HIERARCHICAL MODELS FOR INFORMING GENERAL BIOMASS EQUATIONS WITH FELLED TREE DATA

Brian J. Clough, Matthew B. Russell, Christopher W. Woodall, Grant M. Domke, and Philip J. Radtke[1]

Abstract— We present a hierarchical framework that uses a large multispecies felled tree database to inform a set of general models for predicting tree foliage biomass, with accompanying uncertainty, within the FIA database. Results suggest significant prediction uncertainty for individual trees and reveal higher errors when predicting foliage biomass for larger trees and for conifers. Consequently, we found large uncertainties when applying the fitted models to predict plot-scale foliage biomass for FIA data within Minnesota. These results suggest that applying general equations with fixed parameters may ignore significant error when used to estimate foliage biomass within the FIA database.

INTRODUCTION

The National Greenhouse Gas Inventory (NGHGI) requires that forest biomass component pools are quantified at the national scale, and within FIA this is accomplished by aggregating biomass estimates calculated for individual trees (Woodall et al., 2011). Currently these are derived from a set of equations with fixed parameters (Jenkins et al., 2003), which fail to account for uncertainty when estimating biomass pools at the tree scale (Wayson et al., 2014; Weiskittel et al., 2015). This limitation may be particularly problematic when estimating variable and dynamic biomass components such as tree foliage. Recently a large felled-tree database has been compiled by the USFS Volume Biomass Project, providing the opportunity to inform uncertainty surrounding biomass models with field-measured data for many North American tree species. We used these data to address two specific objectives: (1) assess the expected uncertainty range of foliage biomass at the tree-scale; and (2) quantify the effect of these errors on plot-level estimates of foliage biomass within a set of FIA data.

[1] Postdoctoral Research Associate (BJC), Department of Forest Resources, University of Minnesota, St. Paul, MN 55108; Assistant Professor of Forest Ecosystem Health (MBR), University of Minnesota, St. Paul, MN 55108; Research Forester (CWW), Northern Research Station, Forest Inventory and Analysis, USDA Forest Service, St. Paul, MN 55108; Research Forester (GMD), Northern Research Station, Forest Inventory and Analysis, USDA Forest Service, St. Paul, MN 55108; Associate Professor of Forest Biometrics and Modeling (PJR), Virginia Polytechnic Institute, Blacksburg, VA 24061; BJC is corresponding author: to contact, call (609) 480 7952 or e-mail at bclough@umn.edu.

STUDY AREA

The felled tree data, which were compiled from many previously published and unpublished studies (hereafter referred to as "legacy data"), come from 130 unique locations spanning the United States and Canada. Models fitted to these data were applied to estimate foliage biomass and associated uncertainty across the state of Minnesota, United States.

METHODS

Data

The legacy data we utilized consist of 5690 observations of foliage biomass (kg), total biomass (kg), and diameter at breast height (dbh; cm). These data cover a range of tree sizes (1.0-115.4 cm) and represent 99 species spread across all 10 species groups used by Jenkins et al. (2003). For prediction, we utilized the most recent cycle (2009-2013) of FIA measurements for Minnesota (N=174,883 across 6,144 plots). We included both adult trees and saplings in this set and filtered the data to remove dead trees.

Model

Our empirical model follows a "component ratio" approach (Chojnacky et al., 2013; Domke et al., 2012) where:`

$$\ln(BM_{total}) = \beta_0 + \beta_1 \ln(dbh) + \varepsilon, \quad [1]$$
$$\ln(FR) = \alpha_0 + \alpha_1 \ln(dbh) + \tau, \quad [2]$$
$$BM_{fol} = BM_{total} * FR, \quad [3]$$

where BM_{total} is total aboveground biomass (kg), *FR* is a foliage component ratio, BM_{fol} is foliage biomass (kg), *dbh* is diameter-at-breast height (cm), and the remaining terms are model parameters. Note that while our target is foliage biomass, the component ratio approach requires fitting a model for total biomass as well. Observed foliage ratios (FR) were calculated as observed foliage biomass (kg)/observed total aboveground biomass (kg).

