of radar backscatter on coniferous forest biomass, *IEEE Trans. Geosci. Remote Sens.*, *30*(2), 412–415.

Drake, J. B., R. O. Dubayah, D. B. Clark, R. G. Knox, J. B. Blair, M. A. Hofton, R. L. Chazdon, J. F. Weishampel, and S. Prince (2002), Estimation of tropical forest structural characteristics using large-footprint lidar, *Remote Sens. Environ.*, *79*(2–3), 305–319.

Dubayah, R. O., and J. B. Drake (2000), Lidar remote sensing for forestry, *J. For.*, 98, *44–46*(3).

Foley, J. A., et al. (2005), Global consequences of land use, *Science*, *309*(5734), 570–574.

Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler (2002), The national elevation dataset, *Photogramm. Eng. Remote Sens.*, *68*, 5–11.

Goodale, C. L., et al. (2002), Forest carbon sinks in the northern hemisphere, *Ecol. Appl.*, *12*(3), 891–899.

Hagberg, J. O., L. M. H. Ulander, L. M. H. Ulander, J. Askne, and J. Askne (1995), Repeat-pass sar interferometry over forested terrain, *IEEE Trans. Geosci. Remote Sens.*, *33*(2), 331–340.

Heo, J., J. W. Kim, S. Pattnaik, and H.-G. Sohn (2006), Quality improvement of loblolly pine (pinus taeda) plantation inventory gis using shuttle radar topography mission (STRM) and the national elevation dataset (NED), *For. Ecol. Manage.*, *233*(1), 61–68.

Homer, C., C. Q. Huang, L. Yang, B. Wylie, and M. Coan (2004), Development of a 2001 national land-cover database for the United States, *Photogramm. Eng. Remote Sens.*, *70*(7), 829–840.

Horler, D., and F. Ahern (1986), Forestry information-content of thematic mapper data, *Int. J. Remote Sens.*, *7*(3), 405–428.

Houghton, R. A., J. L. Hackler, and K. T. Lawrence (1999), The U.S. carbon budget: Contributions from land-use change, *Science*, *285*(5427), 574–578.

Huete, A., K. Didan, T. Miura, E. Rodriguez, X. Gao, and L. Ferreira (2002), Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sens. Environ.*, *83*(1–2), 195–213.

Hurtt, G. C., R. Dubayah, J. Drake, P. R. Moorcroft, S. W. Pacala, J. B. Blair, and M. G. Fearon (2004), Beyond potential vegetation: Combining lidar data and a height-structured model for carbon studies, *Ecol. Appl.*, *14*(3), 873–883.

Irving, W., and A. Willsky (2001), A canonical correlations approach to multiscale stochastic realization, *IEEE Trans. Autom. Control*, *46*(10), 1514–1528.

Keeling, C. D., J. F. S. Chin, and T. P. Whorf (1996), Increased activity of northern vegetation inferred from atmospheric CO_2 measurements, *Nature*, *382*(6587), 146–149.

Kellndorfer, J., W. Walker, L. Pierce, C. Dobson, J. A. Fites, C. Hunsaker, J. Vona, and M. Clutter (2004), Vegetation height estimation from shuttle radar topography mission and national elevation datasets, *Remote Sens. Environ.*, *93*(3), 339–358.

Kobayashi, Y., K. Sarabandi, K. Sarabandi, L. Pierce, L. Pierce, and M. C. Dobson (2000), An evaluation of the jpl topsar for extracting tree heights, *IEEE Trans. Geosci. Remote Sens.*, *38*(6), 2446–2454.

Lefsky, M. A., W. B. Cohen, G. G. Parker, and D. J. Harding (2002), Lidar remote sensing for ecosystem studies, *BioScience*, *52*(1), 19–30.

Lefsky, M., D. Harding, M. Keller, W. Cohen, C. Carabajal, F. Espirito-Santo, M. Hunter, and R. de Oliveira (2005), Estimates of forest canopy height and aboveground biomass using ICESat, *Geophys. Res. Lett.*, *32*, L22S02, doi:10.1029/2005GL023971.

Le Toan, T., and N. Floury (1998), On the retrieval of forest biomass from SAR data, paper presented at Workshop on Retrieval of Bio- and Geo-Physical Parameters From SAR Data for Land Applications, Eur. Space Res. and Technol. Cent., Noordwijk, Netherlands, 21–23 Oct.

