Optimizing efficiency of height modeling for extensive forest inventories

T.M. Barrett

Abstract: Although critical to monitoring forest ecosystems, inventories are expensive. This paper presents a generalizable method for using an integer programming model to examine tradeoffs between cost and estimation error for alternative measurement strategies in forest inventories. The method is applied to an example problem of choosing alternative height-modeling strategies for 1389 plots inventoried by field crews traveling within an 82.5×10^6 ha region of the west coast of North America during one field season. In the first part of the application, nonlinear regional height models were constructed for 38 common species using a development data set of 137 374 measured tree heights, with root mean square error varying from 6.7 to 2.1 m. In the second part of the application, alternative measurement strategies were examined using a minimal cost objective subject to constraints on travel time and estimation error. Reduced travel time for field crews can be a significant portion of the cost savings from modeling tree heights. The optimization model was used to identify a height-modeling strategy that, given assumptions made, resulted in <10% of maximum average plot volume error, >33% of potential measurement cost savings, and small bias for estimates of regional volume and associated sampling error (0.1% and 0.4%, respectively).

Résumé : Les inventaires de suivi des écosystèmes forestiers sont coûteux, mais essentiels. Cet article présente une méthode qui peut être généralisée pour l'utilisation d'un modèle de programmation linéaire en nombres entiers conçu pour examiner les compromis entre le coût et l'erreur d'estimation de stratégies alternatives d'inventaire forestier. La méthode est appliquée à un problème typique de choix de stratégies alternatives pour modéliser la hauteur des arbres dans 1 389 placettes inventoriées par une équipe de terrain qui se déplace à l'intérieur d'une région de $82,5 \times 10^6$ ha sur la côte ouest de l'Amérique du Nord durant une saison de terrain. Dans la première partie de l'application, une base de données contenant 137 374 mesures de hauteur d'arbre, avec une erreur quadratique moyenne variant de 6,7 à 2,1 m selon l'essence, a été utilisée pour calibrer les modèles régionaux de hauteur pour 38 essences communes. Dans la deuxième partie de l'application, des stratégies alternatives de mesure ont été examinées en fonction d'un objectif de coût minimal sujet aux contraintes de temps de déplacement et d'erreur d'estimation. La réduction du temps de déplacement de l'équipe d'inventaire peut représenter une partie importante de la réduction du coût pour la modélisation de la hauteur des arbres. Compte tenu des hypothèses stipulées, le modèle d'optimisation a été utilisé pour identifier une stratégie de modélisation de la hauteur qui permet d'obtenir une erreur moyenne maximale inférieure à 10 % par placette pour le volume de bois, une réduction potentielle de plus de 33 % du coût des mesures ainsi qu'un faible biais pour les estimations du volume de bois à l'échelle régionale associé à une faible erreur d'échantillonnage (respectivement 0,1 % et 0,4 %).

[Traduit par la Rédaction]

Introduction

Forest inventories are a necessary but expensive part of forest assessment, monitoring, planning, and management. Using modeled tree heights in lieu of measured tree heights is a frequently used technique for reducing costs in inventories, and numerous height–diameter regression models have been proposed and used (Avery and Burkhart 2002). In this paper, a mixed integer optimization model is used to examine accuracy and efficiency tradeoffs that occur for different height-modeling strategies applied to 1389 plots measured during one field season in an 82.5×10^6 ha region.

Received 31 August 2005 Accepted 12 May 2006. Published on the NRC Research Press Web site at http://cjfr.nrc.ca on 7 September 2006.

T.M. Barrett. Pacific Northwest Research Station, Forest Service, 3301 C Street, Suite 200, Anchorage, AK 99503 USA (e-mail: tbarrett@fs.fed.us).

When height models are used, a common approach is to apply models with fixed parameters for individual species or species groups (e.g., Curtis 1967; Hanus et al. 1999; Huang et al. 1992; Huang et al. 2000; Larsen and Hann 1987; Parresol 1992). Numerous linear and nonlinear forms have been examined for these regional height models, with many forms comparable in terms of model fit (Huang et al. 1992; Zhang 1997).

Regardless of model form, least squares methods seek to minimize error in predicted tree height. However, for most forestry applications, tree height is merely an intermediate step in estimating attributes of interest, such as area by stand structural type or canopy cover class, tree biomass, or timber volume. Using a correctly fitted model to make predictions that are then input to a second model can be problematic. In one study examining this problem, modeled heights increased root mean square error (RMSE) of individual tree volume prediction by roughly 35–38% (Gregoire and Williams 1992). In recognition of this issue, a few height-model

 doi:10.1139/X06-128

for this inventory (e.g., a separator for managed and unmanaged stands) or have large numbers of missing observations (e.g., site index). Initial tests using ecoregion-specific models, plot density, site productivity, or crown ratio resulted in only modest improvement of height prediction. Plot-level approaches showed some promise for managed single-species, even-age, single-stand plots (Monleon 2003), but most plots in this inventory do not fit that profile. In addition, there are advantages to using a method that can be easily understood and replicated by the many users of the forest inventory data. Finally, past experience with inventories designed for long-term monitoring had indicated that simple models were more likely to have smooth transitions as personnel and programming languages change over the decades.

The tree records consisted of trees of at least 10 cm diameter at breast height (DBH, 1.3 m) with field measured heights and diameter. Records were screened to remove any tree records where field crew had recorded broken tree tops. Models were only created for species with at least 250 tree observations. Records for 137 374 trees from 38 species remained. A variety of nonlinear model forms was evaluated, with Richard's model (Richards 1959) chosen based on good performance both for this data set and in previous research applications. This model form is

$$[9] \qquad H = 1.37 + \alpha(1 - e^{-\beta D})^{\gamma}$$

where H is total tree height in metres, D is DBH in centimetres, and α, β, and γ are species-specific parameters.