Model Fitting

Models [1] and [2] were fit to the legacy data using a hierarchical Bayes approach. We used weakly informative normal prior distributions (i.e.,~*N*(0,20)) on the regression coefficients (β_0, β_1, α_0, α_1). Model variances were specified with vague gamma priors (i.e., $\sigma^2 \sim$ Gamma(0.001, 0.001)). In addition, we placed vague "hyper-prior" distributions on the priors of the regression coefficients, allowing the model parameters from all groups to arise from a set of common distributions. Models were fit via Markov chain Monte Carlo (MCMC) methods using Stan, called from R via the RStan package (Stan Development Team, 2014). Our program generated posterior predictions from [1], [2], and [3] simultaneously, which allows us to assess prediction uncertainty in foliage biomass at both tree and plot scales.

Assessing tree-scale uncertainty

We characterized the range of tree-scale uncertainty within our model by performing Bayesian posterior predictive checks (Gelman et al., 1996). We generated 1,000 simulated datasets, of the same dimensions as the legacy data, by taking draws from the posterior predictive distribution, resulting in a marginal posterior distribution for every tree within the dataset. We compared the simulated means, as well as tree-scale 95% uncertainty ranges, to observed foliage biomass from the legacy data.

Application to FIA data

The fitted hierarchical model was then applied to generate posterior predictive distributions, based on 500 simulations, for every tree within the Minnesota FIA data. These were aggregated into plot estimates by multiplying predicted foliage biomass with an adjustment factor to standardize biomass estimates on a per hectare basis and summing this product within plots. This procedure resulted in a distribution of predicted foliage biomass stock (kg*ha^{-1}) at each plot, which we summarized by its mean and 95% uncertainty interval range.

RESULTS

For individual trees within the legacy data, overall mean posterior predicted foliage biomass was 13.3 kg for conifers and 5.5 kg for hardwoods. The corresponding mean uncertainty bounds (95% credible intervals from the posterior simulations) were ±47.12 kg and ±19.44 kg respectively. These uncertainties are large relative to the mean, but for both groups there is much higher error around predicted foliage biomass for large trees than for smaller individuals (Figure 1). In general, uncertainty is higher for conifers than for hardwoods within the legacy data, though hardwoods in these data generally had less foliage biomass.

When applied to predict foliage biomass for FIA data, the fitted models resulted in an overall mean of 3932.2 kg*ha^{-1} across all plots. The large tree-scale uncertainties noted in our first analysis led to considerable error at the plot level, with an average uncertainty interval of 3492.4 kg*ha^{-1}. The distributions of both the plot-scale means and uncertainties are skewed to the left (Figure 2), with most plots predicted to have relatively little foliage biomass, and a smaller number of plots possessing larger stocks with accompanying large error bounds.

