APPLYING INVENTORY METHODS TO ESTIMATE ABOVEGROUND BIOMASS FROM SATELLITE LIGHT DETECTION AND RANGING (LiDAR) FOREST HEIGHT DATA

Sean P. Healey, Paul L. Patterson, Sassan Saatchi, Michael A. Lefsky, Andrew J. Lister, Elizabeth A. Freeman, and Gretchen G. Moisen[1]

Abstract.—Light Detection and Ranging (LiDAR) returns from the spaceborne Geoscience Laser Altimeter (GLAS) sensor may offer an alternative to solely field-based forest biomass sampling. Such an approach would rely upon model-based inference, which can account for the uncertainty associated with using modeled, instead of field-collected, measurements. Model-based methods have been thoroughly described in the statistical literature, and an increasing number of model-based forestry applications use tactically acquired airborne LiDAR. Adapting these methods to GLAS's irregular acquisition pattern requires a strategy for identifying a subset of GLAS "shots" that can be considered a simple random sample. We have developed a flexible method of dividing the landscape into equal-area polygons from which a GLAS shot can be chosen at random as a member of the sample. This process bears similarities to the approach used by the Forest Inventory and Analysis (FIA) Program as it moved toward its current hexagonal sample grid.

Although the ultimate application of this approach would be production of consistent biomass estimates across different countries, well-calibrated FIA estimates over the United States provide a convenient testing ground. Applied to California, this approach produced almost exactly the same estimate of biomass density (Mg/ha) as the FIA sample. The GLAS-based estimate had a considerably higher standard error than FIA's estimate, but it comes at a much lower cost and is based upon globally available GLAS measurements.

INTRODUCTION AND METHODS

Most official national and international forest carbon reporting mechanisms have traditionally relied upon information from field-based inventories. However, many countries do not have such inventories, and even among those that do, significant discrepancies exist in methods and definitions. A potential alternative to purely field-based sampling may lie in the 70-m circular waveform Light Detection and Ranging (LiDAR) returns from the spaceborne Geoscience Laser Altimeter System (GLAS) sensor. Such returns have been shown to be sensitive to biomass, and there may be a chance to use a smaller subset of co-located field plots to create modeled biomass "samples" over the areas sampled by GLAS.

In this paper, we describe a method for identifying a subset of GLAS shots which can be treated as a simple random sample for the purposes of a forest biomass inventory. This process is necessary because GLAS acquisition patterns are irregular and, in aggregate, cannot strictly be considered either random or systematic. We then demonstrate the use of such a

[1] Ecologist (SH) and Statistician (PLP), U.S. Forest Service, Rocky Mountain Research Station, 507 25th St., Ogden, UT 84401; Senior Scientist (SS), California Institute of Technology; Assistant Professor (MAL), Colorado State University; Research Forester (AJL), U.S. Forest Service, Northern Research Station; Ecologist (EAF) and Forester (GCM), U.S. Forest Service, Rocky Mountain Research Station. SH is corresponding author: to contact, call 801-625-5770 or email at seanhealey@fs.fed.us.

sample over California with a model-based estimator similar to that used by Ståhl et al. (2011) to estimate aboveground biomass. Model-based estimation allows us to predict, instead of measure, biomass at each sample point using relationships derived from a separate set of co-located ground and LiDAR measurements. Variance estimators used in this process take into account the uncertainty associated with the models used. For full methodological detail of this estimation process, consult Healey et al. (in press).

The sample design we describe is similar to that used by the Forest Inventory and Analysis (FIA) Program. Prior to a move toward a national sampling framework in the late 1990s, FIA plots were distributed and measured in slightly different ways in different regions of the country. The move to a nationally coherent sampling frame was accomplished by superimposing a hexagonal grid over the entire country, with the area of each grid cell equal to the nominal area represented by each FIA sample (Reams et al. 2005). In cells where one existing plot fell, that plot was kept. In those with more than one plot, only one was selected at random for retention. In those with no existing plots, a plot was established in a random location.