Nonlinear fitting was done with NLIN procedure of SAS version 8.02 (SAS Institute Inc. 2001) using ordinary nonlinear least squares. The variance of error terms did increase with diameter; however, as other researchers have found in developing regional height models, the increase was not substantial enough to require weighting (Huang et al. 1992, Zhang 1997). Models for each species were evaluated using RMSE and R^2.

For this study, R^2 statistics were calculated as

$$[10] \qquad R^2 = 1 - \frac{\sum_k (H_k - \hat{H}_k)^2}{\sum_k (H_k - \overline{H})^2}$$

where H_k is observed height for tree k, $\hat{H}_k$ is modeled height, and $\overline{H}$ is the average height for the species.

For use with the optimization problem, height models were applied to an application data set separate from the development data set. The application data set included 34 325 trees with DBH $\geq$ 12.5 cm that had been measured in California, Oregon, and Washington during the 2002 field season. Although some trees appeared in both the development and application data sets, different plot selection and sampling intensity, switches from variable radius to fixed-area tree sampling, a change in plot design from five or four subplots, and 2–14 years of growth and mortality result in essentially independent data sets.

In the application data set, there were 46 uncommon species with insufficient observations to build robust height models. The trees of these species made up less than 3% of the total inventory data. Each of these uncommon species was assigned a height model of a similar species. Measured heights were retained for trees where field crew had recorded broken tops, under the assumption that modeled heights would never be appropriate for such trees. Measurement error of heights does occur; however, for this application, field measured heights were taken as truth.

Individual-tree volume was calculated with direct-volume equations used by the FIA program for this region. For both height and volume, model fit to the application data set was evaluated by R^2, mean bias, and mean absolute deviation. Height model prediction accuracy for large trees is a concern, so bias is also reported for the 80–100% DBH quantile. Trees with broken tops were excluded from reported results.

Results

When nonlinear optimization is used with Richard's height model (eq. 9) and small data sets, it is possible to have difficulty finding the global optimum. In this case, with at least 250 observations for any species and prior information on initialization values for parameters from previously published literature (Garman et al. 1995; Huang et al. 1992; Zhang 1997), solutions were easily found with the Gauss–Newton method in the SAS NLIN procedure. As expected because of the high heterogeneity of samples from such an extensive geographic region, the R^2 values were low compared with many previously published height models (Table 1). The R^2 values were higher for conifers than hardwoods. Among conifers, those species with highly variable tree form (lodgepole pine, *Pinus contorta* Dougl. ex Loud., and western juniper, *Juniperus occidentalis* Hook.) had lower R^2 values. RMSE values were generally greater for species that can obtain substantial heights (e.g., coastal redwood, *Sequoia sempervirens* (D. Don) Endl.; Douglas-fir, *Pseudotsuga menziesii* (Mirb.) Franco; and Sitka spruce, *Picea sitchensis* (Bong.) Carr.) than for species that cannot (e.g., blue oak, *Quercus douglasii* Hook. & Arn.; incense-cedar, *Calocedrus decurrens* (Torrey) Florin; and western juniper).

When applied to the test data set, modeled heights were again generally more accurate for conifers than for hardwoods (Table 2). The five species with the most samples (Douglas-fir, ponderosa pine, *Pinus ponderosa* Dougl. ex P. Laws. & C. Laws.; western hemlock, *Tsuga heterophylla* (Raf.) Sarg.; white fir, *Abies concolor* (Gord. & Glend.) Lindl.; and lodgepole pine) had relatively small bias of mean height ranging from –0.34 to 0.32 m. For these same species, average absolute errors ranged from 3.17 to 4.41 m, which was 16% to 22% of average tree height. In general, large average height bias of ≥1 m occurred only with small sample sizes, of a few hundred trees or less.

The R^2 values for individual tree volume were considerably higher than corresponding R^2 for tree heights. For the five most common species (Douglas-fir, ponderosa pine, western hemlock, white fir, and lodgepole pine), average volume bias ranged from –0.004 to 0.073 m^3. Average absolute errors for these species ranged from 0.13 to 0.42 m^3, which was 14–19% of mean tree volume. Mean bias and mean absolute deviation for both volume and height were considerably larger for trees in the 80–100% DBH quantile (Table 2). For example, mean Douglas-fir height modeling error was 4.4 m, but increased to an mean height modeling error of 7.3 m for trees within the 80–100% DBH quantile.

Table 1. Number of trees, Richards height model parameters ($H = 1.37 + \alpha(1 - e^{-\beta D})^{\gamma}$), root mean square error (RMSE), and R^2 for 38 common species; 1988–2000 inventory data from California, Oregon, and Washington.