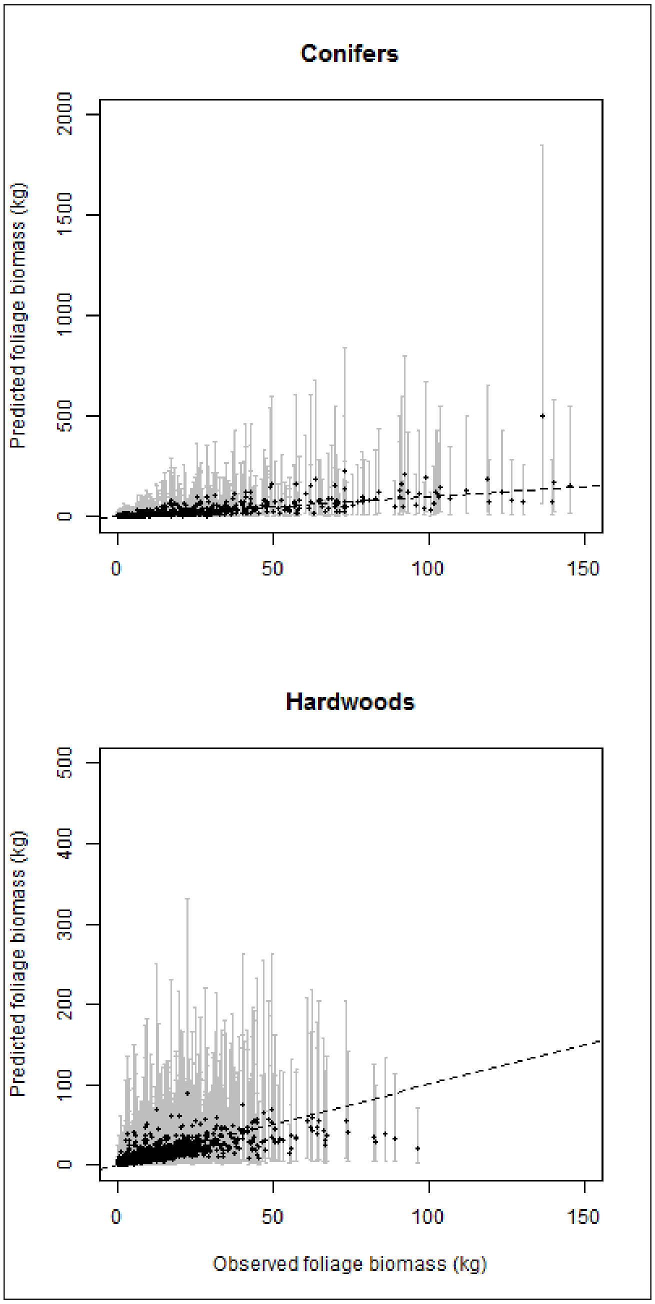

Figure 1—Observed vs. predicted foliage biomass for the legacy data. Error bars represents the 95% uncertainty intervals resulting from the posterior predictive checks we performed.

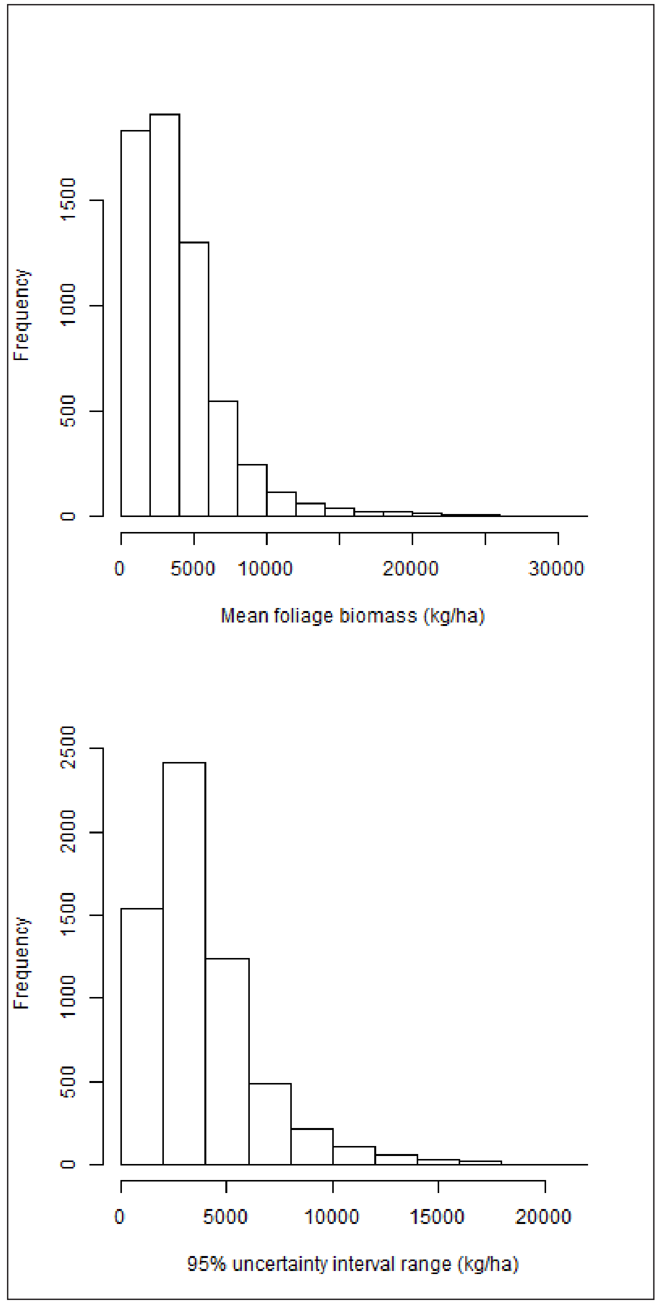

Figure 2—Predicted means and 95% uncertainty intervals of plot-scale foliage biomass (kg*ha-1) for 6,115 FIA plots within Minnesota.

DISCUSSION

Per reporting guidelines for NGHGIs outlined by the United Nations' Framework Convention on Climate Change, the United States is required to provide quantitative estimates of uncertainties surrounding standing forest carbon stocks (IPCC, 2006). In order to best support development of international policy and decision-making, all nations should aim for these error estimates to be both reasonable and all-inclusive. Our results suggest that the current approach, where biomass pools are estimated via equations with fixed parameters, ignores substantial uncertainties associated with allometric functions for estimating foliage biomass.

