USING SMALL AREA ESTIMATION AND LIDAR-DERIVED VARIABLES FOR MULTIVARIATE PREDICTION OF FOREST ATTRIBUTES

F. Mauro[1], V.J. Monleon[2], and H. Temesgen[1]

Abstract— Small area estimation (SAE) techniques have been successfully applied in forest inventories to provide reliable estimates for domains where the sample size is small (i.e. small areas). Previous studies have explored the use of either Area Level or Unit Level Empirical Best Linear Unbiased Predictors (EBLUPs) in a univariate framework, modeling each variable of interest at a time, and not considering their potential correlation. Yet most forest inventory variables such as basal area (G) and volume (V) are strongly correlated. In this situation, EBLUPs for multivariate responses can improve the quality of the estimates. In this study, we apply multivariate SAE techniques in a LiDAR assisted forest inventory. We compare the resulting estimates to those obtained using traditional univariate SAE techniques and other synthetic estimates widely used in forest inventories. The study area is a set of Bureau of Land Management (BLM) and Bureau of Indian Affairs (BIA) owned forest lands in Southwestern Oregon. The small areas are the subsets of the BLM\BIA lands in the study area contained in each 12 level Hydrologic Unit Codes (HUC12). Variables of interest were G and V. A total of 899, 0.125 acre plots were measured in the field. Univariate and multivariate fixed effects and mixed effects regression models were developed. Preliminary results show that correlation between HUC12 level random effects for different variables is moderate while residuals for different variables are highly correlated.

Forest planning needs information about forest structure, available stock, health status and other variables at different scales in order to make informed and better management decisions. This information is usually obtained through sampling and is therefore subject to certain amount of uncertainty. Increasing the sample size of field surveys is a possibility to reduce the uncertainty of the estimates, but this solution may not be affordable. A large variety of techniques have been developed to use of inexpensive or easier to obtain auxiliary information to increase sampling efficiency. These techniques can be either design based such as stratification (Hawbaker et al., 2009), ratio/regression estimators, generalized regression estimators (GREG) (Breidenbach and Astrup, 2012), or model based, such as synthetic prediction (Breidenbach and Astrup, 2012; Næsset et al., 2011; Næsset, 2002)2012; N\\uc0\\u230{}sset et al., 2011; N\\uc0\\u230{}sset, 2002. In general, these methods provide accurate estimates for large populations.

However, in addition to estimates of means and totals for the whole population, where sample sizes are usually large, estimates are needed for subpopulations where the sample size is reduced. Depending on the context, these subpopulations can be stands, counties, species or other units, and they are usually referred to as small areas. Design based and model assisted estimators are, in general, unbiased or nearly so for these small areas. Unfortunately, these estimators are unreliable for small areas because of their large variances when the sample size is small. On the other hand, model based synthetic estimators for subpopulations may not suffer the abovementioned problem of high variance, but they assume that all small areas follow a model that is common for the whole population. If there is significant variability

[1] Post Doctoral Scholar (FM), Department of Forest Engineering, Resources and Management College of Forestry, Oregon State University 023 Peavy Hall, Corvallis, OR 97331, USA.; and Professor of Forest Biometrics (HT), Department of Forest Engineering, Resources and Management College of Forestry, Oregon State University 232 Peavy Hall, Corvallis, OR 97331, USA. (FM) is corresponding author: to contact, call (541) 602-7694 or e-mail at francisco.mauro@oregonstate.edu.