		Parameter				
Species	No. of trees	α	β	γ	RMSE	R^2
Pacific silver fir (*Abies amabilis* (Dougl. Forbes)	4 682	54.74	0.0202	1.3666	4.80	0.83
White fir (*Abies concolor* (Gord. & Glend.) Lidl.)	8 308	51.11	0.0150	1.1510	5.46	0.76
Grand fir (*Abies grandis* (Dougl.) Lindl.)	4 115	44.40	0.0226	1.2217	4.64	0.76
Subalpine fir (*Abies lasiocarpa* (Hook.) Nutt.)	2 239	28.55	0.0492	1.8173	4.27	0.67
Red fir (*Abies magnifica* A. Murr.)	2 790	45.60	0.1788	1.3217	5.34	0.79
Noble fir (*Abies procera* Rehd.)	807	65.53	0.0132	1.1224	5.55	0.80
Bigleaf maple (*Acer macrophyllum* Pursh.)	1 463	28.99	0.0371	0.9646	5.30	0.48
Red alder (*Alnus rubra* Bong.)	3 430	28.26	0.0590	1.3040	4.72	0.45
Pacific madrone (*Arbutus menziesii* Pursh)	1 762	23.12	0.0276	0.8787	3.84	0.51
Incense-cedar (*Calocedrus decurrens* (Torr.) Florin)	3 336	53.97	0.0101	1.0974	4.21	0.81
Port-Orford-cedar (*Chamaecyparis lawsoniana* (A. Murr.) Parl.)	257	54.51	0.0140	1.1258	5.02	0.80
Alaska cedar (*Chamaecyparis nootkatensis* (D. Don) Spach.)	256	42.81	0.0210	1.3220	3.87	0.80
Western juniper (*Juniperus occidentalis* Hook.)	1 122	16.25	0.0133	0.7400	2.86	0.44
Western larch (*Larix occidentalis* Nutt.)	1 977	45.56	0.0226	0.9050	4.44	0.72
Tanoak (*Lithocarpus densiflorus* (Hook. & Arn.) Rehd.)	1 451	32.17	0.0193	0.9017	3.47	0.70
Engelmann spruce (*Picea engelmannii* Parry ex Engelm.)	2 239	42.44	0.0269	1.2909	4.68	0.77
Sitka spruce (*Picea sitchensis* (Bong.) Carr.)	705	49.62	0.0163	0.9973	6.75	0.64
Whitebark pine (*Pinus albicaulis* Engelm.)	326	14.68	0.0658	1.8074	3.54	0.46
Lodgepole pine (*Pinus contorta* Dougl. ex Loud.)	5 542	22.78	0.0760	1.8067	4.57	0.47
Jeffrey pine (*Pinus jeffreyi* Grev. & Balf.)	3 478	49.48	0.0105	1.1339	4.33	0.79
Sugar pine (*Pinus lambertiana* Dougl.)	1 928	77.70	0.0066	0.9542	5.95	0.80
Singleleaf pinyon (*Pinus monophylla* Torr. & Frem.)	368	12.66	0.0414	1.6531	2.14	0.58
Western white pine (*Pinus monticola* Dougl.)	915	29.98	0.0461	1.7471	6.73	0.56
Ponderosa pine (*Pinus ponderosa* Dougl. ex P. Laws. & C. Laws.)	16 347	58.81	0.0123	1.1766	4.56	0.81
Ghost pine (*Pinus sabiniana* Dougl. ex D. Laws)[a]	378	41.28	0.0117	1.0008	3.97	0.67
Black cottonwood (*Populus balsamifera* L.)	277	45.99	0.0203	0.9477	7.08	0.57
Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco)	41 231	62.57	0.0131	0.9936	6.26	0.75
Coast live oak (*Quercus agrifolia* Nee)	423	23.15	0.0087	0.6139	3.03	0.45
Canyon live oak (*Quercus chrysolepis* Liebm.)	2 131	19.12	0.0296	1.1314	3.19	0.60
Blue oak (*Quercus douglasii* Hook. & Arn.)	828	13.47	0.0282	0.8303	2.29	0.48
Oregon white oak (*Quercus garryana* Dougl.)	1 029	19.60	0.0467	1.4919	4.27	0.49
California black oak (*Quercus kelloggii* Newb.)	2 797	27.93	0.0129	0.6968	4.14	0.52
Interior live oak (*Quercus wislizeni* A. DC.)	376	16.85	0.0242	0.8143	2.66	0.51
Redwood (*Sequoia sempervirens* (D. Don) Endl.)	992	89.17	0.0056	0.8684	6.07	0.77
Western redcedar (*Thuja plicata* Donn.)	3 221	60.58	0.0089	0.8721	4.61	0.80
Western hemlock (*Tsuga heterophylla* (Raf.) Sarg.)	10 297	53.85	0.0201	1.1832	5.20	0.79
Mountain hemlock (*Tsuga mertensiana* (Bong.) Carr.)	3 287	35.24	0.0306	1.6109	4.18	0.78
California laurel (*Umbellularia californica* (Hook. & Arn.) Nutt.)	264	22.90	0.0355	1.0353	3.84	0.55

[a]Data included 122 Coulter pine (*Pinus coulteri* D. Don) observations.

Optimization for 2002 field season

Methods

In 2002, the inventory of California, Oregon, and Washington consisted of 3482 plots. Of these, 1389 plots contained at least one tree with DBH greater than 12.5 cm and had been visited by an inventory crew during that year. The remaining plots were excluded from this application. Bias coefficients (b_{ij}) were calculated as

$$[11] \quad b_{ij} = F_j \sum_k tpa_k [v_k(D_k, \hat{H}_k) - v_k(D_k, H_k)]$$

for k in the set of trees modeled on plot j when measurement strategy i is implemented, where tpa_k is a tree expansion factor that depends on the plot size used to measure tree k. F_j is a plot expansion value that expands the plot observation to the population level (for details, see appendices of Bechtold and Patterson (2005)), $v_k(\cdot)$ is the function to calculate tree volume for tree k, D is tree DBH, H is measured tree height, and $\hat{H}$ is modeled tree height.

Travel and measurement times for plots were not collected in the field, so values were based on plot characteristics. Each crew of three people travel from location to location throughout the field season, measuring all accessible plots within the vicinity of their lodging before moving to a new location. Travel time (v_j) was thus based on distance between the plot and the nearest town, with additional time based on recorded values for the distance to the nearest improved road and the type of access to the plot (road, unimproved road, human access trail, or no road or trail within 1.6 km). It was assumed that the crew could locate the plot an average of

Table 2. Height and volume fit of height models (Table 1) for trees measured during 2002 in California, Oregon, and Washington.

