

Minimizing Bias in Biomass Allometry: Model Selection and Log-Transformation of Data

Joseph Mascaro^{1,2,3,6}, Creighton M. Litton⁴, R. Flint Hughes⁵, Amanda Uowolo⁵, and Stefan A. Schnitzer^{2,3}

¹ Department of Global Ecology, Carnegie Institution for Science, Stanford, California 96025, U.S.A.

² Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Republic of Panama

³ Department of Biological Sciences, University of Wisconsin – Milwaukee, Milwaukee, Wisconsin 53211, U.S.A.

⁴ Department of Natural Resources and Environmental Management, University of Hawai'i at Mānoa, Honolulu, Hawaii 96822, U.S.A.

⁵ Institute of Pacific Islands Forestry, USDA Forest Service, Hilo, Hawaii 96720, U.S.A.

ABSTRACT

Nonlinear regression is increasingly used to develop allometric equations for forest biomass estimation (i.e., as opposed to the traditional approach of log-transformation followed by linear regression). Most statistical software packages, however, assume additive errors by default, violating a key assumption of allometric theory and possibly producing spurious models. Here, we show that such models may bias stand-level biomass estimates by up to 100 percent in young forests, and we present an alternative nonlinear fitting approach that conforms with allometric theory.

Key words: allometry; Hawai'i; heteroscedasticity; linear regression; nonlinear regression analysis; *Psidium cattleianum*.

FOREST ECOSYSTEMS EXERT A LARGE INFLUENCE ON THE GLOBAL CARBON CYCLE through the flux and storage of carbon in plant biomass, and quantification of these carbon fluxes and stocks depends to a great degree on allometric models used to estimate aboveground tree biomass (Chave *et al.* 2005). The relationship between tree diameter and biomass is highly conserved, with idealized trees exhibiting a power-law relationship: aboveground biomass = $a \times \text{diameter}^b$ where a and b are regression coefficients (Niklas 2006). Traditionally, linear models have been fit to log-transformed diameter and biomass data, but the increasing availability of advanced statistical packages has led to greater use of nonlinear models fit directly to untransformed diameter and biomass data (e.g., Litton *et al.* 2006). This approach may be favored in part because its use avoids the need for transformation (and back-transformation; Baskerville 1972). In practice, however, the default nonlinear technique used by most statistical packages assumes homogeneity of errors, which cannot be safely assumed with most allometry data. For trees, in particular, ideal allometry data are strongly heteroscedastic, exhibiting increasing variation in biomass with increasing diameter (Chave *et al.* 2005). When nonlinear fitting techniques are applied without accounting for heteroscedasticity, the resulting models may include substantial biases even while maintaining high r^2 and low mean square errors.

Here, we contrast linear and nonlinear fitting approaches for estimating aboveground biomass using harvest data from six tree species of various habits, including one species with no previously published allometric model: *Psidium cattleianum* (strawberry guava). *Psidium cattleianum* is native to Brazil and has been widely intro-

duced in Oceania, causing particularly dramatic alterations to the forest ecosystems of Hawai'i where it is considered one of the state's most disruptive introduced plants (Little & Skolmen 1989, Wagner *et al.* 1999). The species has a high potential to alter ecosystem carbon storage by causing changes to forest structure and nutrient dynamics (Hughes & Denslow 2005, Asner *et al.* 2008), and thus its allometric characteristics are of considerable importance.

METHODS

STUDY SPECIES AND HARVEST PROCEDURES.—We reanalyzed five published models for predicting total aboveground biomass, four of which were fit using nonlinear regression by ordinary least squares on untransformed biomass and diameter data (diameter at breast height [dbh] at 1.3 m from the ground, or basal diameter): (1) *Metrosideros polymorpha* (ohi'a); (2) *Diospyros sandwicensis* (lama); (3) *Psydrax odorata* (alaha'e); and (4) *Dodonaea viscosa* (a'ali'i) (Litton *et al.* 2006, Litton & Kauffman 2008), and a fifth using linear regression by ordinary least squares on ln-transformed biomass and dbh data: (5) *Rhamnus cathartica* (common buckthorn) (Mascaro & Schnitzer 2011). We also analyzed one unpublished dataset for (6) *P. cattleianum* (strawberry guava). The species selected grow as shrubs (1, 4–6), small trees (1–6), or large trees (1–2), and occur across a variety of habitats, including tropical dry forests (1–4), tropical mesic and wet forests (1–4, 6), and temperate to subtropical forests (5).