[2] Research Mathematical Statistician (VM), Forest Service, Pacific Northwest Research Station Forest Service, Pacific Northwest Research Station Resource Monitoring and Assessment Program, 3200 SW Jefferson Way Corvallis, OR 97331.

among small areas, then there may be a high risk of obtaining biased estimates for subpopulations. For those cases, Empirical Best Linear Unbiased Predictors (EBLUP's), based on linear mixed models, have been proposed (Rao and Molina, 2015) as one of the main alternatives to gain efficiency by using auxiliary information and correct both the high variance of design based estimators and the bias of model based synthetic estimators

The need for suitable methods to provide estimates for small subpopulations has been recognized by Næsset, et al. (2011 p. 3612) and different studies have explored the use of EBLUPs in forest inventories assisted by remote sensing auxiliary information. Area level EBLUPs were used by Goerndt et al., (2011), and unit level EBLUPs have been tested by Breidenbach and Astrup, 2012; Goerndt et al., 2013; Magnussen et al., 20142012; Goerndt et al., 2013; Magnussen et al., 2014. In those studies, EBLUP's were derived from univariate linear mixed models where only one response variable was considered at each time. Certain results from the SAE literature show that, when the residuals and small area random effects for different variables of interest are correlated, an additional gain in efficiency can be obtained if the EBLUP's are based on multivariate models (i.e Multivariate EBLUP's MEBLUP's), that correctly account for those correlations (Datta et al., 1999; Molina, 2009). Certain attributes of interest for forest inventories such as basal areas (G) and volumes (V) might be highly to moderately correlated, so that potential gains in efficiency can be expected if estimates for these variables are based on MEBLUP's. No study to the date has analyzed this potential improvement in forest inventories.

This paper explores with a case study how estimated means for G and V can be improved when using MEBLUP's instead of univariate EBLUP's (UEBLUP's hereafter), based on LiDAR auxiliary information.

STUDY AREA

The study area is the set of Bureau of Land Management (BLM) and Bureau of Indian Affairs (BIA) owned lands covered by the DOGAMI LiDAR survey carried out in Southwestern Oregon. Approximately 1,630,000 acres were covered by the LiDAR flight, of which 254,389 acres are managed by BLM and BIA. Coastal coniferous forest with variable degree of species mixing is prevalent, being Douglas fir (Pseudotsuga Menziesii (Mirb) Franco) the dominant species. Other softwoods such as western hemlock (Tsuga hetorphila (Raf.) Sarg.), sitka spruce (Picea sitchensis (Bong.) Carr.) and red cedar (Thuja plicata Donn ex D.Don) are relatively frequent. The most frequent hardwoods species are red alder (Alnus rubra Bong.), bigleaf maple (Acer macrophyllum Pursh), Oregon myrtille (Umbellularia californica (Hook. & Arn.) Nutt.) and tanoak (Notholithocarpus densiflorus (Hook. & Arn.) Manos, Cannon & S.H.Oh). In general, hardwood species are frequent in the understory and play a secondary role in the upperstory, where conifers are by far more abundant.

LIDAR DATA

LiDAR data were collected during the spring and summer of 2008 and 2009 using a Leica ALS Phase II laser. The average return density was 0.761returns/ft^2. Flight and laser sensor specifications are provided in Table 1.

Table 1—Flight parameters and sensor specification

	Description
Sensor	Leica ALS50 Phase II
Flying altitude	3000 ft above ground level
Field of view	28° (±14° from nadir*)
Pulse rate	> 105 kHz
Pulse mode	Single
Mirror scan rate	52.5 Hz
Overlap	100 % (50% side-lap)

A grid was overlaid on the study area and traditional LiDAR height covariates, including extreme values, percentiles, and fractions of pulses above different thresholds were computed for each pixel using FUSION (Mc Gaughey, 2010). The pixel size was 0.125 acres (75 ft side). Only returns with heights of at least 3.28 ft above the ground were used to compute LiDAR covariates. The return threshold for computing cover covariates was set at a height of 6.56 ft. Pixels where the 80th percentile of the LiDAR returns was higher than 9.85 ft and where more than 2% of the first returns were above 19.69 ft were regarded as forested and pixels that did not meet these criteria were removed. This forest mask was visually assessed to ensure that omission and commission errors were low. Total area of the BLM\BIA owned lands classified as forest according to those criteria was 251,000 ac (≈2 million pixels).