Little, E. L. J. (1979), Checklist of United States trees (native and naturalized), Dep. of Agric., For. Serv., Washington, D. C.

Naesset, E. (1997), Determination of mean tree height of forest stands using airborne laser scanner data, *ISPRS J. Photogramm. Remote Sens.*, *52*(2), 49–56.

Papathanassiou, K. P., S. R. Cloude, and S. R. Cloude (2001), Single-baseline polarimetric sar interferometry, *IEEE Trans. Geosci. Remote Sens.*, *39*(11), 2352–2363.

Puhr, C. B., and D. N. M. Donoghue (2000), Remote sensing of upland conifer plantations using Landsat tm data: A case study from Galloway, south–west Scotland, *Int. J. Remote Sens.*, *21*(14), 633–646.

Reigber, A., A. Moreira, and A. Moreira (2000), First demonstration of airborne SAR tomography using multibaseline l-band data, *IEEE Trans. Geosci. Remote Sens.*, *38*(5), 2142–2152.

Rodriguez, E., C. S. Morris, and J. E. Belz (2006), A global assessment of the SRTM performance, *Photogramm. Eng. Remote Sens.*, *72*(3), 249–260.

Saatchi, S., and K. McDonald (1997), Coherent effects in microwave backscattering models for forest canopies, *IEEE Trans. Geosci. Remote Sens.*, *35*(4), 1032–1044.

Saatchi, S. S., M. Moghaddam, and M. Moghaddam (2000), Estimation of crown and stem water content and biomass of boreal forest using polarimetric sar imagery, *IEEE Trans. Geosci. Remote Sens.*, *38*(2), 697–709.

Saatchi, S., M. Marlier, R. L. Chazdon, D. B. Clark, and A. E. Russell (2010), Impact of spatial variability of forest structure on radar estimation of aboveground biomass in tropical forests, *Remote Sens. Environ.*, in press.

Sarabandi, K., and Y. C. Lin (2000), Simulation of interferometric SAR response for characterizing the scattering phase center statistics of forest canopies, *IEEE Trans. Geosci. Remote Sens.*, *38*(1), 115–125.

Schimel, D. S., et al. (2001), Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems, *Nature*, *414*(6860), 169–172.

Simard, M., K. Zhang, V. H. Rivera-Monroy, M. S. Ross, P. L. Ruiz, E. Castaneda-Moya, R. R. Twilley, and E. Rodriguez (2006), Mapping height and biomass of mangrove forests in Everglades National Park with SRTM elevation data, *Photogramm. Eng. Remote Sens.*, *72*(3), 299–311.

Slatton, K. C., K. Nagarajan, V. Aggarwal, H. Lee, W. Carter, and R. Shrestha (2005), Multiscale estimation of terrain complexity using alsm point data on variable resolution grids, in *Gravity, Geoid and Space Missions: International Association of Geodesy Symposia*, Springer, Heidelberg, Germany.

Smith, W. B., P. D. Miles, J. S. Vissage, and S. A. Pugh (2004), Forest resources of the United States, 2002, *Gen. Tech. Rep. NC-241*, U.S. Dep. of Agric., For. Serv., North Central Res. Stn., St. Paul, Minn.

Walker, W. S., J. M. Kellndorfer, E. LaPoint, M. Hoppus, and J. Westfall (2007a), An empirical InSAR-optical fusion approach to mapping vegetation canopy height, *Remote Sens. Environ.*, *109*(4), 482–499, doi:10.1016/j.rse.2007.02.001.

Walker, W. S., J. M. Kellndorfer, and L. E. Pierce (2007b), Quality assessment of SRTM C- and X-band interferometric data: Implications for the retrieval of vegetation canopy height, *Remote Sens. Environ.*, *106*(4), 428–448.

Wenger, K. F. (Ed.) (1984), *Forestry Handbook*, 2nd ed., John Wiley, Hoboken, N. J.

Wofsy, S. C., and R. C. Harriss (2002), The North American carbon cycle program (NACP) report of the NACP committee of the U.S. interagency carbon cycle science program, technical report, U.S. Global Change Res. Program, Washington, D. C.

L. S. Heath and E. LaPoint, Forest Inventory and Analysis, Forest Service, U.S. Department of Agriculture, Durham, NH 03824, USA.

Y. Knyazikhin and R. Myneni, Department of Geography and Environment, Boston University, Boston, MA 02215, USA.

S. Saatchi, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA.

Y. Yu, Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095, USA. (yifany@ucla.edu)