For each dataset, 15–34 trees for each species were cut at ground level and all aboveground parts (i.e., wood, twigs, leaves, fruit) were separated and weighed with spring scales (Table 1). Subsamples of each tissue type were collected and dried to constant mass in a forced

Received 6 January 2011; revision accepted 19 April 2011.

⁶Corresponding author; e-mail: jmascaro@stanford.edu

© 2011 The Author(s)

Journal compilation © 2011 by The Association for Tropical Biology and Conservation

air oven at 60°–90°C (temperatures were consistent within a given species) to correct for moisture content and determine the total aboveground dry weight of each tree (Appendix S1). *Psidium cattleianum* individuals were harvested from the Laupahoehoe Unit of the Hilo Forest Reserve on Hawai'i Island in 2008. Individuals with branching below 1.3 m on the main stem were excluded, and thus the proposed model here should be applied to shrub-form *Psidium* with caution. Details on the published models for the other five species are available in the original publications (Litton *et al.* 2006, Litton & Kauffman 2008, Mascaro & Schnitzer 2011).

STATISTICAL ANALYSES.—We used three techniques to fit allometric models to biomass and diameter data. First, we applied the traditional allometric approach (*sensu* Baskerville 1972) by fitting a linear model to each dataset of the form

$$\ln(y) = \ln(a) + b \times \ln(x) + \varepsilon \quad (1a)$$

where y is the dependent variable (total aboveground biomass), x is the independent variable (diameter), a and b are regression coefficients and ε is regression error. Each linear model was then back-transformed to a power function of the form:

$$y = ax^b \times CF \quad (1b)$$

where CF is a correction factor computed as

$$CF = e^{(MSE/2)} \quad (1c)$$

and MSE is the mean square error of the regression. The CF accounts for the back-transformation of the regression error, and is a requisite step in the use of linear models and ln-transformed data in allometry (Baskerville 1972).

Second, we applied nonlinear regression to fit a power model to each dataset of the form

$$y = ax^b + \varepsilon \quad (2)$$

In the case of four of the five published models, these fits served to replicate the published results of Litton and Kauffman (2008) and Litton *et al.* (2006). Additionally, model 2 replicates the 'default' power-law fit for commonly used statistical packages (*e.g.*, SigmaPlot, JMP, and R – specifically the nlm function).

Finally, we applied a variant of model 2 with an additional term to account for heteroscedasticity that is common in allometric models. Specifically, we allowed regression error to scale as a function of tree diameter by introducing an additional constant (k)

$$y = ax^b + x^k \times \varepsilon \quad (3)$$

All model parameters were estimated using maximum likelihood estimation in SAS (2008). For comparative purposes, r^2 and MSE were assessed in the power-law form for all models (*i.e.*, equation 1b for model 1). Corrected (*i.e.*, for sample size) Akaike Information Criterion scores (AICc) were assessed for all three models (follows Draper & Smith 1998).

RESULTS AND DISCUSSION

All three model fits were significant for all six species (*i.e.*, $r^2 > 0.73$, $P < 0.0001$) (Table S1; Fig. 1). For all species, however, very high Δ AICc scores suggested that model 2 was the least effective of the three in predicting biomass (Table S1). In four of the six species, model 2 introduced a substantial and consistent bias for smaller diameter stems compared with harvested biomass. For *M. polymorpha*, the error averaged 223 percent for individuals < 8 cm dbh, and reached 400 percent error within that span (Fig. 2A). Similar biases were found for *D. viscosa* (80% for stems < 1.5 cm basal diameter), *P. cattleianum* (51% for stems < 14 cm dbh), and *R. cathartica* (74% for stems < 18 cm dbh). Thus, although model 2 is the ideal form for tree allometry (Niklas 2006), fitting this model without controlling for variant error structure over the range of diameters sampled (a near ubiquitous feature of allometric datasets) may result in spurious equations that are not applicable across the entire size range used in generating the models.

The equation biases described almost exclusively affect the smaller stems in the datasets. However, in practice we have found that such biases can result in substantial error in estimating plot-level aboveground biomass in forests that contain many small trees (*e.g.*, young, regenerating stands), and early successional stands dominated by small diameter individuals are an increasingly important component of forested landscapes across the globe (Swanson *et al.* 2011). Importantly, three of the species for which the biases were detected are capable of forming large, monotypic stands

TABLE 1. Summary of harvest data for six tree allometry datasets. Diameter (cm) is diameter at breast height (dbh, 1.3 m from the ground), except for *Dodonaea viscosa*, which is basal diameter.