		Height of all trees				Height of large trees (m)[a]		Volume of all trees[b]				Volume of large trees (m^3)[b]	
Species	N	R^2	Mean height (m)	Mean bias (m)	AAD (m)	Mean bias	AAD	R^2	Mean volume (m^3)	Mean bias (m^3)	AAD (m^3)	Mean bias	AAD
Pacific silver fir	926	0.86	18.78	–0.48	2.97	–1.06	5.05	0.97	1.60	–0.02	0.23	–0.08	0.91
White fir	2308	0.84	20.38	0.14	3.31	1.51	5.55	0.95	1.92	0.07	0.30	0.34	1.06
Grand fir	800	0.79	19.10	0.12	3.51	–0.36	4.73	0.96	1.18	0.03	0.18	0.11	0.65
Subalpine fir	409	0.54	14.03	–0.72	2.88	–1.90	3.79	0.91	0.34	–0.03	0.07	–0.13	0.22
Red fir	843	0.81	23.16	1.10	3.98	–0.17	5.78	0.93	3.45	0.11	0.65	0.03	1.88
Noble fir	107	0.85	26.68	–2.03	3.78	–4.22	6.10	0.96	4.20	–0.35	0.61	–1.05	1.88
Bigleaf maple	472	0.40	17.96	–0.47	3.35	–1.60	3.74	0.99	0.65	–0.02	0.08	–0.10	0.25
Red alder	1067	0.48	19.08	–1.00	3.50	–0.27	4.04	0.94	0.49	–0.01	0.09	0.00	0.23
Pacific madrone	498	0.42	13.60	–0.34	2.91	–1.51	4.08	0.88	0.63	–0.03	0.13	–0.17	0.46
Incense-cedar	947	0.85	16.30	–0.24	2.81	–1.21	5.31	0.95	1.60	–0.03	0.24	–0.16	1.00
Port-Orford-cedar	38	0.81	21.66	3.35	4.29	5.42	5.42	0.99	1.50	0.17	0.18	0.68	0.68
Alaska cedar	65	0.81	16.10	0.58	2.60	–0.89	3.36	0.96	0.83	–0.01	0.11	–0.05	0.41
Western juniper	510	0.35	8.52	0.28	2.08	–0.72	3.26	0.85	0.31	0.00	0.07	–0.03	0.29
Western larch	288	0.67	24.57	0.39	3.17	–1.74	3.62	0.97	0.98	–0.03	0.11	–0.20	0.39
Tanoak	825	0.57	15.64	1.26	3.08	1.50	4.17	0.95	0.55	0.05	0.11	0.19	0.42
Engelmann spruce	263	0.79	22.10	0.47	3.45	1.22	5.24	0.96	1.65	0.05	0.23	0.25	0.86
Sitka spruce	139	0.71	25.71	0.75	4.34	–1.20	3.75	0.98	2.99	–0.07	0.39	–0.34	1.09
Whitebark pine	104	0.10	9.95	–0.88	3.13	–2.50	3.87	0.91	0.27	–0.03	0.08	–0.12	0.21
Lodgepole pine	1826	0.44	16.15	–0.34	3.54	0.01	4.32	0.95	0.67	0.01	0.13	0.10	0.48
Jeffrey pine	365	0.81	19.29	1.18	3.63	2.73	5.37	0.95	3.17	0.22	0.52	0.80	1.62
Sugar pine	320	0.82	26.56	–0.03	4.06	–0.40	6.04	0.95	4.94	–0.10	0.72	–0.43	2.11
Singleleaf pinyon	226	0.14	6.64	–1.42	2.37	–2.28	3.64	0.82	0.20	–0.03	0.05	–0.09	0.17
Western white pine	166	0.32	17.92	–3.78	6.07	–3.75	8.10	0.84	2.07	–0.34	0.64	–0.73	1.89
Ponderosa pine	3328	0.83	18.74	0.32	3.17	0.01	4.80	0.96	1.64	0.00	0.24	–0.04	0.86
Ghost pine	97	0.59	18.68	0.37	3.29	–0.65	4.06	0.92	1.84	0.01	0.30	–0.07	0.72
Balsam poplar	59	0.59	35.31	3.09	5.20	2.84	5.54	0.95	4.10	0.35	0.55	0.52	0.90
Douglas-fir	8767	0.80	25.71	0.31	4.41	0.39	7.25	0.95	2.56	0.01	0.42	0.05	1.56
Coast live oak	145	0.42	11.28	–0.02	2.69	1.10	4.77	0.90	1.29	0.07	0.25	0.39	1.06
Canyon live oak	1001	0.46	9.31	0.18	2.38	0.87	3.61	0.94	0.31	0.02	0.06	0.11	0.22
Blue oak	419	0.43	8.87	0.04	1.51	–0.48	1.87	0.99	0.46	0.00	0.04	–0.01	0.14
Oregon white oak	298	0.24	8.84	–1.05	2.62	–1.73	4.25	0.90	0.22	–0.01	0.06	0.00	0.22
California black oak	631	0.52	12.46	–0.96	2.72	–1.13	4.15	0.95	0.80	–0.03	0.14	–0.06	0.55
Interior live oak	348	0.43	9.78	0.40	2.34	1.20	2.84	0.97	0.32	0.02	0.05	0.08	0.17
Redwood	518	0.75	27.99	–0.74	4.67	–4.93	6.82	0.95	3.43	–0.21	0.51	–1.19	1.57
Western redcedar	808	0.84	20.12	–0.31	3.06	–1.97	5.19	0.97	2.08	–0.10	0.30	–0.54	1.19
Western hemlock	2431	0.79	22.57	0.17	3.94	–0.48	5.50	0.97	1.71	0.00	0.25	–0.02	0.92
Mountain hemlock	590	0.83	16.83	–0.15	2.77	0.63	5.11	0.92	1.31	0.02	0.27	0.14	1.11
California laurel	225	0.37	12.52	0.03	2.60	–0.13	3.40	0.97	0.28	0.00	0.05	–0.01	0.16

Note: AAD, average absolute deviation.
[a]Large trees were trees in the 80–100% DBH quantile.
[b]Volume is net cubic metre volume for trees greater than 12.5 cm DBH from a 30 cm stump to a 10 cm top, net of cull.