Establishment of this semi-systematic, equal-area sample frame allowed FIA to accommodate existing plot locations while drawing a sample which was spatially balanced across the country but was random with respect to forest conditions (Reams et al. 2005). The sample design we propose for GLAS follows a similar approach. One and only one GLAS shot is retained in each cell of an equal-area (but not equal-shape) tessellation of the area labeled as “forest” in a global land cover map. This tessellation is created following a fractal-based approach, using simple geometric rules to create equal-area clusters (Lister and Scott 2009). Since retroactively “adding” GLAS measurements (the last of which were collected in 2008) is not possible, tessellation cell size (and, inversely, sample number) is limited by the constraint that each equal-area cell must contain at least one GLAS shot.

Given the elimination of all GLAS shots except one in every tessellation cell under this approach, it is of practical interest to know the precision of resulting biomass estimates. The precision (i.e., standard error) of model-based estimates of biomass in California using the GLAS sample will be compared to design-based estimates derived from FIA’s sample of more than 5,500 field measurements in the state.

RESULTS

In view of the constraint that there be at least one GLAS shot in each tessellation cell, the maximum number of cells in California is 182, or one per 48,000 ha. While the minimum number of GLAS shots in a single cell was one, the average was 560, from which a single shot was chosen at random. These randomly selected shots constitute the “S1” model-building sample (Fig. 1). The average distance between each point in the S1 population and its closest neighbor is 19.6 km (median = 13.5 km). The minimum overall distance (i.e., closest pair of neighbors) is 2.4 km.

Thirty-five co-located GLAS/FIA plots were available for use in determining the relationship between a GLAS derivative called Lorey’s height (described as “basal area-weighted height”) and biomass (i.e., the S2 sample; Fig. 2). The most parsimonious applicable model for this relationship was considered to be a model with a single quadratic term and no intercept (biomass = 0.3717 Lorey’s height^2). A no-intercept model was used because of our assumption that forested plots with no biomass should return no Lorey’s height. Significance tests indicated negligible gain from including a linear term in the model.

It should be noted that seven values (less than 4 percent) of the S1 sample exceeded the largest value in the model-building S2 data set shown in Figure 2 (specifically, these values were 45, 46, 48, 50, 52, 54, and 60 m). Ideally, the model-building data set should span the entire range of the values to be modeled. Given the small percentage of Lorey’s heights in S1 not represented in S2, however, we assume that the