Species	N	Minimum diameter	Maximum diameter	References
<i>Metrosideros polymorpha</i>	30	0.3	33.1	Litton and Kauffman (2008)
<i>Diospyros sandwicensis</i>	25	1.8	20.2	Litton <i>et al.</i> (2006)
<i>Psychrax odorata</i>	34	0.5	4.6	Litton <i>et al.</i> (2006)
<i>Dodonaea viscosa</i>	20	0.5 ^a	2.9 ^a	Litton and Kauffman (2008)
<i>Psidium cattleianum</i>	26	0.2	18.2	This study
<i>Rhamnus cathartica</i>	15	1.2	24.7	Mascaro and Schnitzer (2011)

^aBasal diameter (cm).

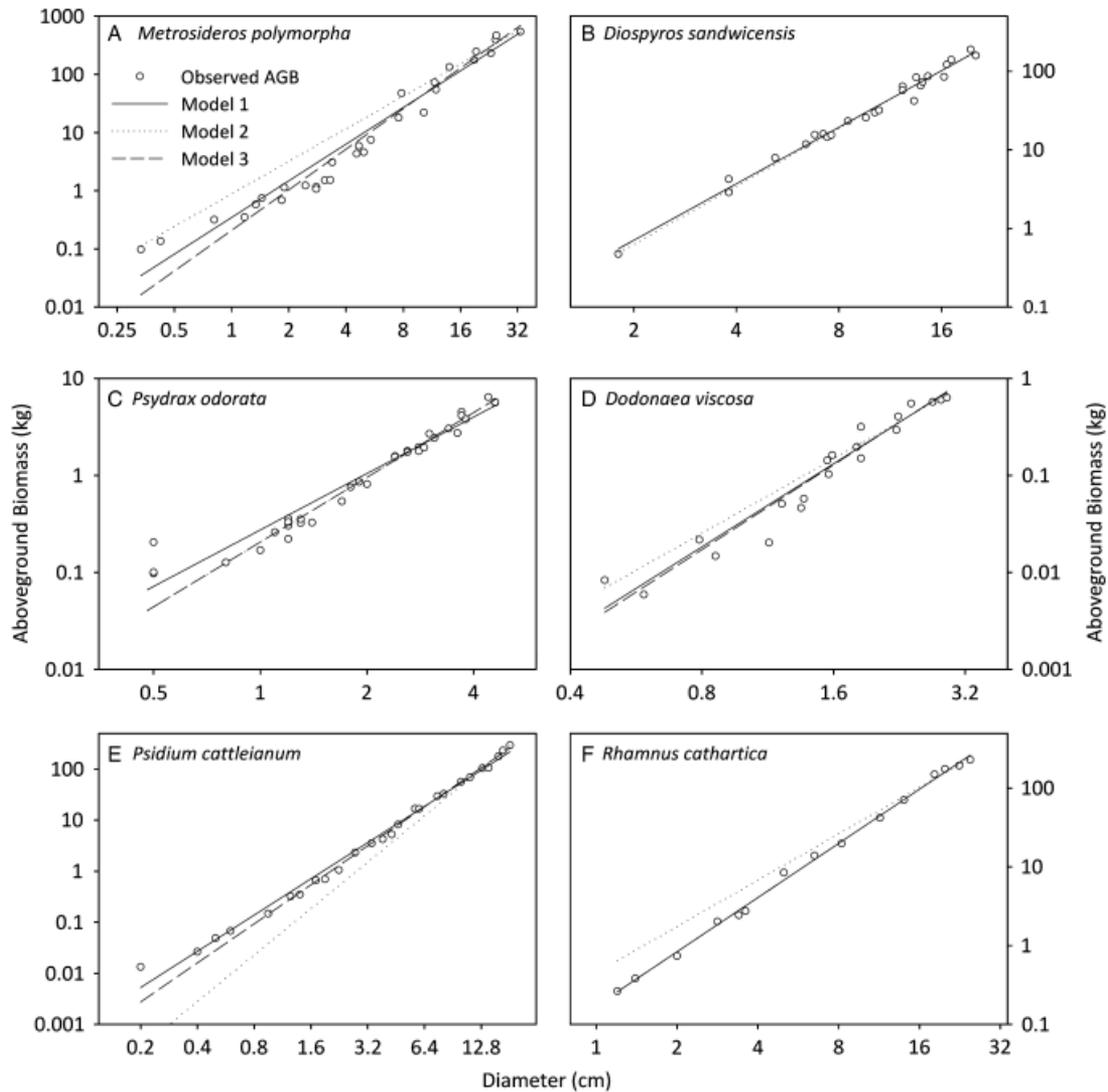


FIGURE 1. Allometric models predicting aboveground biomass based on stem diameter for six woody species. Diameter (cm) is diameter at breast height (dbh, 1.3 m from the ground), except for *Dodonaea viscosa*, which is basal diameter.

of small individuals with very high stem densities (Mueller-Dombois & Fosberg 1998, Mascaro & Schnitzer 2007, Asner *et al.* 2008). We applied all three *M. polymorpha* models to plot inventory data from young, open forests on Hawai'i Island (stand density = 1670 stems/ha; dbh range = 2–12 cm; mean dbh = 3.8 cm; trees in stand > 8 cm dbh = 5%; J. Mascaro & F. Hughes unpubl. data), and found that use of model 2 resulted in a ~100 percent overestimation of stand-level aboveground biomass compared to the other models. However, when applied to an older stand with trees of varying size (stand density = 1140 stems/ha; dbh range = 3.2–32.7 cm; mean dbh = 15.4 cm; trees in stand > 8 cm dbh = 93%; C.M. Litton & J.B. Kauffman unpubl. data), all three models produced similar stand totals, highlighting that the bias outlined here becomes less important as mean tree size increases.

To fit an appropriate model, a multiplicative—rather than additive—error term is required. The simplest way to account for multiplicative errors is to perform a log-transformation on the

diameter and biomass data, thus normalizing the biomass error structure along the range of diameter values as in model 1 (Baskerville 1972). Note that the additive error term (ϵ) in model 1 (eq 1a) represents multiplicative errors on the original scale. An alternative can be accomplished by introducing an additional regression constant that allows errors to scale with diameter (*i.e.*, model 3). Here, model 3 performed better than model 1 (AICc) for three of six species examined, and for two of these the Δ AICc scores from model 3 to 1 were > 10. By contrast, for the three species where model 1 performed better, the Δ AICc from model 1 to 3 was always < 5. Thus, the few examples considered here suggest that model 3 may be more reliable, although we emphasize that models 1 and 3 were very similar overall.

Our results highlight that the decision to log-transform raw data in allometry is more than one of statistical convenience. Packard and Boardman (2008) argued that fitting linear models on log-transformed data leads to results that are 'biased and misleading'