0.50 h more quickly on the second visit. Measurement time (f_j) included travel time, a fixed measurement time for nontree attributes, and a variable measurement time based on the number of trees on the plot. Nontree measurement time includes aspect, slope, elevation, GPS location, site index, seedling counts, downed woody debris transects, and understory vegetation. Per-tree measurement time for things other than tree height accounts for such tasks as diameter measurement, recording damage and disease, and tagging trees for identification at future visits.

Tree height measurement time (t_{ij}) for a given measurement strategy was based on the following assumptions: modeled heights required no field measurement time; each measured tree height required 20 s of crew time to measure and record the height; and each 3 m of tree height required an additional 3 s. This last assumption was included to account for the greater time required to move to where both the base and top of the tree can be seen when the tree is large. Multipliers were also used for steeper slopes, which increase time for height measurement. Cost for the crew of three people was assumed to be US$100/h.

Alternative set A: modeling by whole plot

Two sets of strategies for the optimization model were analyzed. The first set of strategies tested was used to look at what types of plots were selected for height modeling when the mean absolute error constraint was gradually increased from no error to the maximum, where maximum error corresponds to the absolute mean error that results from modeling all heights. The available strategies for each plot were limited to either all heights on the plot are modeled or all heights on the plot are measured. This set of alternatives was not considered practical, because it allows a measurement strategy to vary among plots. However, this approach is useful for understanding what plots show the highest efficiency gains when tree heights are modeled. Results were examined for patterns in geographic location, ownership, measurement and travel time, and stand characteristics.

Alternative set B: modeling by diameter class, species group, and intensity

The second set of strategies was used to examine the cost, error, and bias of different measurement rules that could be applied to the entire region. The available strategies for each plot were based on diameter class, hardwood or conifer group, and degree of modeling used (0% of trees, 50% of trees, or 100% of trees). The following DBH classes were used: (*i*) 12.5–25.0 cm; (*ii*) 25.1–37.5 cm; (*iii*) 37.6–50.0 cm; (*iv*) 50.1–62.5 cm; (*v*) 62.6–75.0 cm; and (*vi*) >75.0 cm. Constraint set 3 was used for the degree of modeling but not for hardwood or conifer group and diameter class as they were not mutually exclusive. To mimic the choice of a single measurement policy for the region, if any strategy was used it was constrained to be used on all plots (constraint 8).

Modeling tree height creates bias in the estimate of volume for the three-state region; contributing factors could be model error from using height equations that are designed to be unbiased for tree height rather than tree volume or having development data sets and application data sets drawn from slightly different populations. Substituting modeled tree heights (expected values) for measured heights will also affect the estimate of variance for the regional volume estimate. Bias in estimated variance can affect interpretation of results for the region, and potentially lead to incorrect conclusions about current conditions or rates of change for forest ecosystems. To provide a more complete picture of the consequences of using height models, the bias in estimated variance from a particular modeling strategy was calculated for each set of optimal height measuring strategies. This was done by taking the combination of modeled and measured heights specified in the optimal solution, calculating the resulting plot level volume for each plot, and then estimating variance of the three-state volume estimate using the method that is the standard for the US national inventory program (Bechtold and Patterson 2005, eq. 4.14).

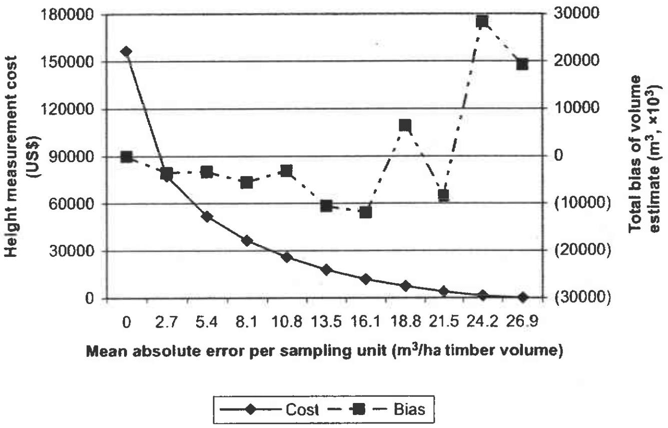
Results

When all heights are measured, total estimated volume for the region was 8.2034×10^9 m^3 with sampling error of 0.226×10^9 m^3. With all heights modeled, volume was overestimated by 0.24%, and sampling error was underestimated by 7.61%.

Alternative set A: modeling by whole plot

With all heights modeled, mean absolute error per plot was 26.9 m^3/ha. To test the effect of increasing tolerance for

Fig. 1. Alternative set A results for volume bias from height modeling and optimized height measurement cost for inventory of California, Oregon, and Washington in 2002 as constrained mean absolute error is varied from 0% to maximum by 10% increments.



error, 10 optimizations were made with average absolute error increased at equal intervals of 10% (0, 2.7, 5.4, 8.1, 10.8, 13.5, 16.1, 18.8, 21.5, 24.2, and 26.9 m^3/ha). The minimized cost for these simulations showed an exponential decline, with annual cost savings of more than US$79 312 for the first 10% allowance of error but minimal cost savings of $1215 for increasing error allowance from 90% of maximum to 100% of maximum (Fig. 1).