GROUND INVENTORY AND VARIABLES OF INTEREST

Forested areas were stratified based on two LiDAR metrics (80th percentile and standard deviation of the LiDAR heights). A total of 30 strata were defined and, within each one, a random sample of 30 pixels was selected to be measured in the field. Field crews visited all the selected locations except one and measured a total of 899 circular plots of approximately the same area as the pixels.

The diameter at breast height (DBH), height and species of every tree larger than 5.5 in were recorded. Trees smaller than 5.5 in were measured only in concentric plots of 0.02 ac. Standing volume for each tree was computed using regional species specific volume equations included in the US National Volume Estimator Library. Tree volumes were aggregated at plot level and then converted to per acre values applying the corresponding expansion factors to large (DBH≥ 5.5 in) and small trees (DBH< 5.5 in). Basal area per acre was similarly computed for each plot.

SUBPOPULATIONS OF INTEREST

The study area was divided in smaller subpopulations using the 12 level Hydrologic Unit Codes (HUC12). The BLM and BIA owned lands included in each HUC12 estimates were considered as small areas, and mean per hectare values of V and G, were obtained for each one. The study area, forest mask and HUC12 used to define subpopulations are shown in Figure 1.

MODEL SELECTION

Univariate Models

A method similar to the one described in Goerndt et al., (2011), based on fitting linear fixed effects models, was applied to select the LiDAR predictors for each response variable. Once the auxiliary variables were selected, linear mixed effects models with the same LiDAR predictors and random intercepts for each HUC12 were obtained. Significance of HUC12 level random effects was tested for each model.

Multivariate Models

The LiDAR predictors associated with each variable of interest were not modified and correlation between HUC12 random effects associated to V and G were considered in the multivariate model. The plot or pixel levels random effects for different variables were considered independent for this analysis.
In addition, multivariate fixed effects models considering the correlation between residuals for different variables were fit.

RESULTS AND CONCLUSION

In both univariate and models the HUC12 random effects were significant. When modelled together, the correlation between the HUC12 random effects for each variable was moderate and negative (-0.155). Multivariate fixed effects models showed that the correlation between residuals for G and V was strong and positive (0.952). Further analyses will develop models that include both the correlation between random effects for each variable within the HUC12 and the correlations between residuals for different variables.