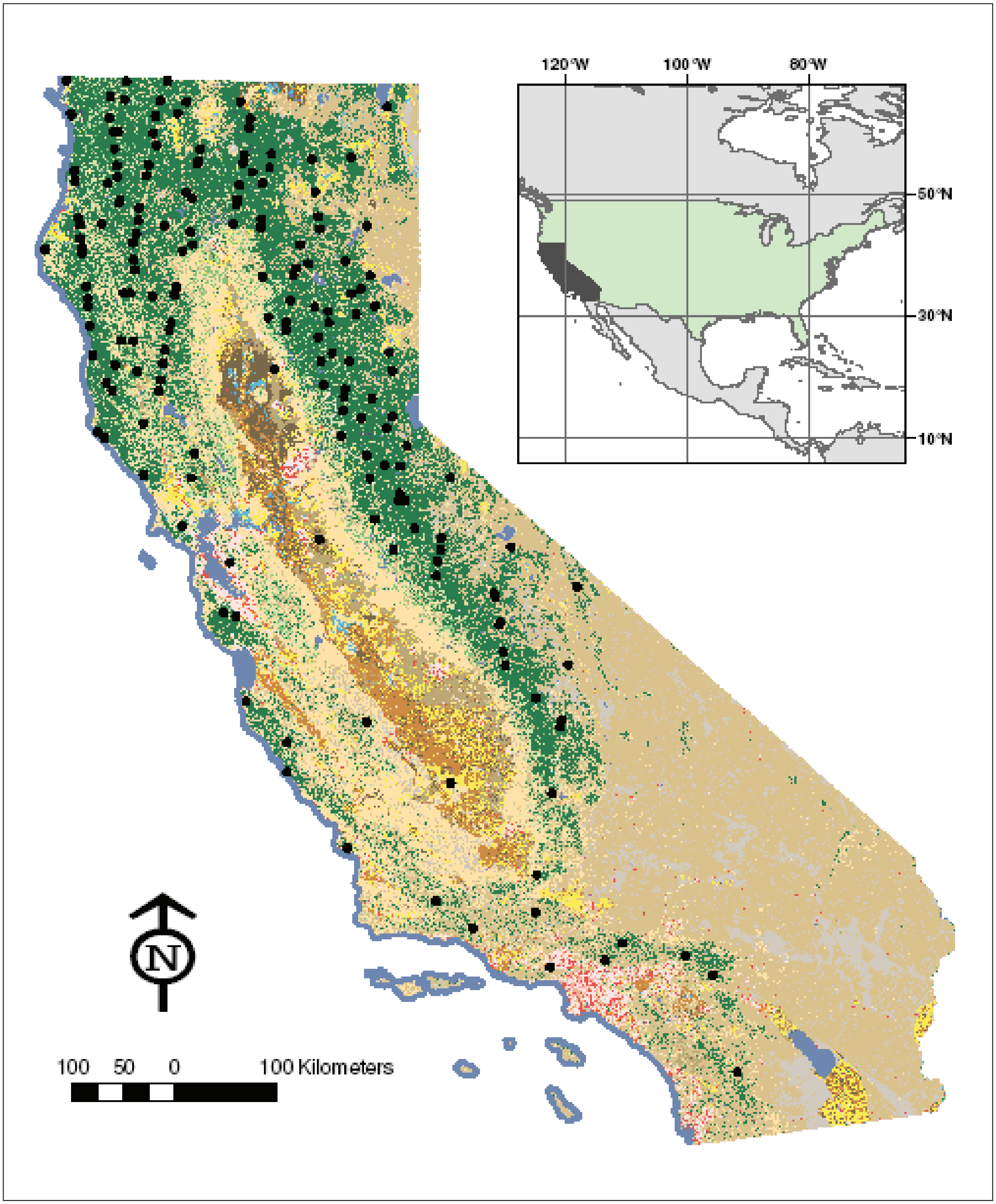

Figure 1.—The 182 GLAS shots selected for inclusion in the S1 sample of California forests. This sample has properties similar to the sample used in the U.S. NFI and is treated here as a simple random sample. A National Land Cover Database cover map is shown for context.

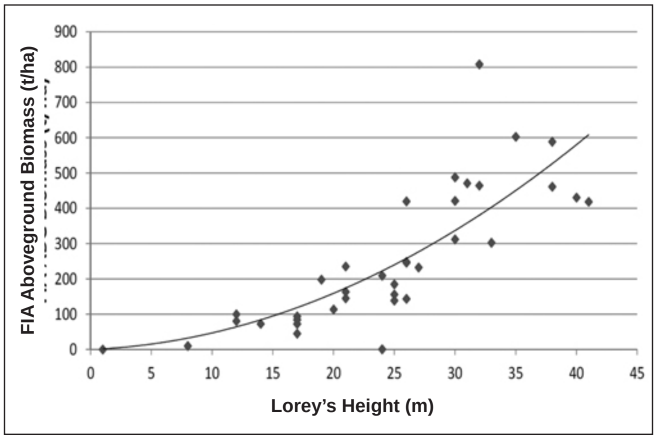

Figure 2.—Relationship between FIA-measured aboveground tree biomass and GLAS-based Lorey's height in California. The line is described by: y = 0.3717 x2.

model is valid for the entire population. We likewise assume no spatial autocorrelation among S1 samples. Our GLAS-based estimate of biomass density in California's forests was 211.11 Mg/ha, which was within standard error bounds (±2.88) of the FIA estimate of 208.95 Mg/ha (Miles 2011). The FIA estimate was derived through a 10-year ground sample of 5,261 forested plots. The standard error of the GLAS-based estimate was 20.70 Mg/ha (Fig. 3). The modeling variance was approximately 0.77 times the variance contributed by the sampling process.