The shape of the optimized minimal cost curve is the result of unequal efficiency gains from modeling among plots. The first 10% relaxation of the error constraint resulted in height modeling for 33% of the plots, with the next two 10% increments resulting in modeling for an additional 15% and 13% of plots, respectively (Fig. 2). Those plots that were selected first as error was relaxed (Fig. 2*a*), which indicates high cost savings relative to contributed error, showed a slight tendency to be found on the drier eastern side of Oregon and Washington, and a slightly higher percentage of plots were selected for modeling in California than in the other states. Selected plots had a mean of 30 trees/plot, compared with a mean of 24 trees for those plots that were not selected. There was no pattern of selection for ownership, volume, stand size class, or productivity class. On average, the plots showing higher efficiency gains for height modeling had an additional 28 min of measurement time for nonheight variables, and an additional 5 min of travel time to the plot. The most distinctive factor in selection of plots was whether modeling heights allowed the plots to be completed in a single day, with plots requiring a second day of measurement dropping from 407 plots to 273 plots as constrained absolute mean plot error was increased from 0% to 10% of maximum. All versions of these alternatives solved optimally in a few seconds using the CPLEX 9.0 mixed integer optimizer (ILOG 2003) on a Sun Fire 240 Unix workstation with the Solaris 5.8 operating system and 8 gigabytes of RAM.

Alternative set B: modeling by diameter class, species group, and intensity

Alternative B allowed separate decisions to model height by DBH class, by hardwood or softwood group, and by in-

Fig. 2. Plots selected for height modeling for the 2002 field season when average absolute volume error was constrained to (*a*) 10%, (*b*) 20%, (*c*) 30%, and (*d*) 100% of maximum error.

tensity level (all trees measured, 50% of heights modeled, or all heights modeled), with constraint 8 used to ensure that any of these strategies would be applied to all plots. Separating choices into these categories greatly increased model complexity. The base model for these alternatives had 16 988 rows, 50 233 columns, and 114 715 nonzero variables. Only 4 of 10 alternatives solved optimally; for those alternatives where an optimum solution was not found, maximums of 22 h of processing time and a 1.5% optimization gap were used to limit the search.

When all trees were modeled, average absolute error per plot within a species group and diameter class was 2.8 m^3/ha. Because these alternatives required that any modeling choice be implemented on all plots in the region, the number of trees chosen for modeling increased fairly smoothly as the error constraint increased from 0% to maxi-

Table 3. Alternative set B height modeling strategies chosen as absolute mean volume error constraint is relaxed.

	Strategy for hardwood DBH class						Strategy for softwood DBH class						
Alternative[a]	1	2	3	4	5	6	1	2	3	4	5	6	No. of modeled trees
0	○	○	○	○	○	○	○	○	○	○	○	○	0
10	□	○	○	○	□	○	●	○	○	○	○	○	15 207
20	●	○	□	○	○	○	●	□	○	○	○	○	20 591
30	●	●	○	○	○	○	●	●	○	○	○	○	24 615
40	●	○	○	●	●	○	●	●	○	○	○	□	25 441
50	●	○	○	○	○	○	●	●	○	○	○	●	27 160
60	●	●	□	○	□	○	●	●	○	○	□	●	29 974
70	●	●	●	○	●	●	●	●	○	○	●	●	31 481
80	●	●	○	○	□	○	●	●	●	○	●	●	33 189
90	●	●	●	○	□	○	●	●	●	□	●	●	34 352
100	●	●	●	●	●	●	●	●	●	●	●	●	35 516

Note: ○, all trees measured; □, 50% of trees measured; ●, all trees modeled. DBH classes: 1, 12.5–25.0 cm; 2, 25.1–37.5 cm; 3, 37.6–50.0 cm; 4, 50.1–62.5 cm; 5, 62.6–75.0 cm; 6, >75.0 cm.
[a]Values are mean absolute errors as a percentage of maximum.

Table 4. Bias, precision, cost savings, and crew days for optimal solutions to alternative B height modeling strategies for 1389 plots in California, Oregon, and Washington inventoried during the 2002 field season.

Alternative[a]	Estimated volume (m^3, $\times 10^6$)	Bias in estimated volume (m^3, $\times 10^6$)	Estimated sampling error of volume (m^3, $\times 10^6$)	Bias in estimated sampling error (m^3, $\times 10^6$)	Direct measurement cost for heights (US$)[b]	Total second-day visits to plots (person-days)	Travel costs attributed to height measurement (US$)
0	8203.4	0.0	225.8	—	99 437	1218	56 915
10	8194.7	–8.6	224.9	–0.9	66 503	1008	36 908
20	8198.0	–5.4	223.4	–2.4	53 019	951	31 626
30	8191.1	–12.3	221.9	–4.0	42 282	894	25 909
40	8195.8	–7.5	219.3	–6.6	38 734	843	21 499
50	8211.4	8.0	217.8	–8.0	27 243	768	13 858
60	8205.1	1.7	215.1	–10.7	23 483	735	10 346
70	8210.2	6.9	212.9	–12.9	13 233	711	7 559
80	8204.7	1.3	211.2	–14.6	7 735	675	3 620
90	8219.0	15.6	210.4	–15.4	3 925	651	913
100	8222.6	19.3	208.6	–17.2	—	642	—

[a]Values are mean absolute errors as a percentage of maximum. See Table 3 for specification of which trees have modeled heights for each alternative.
[b]Does not include travel costs.

mum by 10% increments (Table 3). Small trees (12.5–25.0 cm DBH) were the first to be chosen for modeling as the error constraint was relaxed, with small conifers selected before small hardwoods. The last trees that would be selected for modeling were large hardwoods (DBH ≥ 50.0 cm) and conifers in the 50–62.5 cm DBH range.