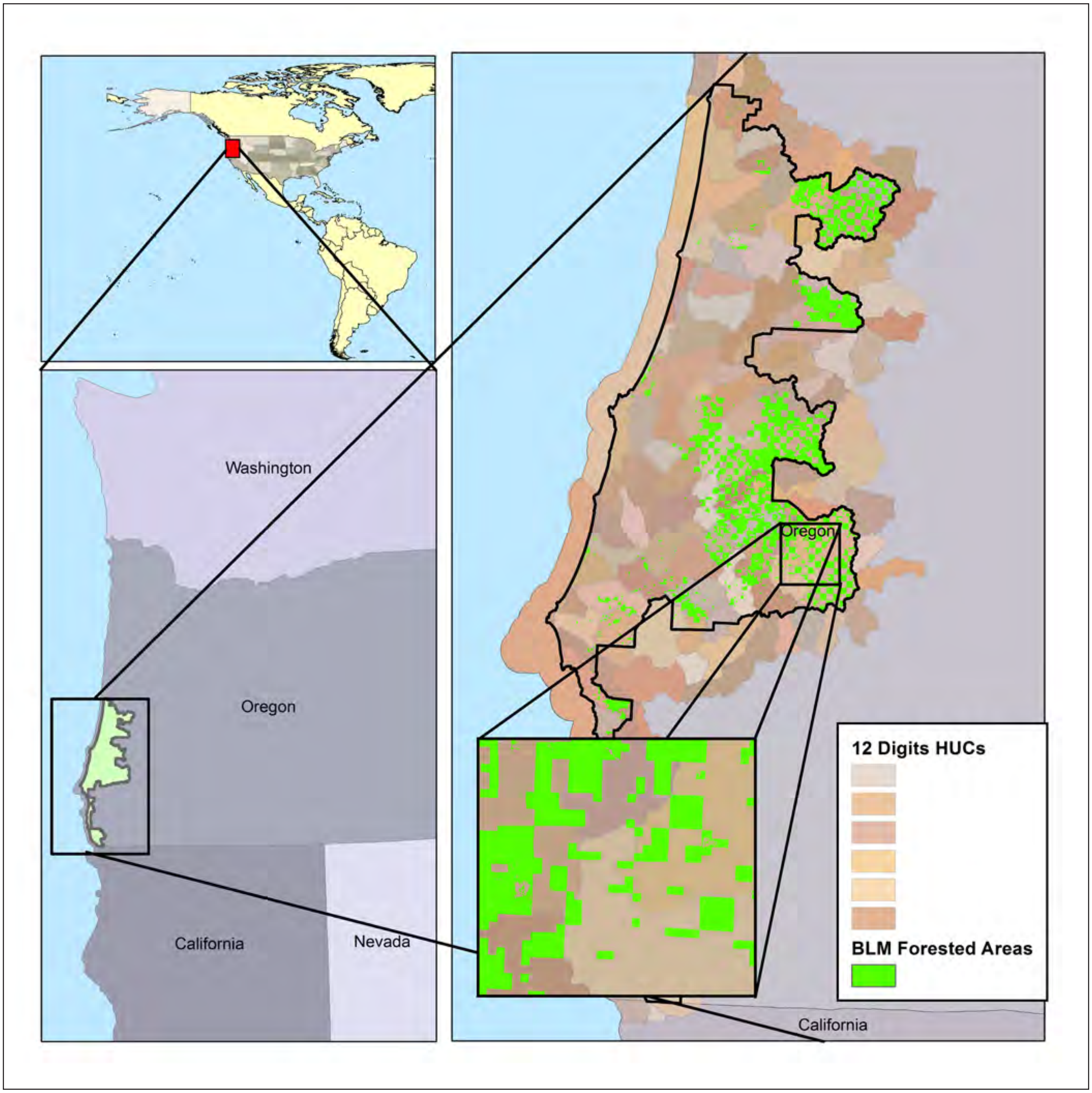

Figure 1—Study area. BLM\BIA owned forest lands (green) and HUC12 employed to define subpopulations of interest.

LITERATURE CITED

Breidenbach, J., Astrup, R., 2012. Small area estimation of forest attributes in the Norwegian National Forest Inventory. Eur. J. For. Res. 131 (4), 1255-1267.

Datta, G.S., Day, B., Basawa, I., 1999. Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation. J. Stat. Plan. Inference 75 (2), 269–279.

Goerndt, M.E., Monleon, V.J., Temesgen, H., 2011. A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. Can. J. For. Res. 41 (6), 1189–1201.

Goerndt, M.E., Monleon, V.J., Temesgen, H., 2013. Small-Area Estimation of County-Level Forest Attributes Using Ground Data and Remote Sensed Auxiliary Information. For. Sci. 59 (5), 536–548.

Magnussen, S., Mandallaz, D., Breidenbach, J., Lanz, A., Ginzler, C., 2014. National forest inventories in the service of small area estimation of stem volume. Can. J. For. Res. 44 (9), 1079–1090.

Mc Gaughey, R.J., 2011. FUSION/LDV: Software for LIDAR Data Analysis and Visualization.

Molina, I., 2009. Uncertainty under a multivariate nested-error regression model with logarithmic transformation. J. Multivar. Anal. 100 (5), 963–980. doi:10.1016/j.jmva.2008.09.007

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80 (1), 88–99.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G., Weydahl, D., 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. Remote Sens. Environ. 115 (12), 3599–3614. doi:10.1016/j.rse.2011.08.021

Rao, J.N.K and Molina, I., 2015. Small Area Estimation, 2nd Ed. London, Wiley & Sons, Inc.