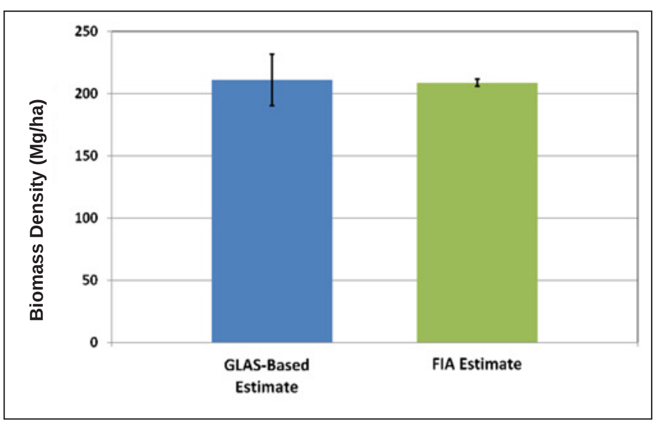

Figure 3.—Comparison between the FIA carbon density estimate for California's forests and the estimate made here using GLAS- and model-based estimation. The estimates are nearly identical, although FIA's estimate has significantly less uncertainty (bars indicate standard error).

DISCUSSION

Model-based estimation using the sample design we describe provides a transparent method for estimating biomass for particular spatial domains. This sample design, in which one arbitrarily located sample point is drawn from equal-area sample units distributed across the landscape, is similar to that used by FIA, and our estimate of biomass density in California closely matched FIA's design-based estimate. The standard error of our estimate (approximately 9.8 percent of the estimate) was substantially larger than that of the FIA estimate (1.4 percent) and that derived through model-based estimation by Andersen et al. (2011) using specifically acquired airborne LiDAR data (8 percent). However, the cost of the FIA estimate was approximately $10.5 million (based on a commonly used valuation of $2,000 per plot), and the LiDAR acquisition alone in Anderson et al.'s (2001) much smaller study area cost $60,000. Future use of GLAS data in the process described here represents an almost no-cost option for providing consistent, moderate-precision biomass estimates across the globe.

A primary advantage of the model-based inference used here is the capacity to apply models developed in areas of rich inventory data to GLAS shots, informing estimates in ecologically similar areas where field data are sparse. For example, Nelson et al. (2009) used relationships observed in a limited area of co-located biomass/GLAS observations to estimate biomass for all of Quebec, following a modified model-based approach. However, the validity of inference in model-based approaches depends upon how well the stipulated models accord with the population of interest (Gregoire 1998). The question of how well the model applies to the population of interest is a critical consideration in the application of our approach, whether the model was developed *in situ* or from a spatially remote but perhaps ecologically similar area. Since our model was created from an arbitrary subset of FIA's presumably unbiased ground sample, there is a compelling argument that the model is appropriate for the forests of California.

The degree to which this model may apply beyond California remains an open question. Saatchi et al. (2011) noted regional differences in the relationship between biomass and GLAS-based Lorey's height in their pan-tropical study. Data collected to support biomass estimation using the global GLAS data set would at least have to span major ecological systems. The consolidation of ground data needed to support a global GLAS-based biomass inventory would require significant international cooperation and, as illustrated by our results, would likely not improve the precision of biomass estimates available in countries with established national forest inventories (NFIs). Those inventories typically rely upon a denser sample than is available from GLAS and do not have to account for model variance, which in our example made up approximately 44 percent of the total variance.