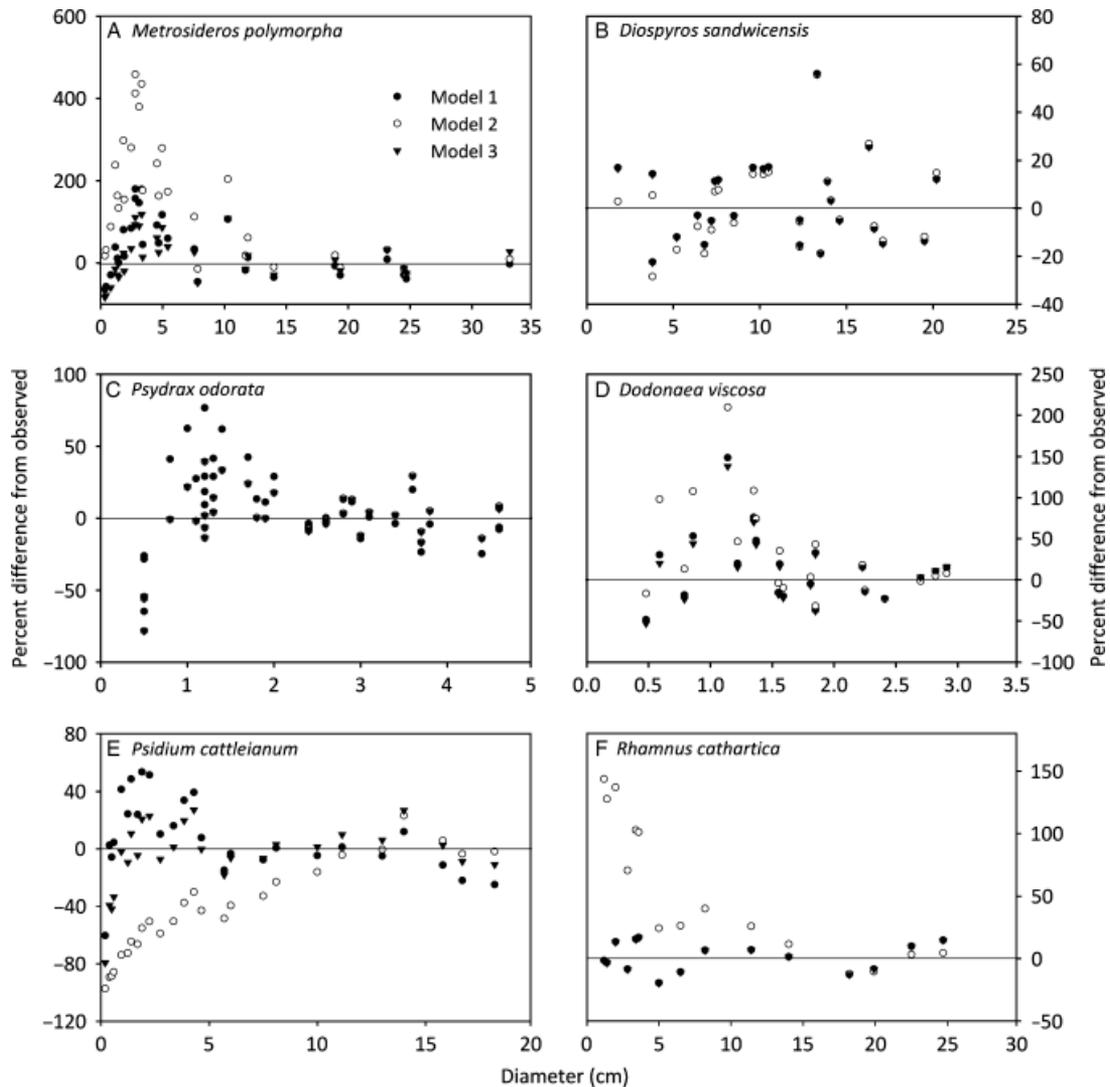


FIGURE 2. Percent error in allometric models predicting aboveground biomass based on stem diameter for six woody species. Diameter (cm) is diameter at breast height (dbh, 1.3 m from the ground), except for *Dodonaea viscosa*, which is basal diameter.

because such models operate in geometric rather than arithmetic space, and that analyses should be performed on the original scale. However, Kerkhoff and Enquist (2009) note that many allometric characteristics of organisms are ‘multiplicative by nature’ and thus fitting models to log-transformed data is perfectly acceptable because accounting for proportional rather than absolute variation is most important. Our results support the latter view. Models that assumed multiplicative errors (1 and 3) described the data very well, although the results of model 3 indicate that log-transformation is not the only way to satisfy this assumption.

In conclusion, fitting nonlinear biomass allometry models while assuming additive errors can produce systematic biases in estimates for smaller diameter trees. This bias can lead to large errors in landscape-scale biomass estimation from stand-level datasets that are dominated by small trees. In light of these results, we believe that added care should be used to determine whether an allometric model is appropriate for the error structure of the data. As a solution, we propose that

published allometric models should be accompanied by: (1) the raw data used to generate the regression equations (*e.g.*, as an appendix); and (2) plots of the residuals. Such an approach would allow individual investigators to judge whether the model is adequate for their needs as published, or if alternative approaches are more appropriate for estimating biomass for a given tree or stand.

ACKNOWLEDGMENTS

Jim Baldwin (USFS-PSW) provided assistance fitting models. J. Mascaro was supported by a grant from Applied Ecological Services Inc., an NSF Graduate Research Fellowship, and a UWM Golda Meir Library Scholar Award. Additional support was provided by a grant from the National Science Foundation to C.M. Litton (DEB-0816486). Comments from four anonymous reviewers improved this manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

TABLE S1. *Model parameters for six woody species.*

TABLE S2. *Harvest data for *Metrosideros polymorpha* (ohi'a).*

TABLE S3. *Harvest data for *Diospyros sandwicensis* (lama).*

TABLE S4. *Harvest data for *Psydrax odorata* (alaha'e).*

TABLE S5. *Harvest data for *Dodonaea viscosa* (a'ali'i).*

TABLE S6. *Harvest data for *Psidium cattleianum* (strawberry guava).*

TABLE S7. *Harvest data for *Rhamnus cathartica* (common buckthorn).*

APPENDIX S1. Primary harvest data for total aboveground biomass.