Economic and error effects of the solutions found to the 10 alternative B problems are shown in Table 4. The solution chosen for allowing error up to 10% of maximum error (modeling 50% of hardwoods and all conifers in the 12.5–25 cm DBH range) results in direct measurement cost savings of approximately US$33 000 using the specified assumptions. This height-modeling strategy would also save 70 second-day crew visits to plots, which was equivalent to an additional US$20 000 (Table 4). The other alternative B solutions also showed that reduced travel costs were a substantial portion of potential cost savings: travel costs attributable to height measurement were 18–37% of total height measurement cost.

Modeling choices affected volume estimates less than sampling error estimates (Table 4). As absolute error constraints were tightened, solutions found resulted in declining measurement and travel costs but increased bias of sampling error. However, for the first 10% relaxation of error, which resulted in the greatest increment of cost and time savings, the optimal solution resulted in a modest 0.4% underestimate of sampling error.

Discussion

National inventories of forests have become of increasing importance in local and international debates over sustainability of resources and ecosystems. Such inventories are used for decision-making on issues as diverse as timber supply, biodiversity protection, forestland taxation, global climate change, and fire risk reduction. This paper presented an optimization method that can be used to examine cost and error tradeoffs that result from modeling attributes instead of collecting field measurements in an extensive forest inventory.

The optimization model was first presented as a set of equations that could be used for a variety of modeling versus measuring inventory problems, where minimized cost is the objective and constraints exist for modeling error and field crew time. In the example application that followed, simple nonlinear height models were developed as the alternative to measuring heights in the field. The height models were applied to plots inventoried in a three-state region in 2002, with resulting plot volume errors incorporated into an optimization model. The example application was used to demonstrate how the method can be used to understand opportunities (and disadvantages) from modeling attributes in extensive forest inventories.

The optimization method presented here constrains plot-level error rather than population-level error. Population-level estimation is the nominal purpose of most regional inventory programs. However, many people who use forest inventory data are interested in accuracy at the individual-tree level, for uses such as growth and yield modeling, or in accuracy at the plot level, for uses that include mapping, spatial data development, decision support applications, or forest planning. Constraining for error at the population level does not necessarily lead to greater accuracy at the plot level. Conversely, constraining at the plot level does tend to reduce population level error, as shown in results for the two alternative sets. If the sole concern for managers was population-level estimation, a different optimization modeling approach from the one presented here might be preferable.

Because the optimization model incorporates past data and because numerous assumptions must be made about travel time, measurement time, and economic costs, the model is useful for understanding relative magnitude of choices rather than prescribing an exact strategy and resulting outcomes for future years. Some conclusions from this example, using the assumptions made, are: (*i*) bias of sampling error estimates for regional timber volume was of greater magnitude than bias of the estimates themselves (e.g., 6% vs. 0.02% when all trees are modeled); (*ii*) travel cost can be a significant consideration in choosing whether to model heights in extensive forest inventories, representing 18–37% of the cost of optimized solutions for height modeling choices in this example; (*iii*) modeling heights for small trees (trees 12.5–25.0 cm DBH) had the highest efficiency gain of alternatives examined; (*iv*) modeling heights of large hardwoods showed the lowest efficiency gain of alternatives examined.

Error for individual tree heights is quite high with simple nonlinear regional height models, resulting in mean height errors of 1.5–6.1 m by individual species. However, a more deliberate consideration of height modeling requires considering the efficiency of height modeling for desired applications. Allowing 10% of the maximum possible mean absolute plot error resulted in finding an optimal solution that, using the assumptions made here, could result in cost savings (33% reduction of height measurement costs) and reduction in field time with relatively small bias in estimated regional volume and associated sampling error (0.1% and 0.4%, respectively).

These conclusions are dependent on the inclusion of all relevant aspects of the decision problem and the accuracy of model coefficients. There are aspects of the decision problem that cannot easily be incorporated into the model. For example, negative aspects of height modeling include the possibility of reduced accuracy in recording other individual tree attributes such as damage and decay, crown ratio, or tree cavities because of less time spent examining the tree. There are also positive aspects to height modeling that are difficult to incorporate, such as the possibility of reduced accident rates from shorter field days. Measurement error exists both in the coefficients for plot volume, which includes error from volume models and from imprecision of measuring tree height in the field, and in the cost and travel coefficients. A fruitful area for future work in this area might be to examine how uncertainty in model coefficients affects optimal solutions. However, because solution times for this application were quite long, a smaller application or different problem formulation would be advisable for that purpose.

Because using integer programming optimization to examine tradeoff between cost and error in extensive forest inventory is a new application of operations research methods in forestry, the models presented here have deliberately been kept simple. Possible avenues for future work include improved optimization strategies, more sophisticated representation of travel, allowing variable crew size, and using improved attribute models. For example, there are a variety of stand-level fitting techniques that have been developed for individual tree height prediction (e.g., Jayaraman and Zakrzewski 2001; Monleon 2003; Robinson and Wykoff 2004); although accompanied by increased implementation difficulties, these approaches promise improved height prediction. Improved collection of information about crew travel time, plot completion time, and time of measurement for variables would improve the ability to understand cost and error tradeoffs. Other topics that should be examined before adopting a height-modeling strategy would be the effects of using height models on common regional estimates other than timber volume, such as area by forest structure type or fire risk categories.

Acknowledgements

I thank B.Bechtold, J.Alegria, and D.Marshall for comments on an initial study plan; M.Dailey, K.Georgitis, and V.Monleon for comparisons of height model forms; the associate editor, two reviewers, and C.Scott for helpful comments on the paper; K.Campbell for calculation of tree volumes; and the USDA Forest Service Pacific Northwest FIA field crew for collection of the data used in this research.