A hierarchical framework such as we use here provides an ideal approach for capturing this error. While non-hierarchical simulation-based approaches have been proposed (e.g., Wayson et al., 2014)"type" : "article-journal" }, "uris" : ["http://www.mendeley.com/documents/?uuid=8beb78c2-5742-41c0-8282-826aabf09cae"] }], "mendeley" : { "formattedCitation" : "(Wayson et al., 2014, these require *a priori* decisions about the distributions underlying model parameters. In a hierarchical model, the dimensions of these distributions are determined by the fitting data (Green et al., 1999). Further, when a Bayesian approach is employed, uncertainties in both model parameters and data are seamlessly integrated into predicted estimates. Of course fitting a hierarchical model requires felled-tree data, so projects that aim to compile and enhance existing datasets, such as the USFS Volume Biomass Project, are integral to this approach.

While the uncertainties found by our analysis are large, these results do carry some important caveats. First, the legacy data are sparse relative to the study area, and provide varying coverage across species groups. Ongoing work aims to fill gaps in these data, in order to provide a more representative dataset for the whole United States. Second, while we found large prediction errors in the foliage biomass pool, the extent to which this impacts overall uncertainty in the forest carbon pool remains unclear. Future work will assess whether similar error bounds can be expected for other, larger biomass components (i.e. roots, which are similarly dynamic), as well as for total aboveground biomass.

CONCLUSIONS

By using a hierarchical model fit to a large felled-tree database, we reveal large uncertainty from allometric functions for predicting foliage biomass. Given the need for complete and accurate error estimates to support decision making related to the management of greenhouse gas emissions, these results may have important implications for national and international policy related to climate change. A hierarchical approach and the availability of the legacy data were important in uncovering these uncertainties, and we argue that such a framework should be adopted by future NGHGIs. That said more research is required to assess if the scale of uncertainty we found for foliage biomass is particular to this component, or if it will have a large impact on the estimation of the overall forest carbon pool.

ACKNOWLEDGMENTS

We wish to thank the contributors to the USFS Volume Biomass Project for supplying the felled-tree data used in our analysis. In addition we thank Miranda Curzon, Ram Deo, and Ed Green for their comments and feedback on this manuscript.

LITERATURE CITED

Chojnacky, D.C., Heath, L.S., Jenkins, J.C. 2013. Updated generalized biomass equations for North American tree species. Forestry 87(1):129–151.

Domke, G.M., Woodall, C.W., Smith, J.E., Westfall, J. A., McRoberts, R.E. 2012. Consequences of alternative tree-level biomass estimation procedures on U.S. forest carbon stock estimates. Forest Ecology and Management 270:108–116.

Gelman, A.E., Meng, X.-L., Stern, H. 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica 6:733–807.

Green, E.J., MacFarlane, D.W., Valentine, H.T., Strawderman, W.E. 1999. Assessing Uncertainty in a Stand Growth Model by Bayesian Synthesis. Forest Science 45(4): 528-538.

IPCC, 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Volume 4 Agriculture, Forestry, and Other Land Use, Prepared by the National Greenhouse Gas Inventories Programme. IGES, Japan.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. Forest Science 49(1): 12–35.

Stan Development Team 2014. RStan: the R interface to Stan, Version 2.5. Available: http://mc-stan.org/rstan.html. (Accessed November, 2014).

Wayson, C. A., Johnson, K.D., Cole, J. A., Olguín, M.I., Carrillo, O.I., Birdsey, R. A. 2014. Estimating uncertainty of allometric biomass equations with incomplete fit error information using a pseudo-data approach: methods. Annals of Forest Science. In Press.

Weiskittel, A.R., Macfarlane, D.W., Radtke, P.J., Affleck, D.L.R., Temesgen, H., Westfall, J.A., Woodall, C.W., Coulston, J.W. 2015. A call to improve methods for estimating tree biomass for regional and national assessments. Journal of Forestry. In press.

Woodall, C.W., Heath, L.S., Domke, G.M., Nichols, M.C. 2011. Methods and equations for estimating aboveground volume, biomass, and carbon for trees in the U.S. forest inventory, 2010. US Forest Service, Durham, New Hampshire.