Please note: Wiley-Blackwell Publishing are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

LITERATURE CITED

- ASNER, G. P., R. F. HUGHES, P. M. VITOUSEK, D. E. KNAPP, T. KENNEDY-BOWDOIN, J. BOARDMAN, R. E. MARTIN, M. EASTWOOD, AND R. O. GREEN. 2008. Invasive plants transform the three-dimensional structure of rain forests. *Proc. Natl. Acad. Sci. USA* 105: 4519–4523.
- BASKERVILLE, G. 1972. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* 2: 49–53.
- CHAVE, J., C. ANDALO, S. BROWN, M. A. CAIRNS, J. Q. CHAMBERS, D. EAMUS, H. FOLSTER, F. FROMARD, N. HIGUCHI, T. KIRA, J.-P. LESQUIRE, B. W. NELSON, H. OGAWA, H. PUIG, B. RIERA, AND T. YAMAKURA. 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145: 87–99.
- DRAPER, N. R., AND H. SMITH. 1998. *Applied regression analysis* (3rd edition). Wiley-Interscience, New York, NY.
- HUGHES, R. F., AND J. S. DENSLow. 2005. Invasion by a N₂-fixing tree, *Falcataria moluccana*, alters function, composition, and structure in wet lowland forests of Hawai'i. *Ecol. Appl.* 15: 1615–1628.
- KERKHOFF, A. J., AND B. J. ENQUIST. 2009. Multiplicative by nature: Why logarithmic transformation is necessary in allometry. *J. Theor. Biol.* 257: 519–521.
- LITTLE, E. L. JR., AND R. G. SKOLMEN. 1989. *Common forest trees of Hawaii* (native and introduced). U.S. Department of Agriculture, Washington, DC.
- LITTON, C. M., AND J. B. KAUFFMAN. 2008. Allometric models for predicting aboveground biomass in two widespread woody plants in Hawaii. *Biotropica* 40: 313–320.
- LITTON, C. M., D. R. SANDQUIST, AND S. CORDELL. 2006. Effects of non-native grass invasion on aboveground carbon pools and tree population structure in a tropical dry forest of Hawaii. *For. Ecol. Manage.* 231: 105–113.
- MASCARO, J., AND S. A. SCHNITZER. 2007. *Rhamnus cathartica* L. (common buckthorn) as an ecosystem dominant in Southern Wisconsin forests. *Northeastern Nat.* 14: 387–402.
- MASCARO, J., AND S. A. SCHNITZER. 2011. Dominance by the introduced tree *Rhamnus cathartica* (common buckthorn) may limit aboveground carbon storage in Southern Wisconsin forests. *For. Ecol. Manage.* 261: 545–550.
- MUELLER-DOMBOIS, D., AND F. R. FOSBERG. 1998. *Vegetation of the tropical Pacific islands*. Springer, New York, NY.
- NIKLAS, K. J. 2006. A phyletic perspective on the allometry of plant biomass-partitioning patterns and functionally equivalent organ-categories. *New Phytol.* 171: 27–40.
- PACKARD, G. C., AND T. J. BOARDMAN. 2008. Model selection and logarithmic transformation in allometric analysis. *Physiol. Biochem. Zool.* 81: 496–507.
- SAS. 2008. SAS version 9.2. SAS Institute Inc., Cary, NC.
- SWANSON, M. E., J. F. FRANKLIN, R. L. BESCHTA, C. M. CRISAFULLI, D. A. DELLA SALA, R. L. HUTTO, D. B. LINDENMAYER, AND F. J. SWANSON. 2011. The forgotten stage of forest succession: Early-successional ecosystems on forest sites. *Front. Ecol. Environ.* 9: 117–125.
- WAGNER, W. L., D. R. HERBST, AND S. H. SOHMER. 1999. *Manual of the flowering plants of Hawai'i*. University of Hawai'i Press/Bishop Museum Press, Honolulu, HI.