References

Arabatzis, A.A., and Burkhart, H.E. 1992. An evaluation of sampling methods and model forms for estimating height–diameter relationships in loblolly pine plantations. For. Sci. **39**: 192–198.

Avery, T.E., and Burkhart, H.E. 2002. Forest measurements. 5th ed. McGraw-Hill, New York.

Bechtold, W.A., and Patterson, P.L. (*Editors*). 2005. The enhanced forest inventory and analysis program—national sampling design and estimation procedures. USDA For. Serv. Gen. Tech. Rep. SRS-80.

Bechtold, W.A., Zarnoch, S.J., and Burkman, W.G. 1998. Comparisons of modeled height predictions to ocular height estimates. South. J. Appl. For. **22**: 216–221.

Begin, J., and Raulier, F. 1995. Comparaison de différentes approaches, modèles et tailles d'échantillons pour l'établissement de relations hauteur–diamètre locales. Can J. For. Res. **25**: 1303–1312.

Curtis, R.O. 1967. Height–diameter and height–diameter–age equations for second-growth Douglas-fir. For. Sci. **13**: 365–375.

Ek, A.R. 1973. Performance of regression models for tree height estimation with small sample sizes. *In* Statistics in Forestry Research, Proceedings of the 4th Conference of the Advisory Group of Forest Statisticians, 20–24 August 1973, Vancouver, B.C. International Union of Forest Research Organizations, Vancouver, B.C. pp. 67–80

Garman, S.L., Acker, S.A., Ohmann, J.L., and Spies, T.A. 1995 Asymptotic height–diameter equations for twenty-four species in western Oregon. Forest Research Laboratory, Oregon State University, Corvallis, Ore. Res. Contrib. No. 10.

Green, E.J., and C.T. Scott. 1987. Simulation of subsampling selection rules for hardwood tree heights. *In* Proceedings of the Central Hardwood Forest Conference VI, 24–26 February1987, Knoxville, Tenn. *Edited by* R.L. Hay, F.W. Woods, and H. DeSelm. University of Tennessee Press, Knoxville, Tenn. pp. 443–448.

Gregoire, T.G., and Williams, M. 1992. Identifying and evaluating the components of non-measurement error in the application of standard volume equations. Statistician, **41**: 509–518.

Hanus, M.L., Marshall, D.D., and Hann, D.W. 1999. Height–diameter equations for six species in the coastal regions of the Pacific Northwest. Forest Research Laboratory, Oregon State University, Corvallis, Ore. Res. Contrib. No. 25.

Houghton, D.R., and Gregoire, T.G. 1993. Minimum subsamples of tree heights for accurate estimation of loblolly pine plot volumes. South. J. Appl. For. **17**: 124–129.

Hiserote, B., and Waddell, K. 2005. IDB, a database of forest inventory for California, Oregon, and Washington. USDA Forest Service, Pacific Northwest Research Station, Forest Inventory and Analysis, Portland, Ore.

Huang, S., Titus, S.J., and Wiens, D.P. 1992. Comparison of nonlinear height–diameter functions for major Alberta tree species. Can. J. For. Res. **22**: 1297–1304.

Huang, S., Price, D., and Titus, S.J. 2000. Development of ecoregion-based height–diameter models for white spruce in boreal forests. For. Ecol. Manage. **129**: 125–141.

ILOG. 2003. ILOG CPLEX 9.0 User's manual [online]. Available from http://www.ilog.com/products/cplex/ [accessed Aug. 25, 2005].

Jayaraman, K., and Zakrzewski, W.T. 2001. Practical approaches for calibrating height–diameter relationships for natural sugar maple stands in Ontario. For. Ecol. Manage. **148**: 169–177.

Krumland, B.E., and Wensel, L.C. 1988. A generalized height–diameter equation for coastal California species. West. J. Appl. For. **3**: 113–115.

Lappi, J. 1991. Calibration of height and volume equations with random parameters. For. Sci. **37**: 781–801.

Larsen, D.R., and Hann, D.W. 1987. Height–diameter equations for seventeen tree species in Southwest Oregon. Forest Research Laboratory, Oregon State University, Corvallis, Ore. Res. Pap. No. 49.

Martin, F.C., and Flewelling, J.W. 1998. Evaluation of tree height prediction models for stand inventory. West. J. Appl. For. **13**: 109–119.

Monleon, V.J. 2003. A hierarchical linear model for tree height prediction. *In* Proceedings of the 2003 meeting of the American Statistical Association, 3–7 August 2003, San Francisco, Calif. American Statistical Association, Alexandria, Va. pp. 2865–2869.

Moore, J.A. 1996. Height–diameter equations for ten tree species in the Inland Northwest. West. J. Appl. For. **11**: 132–137.

Parresol, B.R. 1992. Baldcypress height–diameter equations and their prediction confidence intervals. Can. J. For. Res. **22**: 1429–1434.

Richards, F.J. 1959. A flexible growth function for empirical use. J. Exp. Bot. **10**: 290–300.

Robinson, A.P., and Wykoff, W.R. 2004. Imputing missing height measures using a mixed-effects modeling strategy. Can. J. For. Res. **34**: 2492–2500.

SAS Institute Inc. 2001. SAS for Windows, version 9.01 edition. SAS, Cary, N.C.

Williams, M.S., Bechtold, W.A., and LaBau, V.J. 1994. Five instruments for measuring tree height: an evaluation. South. J. Appl. For. **18**: 76–82.

Zhang, L. 1997. Cross-validation of non-linear growth functions for modeling tree height–diameter relationships. Ann. Bot. **79**: 251–257.