However, a GLAS-based biomass inventory would represent an internationally coherent basis for comparison among countries, especially those without established NFIs. Even moderate-precision biomass estimates would be an improvement in many countries (Gibbs et al. 2007), and consistent sample design and estimation methods would remove an important source of uncertainty in international monitoring. GLAS data were acquired in spatial patterns difficult to associate with either a systematic or random process. The sample design presented in this paper allows identification of a subset of GLAS data which may be used as a simple random sample to estimate biomass, perhaps globally, with consistent measures of uncertainty under a model-based estimation framework.

CONCLUSIONS

- The methods presented here constitute a globally extensible approach for generating a simple random sample from the global GLAS data set. The properties of the sample collected by GLAS have hitherto not been strictly identifiable with any particular design.
- Model-based estimation, following Ståhl et al. (2011), based upon GLAS data in California produced an estimate of biomass density (biomass/ha) almost identical to the estimate derived from the design-based NFI.
- Global application of model-based estimation using GLAS, while demanding significant consolidation of training data, would improve inter-comparability of international biomass estimates by imposing consistent methods and a globally coherent sample frame.

ACKNOWLEDGMENTS

This work was supported by the NASA Carbon Monitoring System and by FIA. The authors additionally would like to thank James Menlove for expert help in interpreting FIA biomass data as well as Karen Waddell and Jock Blackard for their help with the FIA database.

LITERATURE CITED

Andersen, H.-E.; Strunk, J.; Temesgen, H. 2011. **Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the Upper Tanana Valley of Interior Alaska.** Western Journal of Applied Forestry. 26: 157-164.

Gibbs, H.K.; Brown, S.; Niles, J.O.; Foley, J.A. 2007. **Monitoring and estimating tropical forest carbon stocks: making REDD a reality.** Environmental Research Letters. 2: 1-13.

Gregoire, T.G. 1998. **Design-based and model-based inference in survey sampling: appreciating the difference.** Canadian Journal of Forest Resources. 28: 1429-1447.

Healey, S.P.; Patterson, P.L., Saatchi, S.; Lefsky, M.A.; Lister, A.J.; Freeman, E.A. [In press]. **A sample design for globally consistent biomass estimation using LiDAR data from the Geoscience Laser Altimeter System (GLAS).** Carbon Balance and Management.

Lister, A.; Scott, C. 2009. **Use of space-filling curves to select sample locations in natural resource monitoring studies.** Environmental Monitoring and Assessment. 149: 71-80.

Miles, P.D. 2011. **Forest Inventory EVALIDator web-application version 4.01 beta.** Available at http://fiatools.fs.fed.us/Evalidator4/tmattribute.jsp. [Date accessed unknown].

Nelson, R.; Boudreau, J.; Gregoire, T.G.; Margolis, H.; Naesset, E.; Gobakken, T.; Stahl, G. 2009. **Estimating Quebec provincial forest resources using ICESat/GLAS.** Canadian Journal of Forest Resources. 39: 862-881.

Reams, G.A.; Smith, W.D.; Hansen, M.H.; Bechtold, W.A.; Roesch, F.A.; Moisen, G.G. 2005. **The Forest Inventory and Analysis sampling frame.** In: Bechtold, W.A.; Patterson, P.L., eds. The enhanced Forest Inventory and Analysis Program—national sampling design and estimation procedure. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 31-36.

Saatchi, S.S.; Harris, N.L.; Brown, S.; Lefsky, M.; Mitchard, E.T.A.; Salas, W.; Zutta, B.R.; Buermann, W.; Lewis, S.L.; Hagen, S.; Petrova, S.; White, L.; Silman, M.; Morel, A. 2011. **Benchmark map of forest carbon stocks in tropical regions across three continents.** Proceedings of the National Academy of Sciences. 108: 9899-9904.

Stähl, G.; Holm, S.; Gregoire, T.G.; Gobakken, T.; Naesset, E.; Nelson, R. 2011. **Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway.** Canadian Journal of Forest Resources. 41: 96-107.