The Sensitivity of Derived Estimates to the Measurement Quality Objectives for Independent Variables

Francis A. Roesch[1]

Abstract.— The effect of varying the allowed measurement error for individual tree variables upon county estimates of gross cubic-foot volume was examined. Measurement Quality Objectives (MQOs) for three forest tree variables (biological identity, diameter, and height) used in individual tree gross cubic-foot volume equations were varied from the current USDA Forest Service Forest Inventory and Analysis specifications in a simulation under alternative error models. Assuming unbiased errors may lead to a different control strategy than assuming unbiased errors. Strengthening the MQO for diameter was shown to help reduce the overall variance of volume estimates if diameter errors are slightly biased. Height errors responded favorably to increased control under both the biased and unbiased models. County volume estimates are somewhat robust to the MQOs for biological identity. However, increased control of biological identity did play a more important role when the underlying distributions for diameter and height were assumed to be biased than when these errors were assumed to be unbiased.

The five USDA Forest Service Forest Inventory and Analysis units (FIA) have adopted a common forest inventory design, including core variables, analysis procedures, and quality standards (USDA 2002). An important part of this national effort has been defining Measurement Quality Objectives (MQOs) including acceptable measurement error (or tolerances) for data collected on field plots. Little or no hard data were available to support the initial development of most of these MQOs. So rather than defining the MQOs to achieve a specified maximum variance due to measurement error, they were defined as the best guess as to what might be the specifications achievable by a well-trained observer.

Derived estimates are often the most important factors in considering the utility and applicability of inventory results for a particular purpose. If we wish to control the quality of derived estimates in forest inventories, we must do so by defining measurement quality objectives (MQOs) for those measured variables that contribute information to the derived estimate. To do this, we need to understand the relationships of the error distributions of the measured variables to the error distribution of the derived estimates, and these relationships are typically complex. This paper shows how one may use a simulation to evaluate the contribution to the mean squared error of a derived variable by the allowed error in measured independent variables. In an example, the measurement error allowed by the FIA's existing Measurement Quality Objectives (MQO) for three forest tree variables (species, diameter, and height) used in individual tree gross cubic-foot volume equations were varied in a simulation to examine the effects of the MQOs upon county estimates of gross cubic-foot volume. The simulations were run under two sets of assumptions for two of the variables, height and diameter. I first assumed that the true underlying error distribution was unbiased for each of these variables and subsequently assumed that the true underlying distributions for height and diameter were both biased and skewed.

Assume we are interested in a county attribute mean per acre for county j:

$$Y_j = \frac{1}{L_j}\sum_{i=1}^{N_j} y_i$$

where: Nj equals the number of trees within county j, Lj equals the land area in acres within county j, and yi equals the value of an attribute of tree i. N is uniquely partitioned into N_G groups, $g = 1,\ldots,N_G$. For each group there is a unique function of an easily measured variable vector **x** to y:

[1] Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Southern Research Station, P.O. Box 2680, Asheville, NC 28802–2680.
E-mail: froesch@fs.fed.us.

$$y_{i(g)} = f_g(\mathbf{x}_i), \qquad \mathbf{x} = x_{1i}, x_{2i}, x_{3i}, \ldots\ldots$$

Assume further that our variable of interest is gross cubic-foot volume (*gcv*) because it is a pivotal quantity at the FIA unit in the Southern Research Station (SRS), as it enters into equations for most other volume estimates. The *gcv* equations for trees by species group are linear functions of the form:

$$v_{i(g)} = a_g + b_g d^2_{i(g)} h_{i(g)}$$

where: $v_{i(g)}$ = *gcv* for tree *i* in species group *g*,

$d_{i(g)}$ = diameter of tree *i* in species group *g* at 4.5' above the ground (d.b.h.),

$h_{i(g)}$ = total height of tree *i* in species group *g*, and

a_g and b_g are regression coefficients for species group *g*.

Note that the functional form of volume equations is one aspect of inventory that is not yet standardized nationally. Therefore, the results of this investigation are directly applicable to equations currently used in the southern United States. However, many volume equations used today contain the Schumacher factor: $b_1 d^{b_2} h^{b_3}$ where b_2 is a parameter usually close to 2.0 and b_3 is a parameter usually close to 1.0. Because the Schumacher factor often has an overriding influence in the equation, it is reasonable to expect similar results if we conducted the same studies using the existing equations at other FIA units. For our purposes, we will assume that the functional relationship is known without error. Therefore, if species is correct, as well as diameter and height, the volume is correct.

Methods

Currently, the MQOs require the data collectors to correctly identify the species of all trees 95 percent of the time, and identify the genera of all trees 99 percent of the time. Note that the biological grouping of species into genera does not exactly match the empirical grouping of species referred to above. Depending on how species are grouped, a biological identity error may or may not affect the volume estimate. In addition, it is required that diameter at breast height be measured to within +/- 0.1 inch per 20 inches of diameter 95 percent of the time. Total tree height must be measured to within +/-10 percent of the true height 90 percent of the time (USDA 2001).

The Simulation

A simulation was used to examine the effects of measurement error allowed by the current as well as alternative MQOs upon county estimates of gross cubic-foot volume per acre (*GCV*). Data from the most recent cycle of the FIA survey measured in South Carolina were used which consisted of five consecutively measured panels. Each panel covered the entire State, and all five panels were measured over a period spanning slightly more than 3 years (1998 to 2001). Assuming the data were measured without error, the "true" *GCV* was calculated for each county *j* (GCV_j). For each set of MQOs, biological identity, diameter and height were randomly perturbed within the defined MQOs and error distribution assumptions. Error was randomly applied to the three volume equation variables (biological identity, diameter, and height) for each tree measured in the survey within the defined MQOs and error distribution assumptions. A small quality assurance (QA) data set from the 2000/2001 Forest Health Monitoring (FHM) field season was used to classify the error distributions under the unbiased and biased assumptions for the error distributions. The *gcv* was calculated from these realizations, and the mean gross cubic-foot volume per acre (   ) was calculated for each county in the State. This was ((GĈV)ng the current MQOs, and the alternative MQOs described in table 1. The specifications for each of the three variables were varied while the current specifications of the other two variables were maintained for comparison. The mean difference (MD), mean absolute difference (MAD), and mean squared differences (MSD) from the county results based on the original "true" data were calculated after 1,000 iterations. Specifically, for the error in each estimator of the county mean: $\varepsilon_j = \hat{Y}_j - Y_j = G\hat{C}V_j - GCV_j$, let *C* equal the number of counties in the State, and form three statistics based on 1,000 iterations:

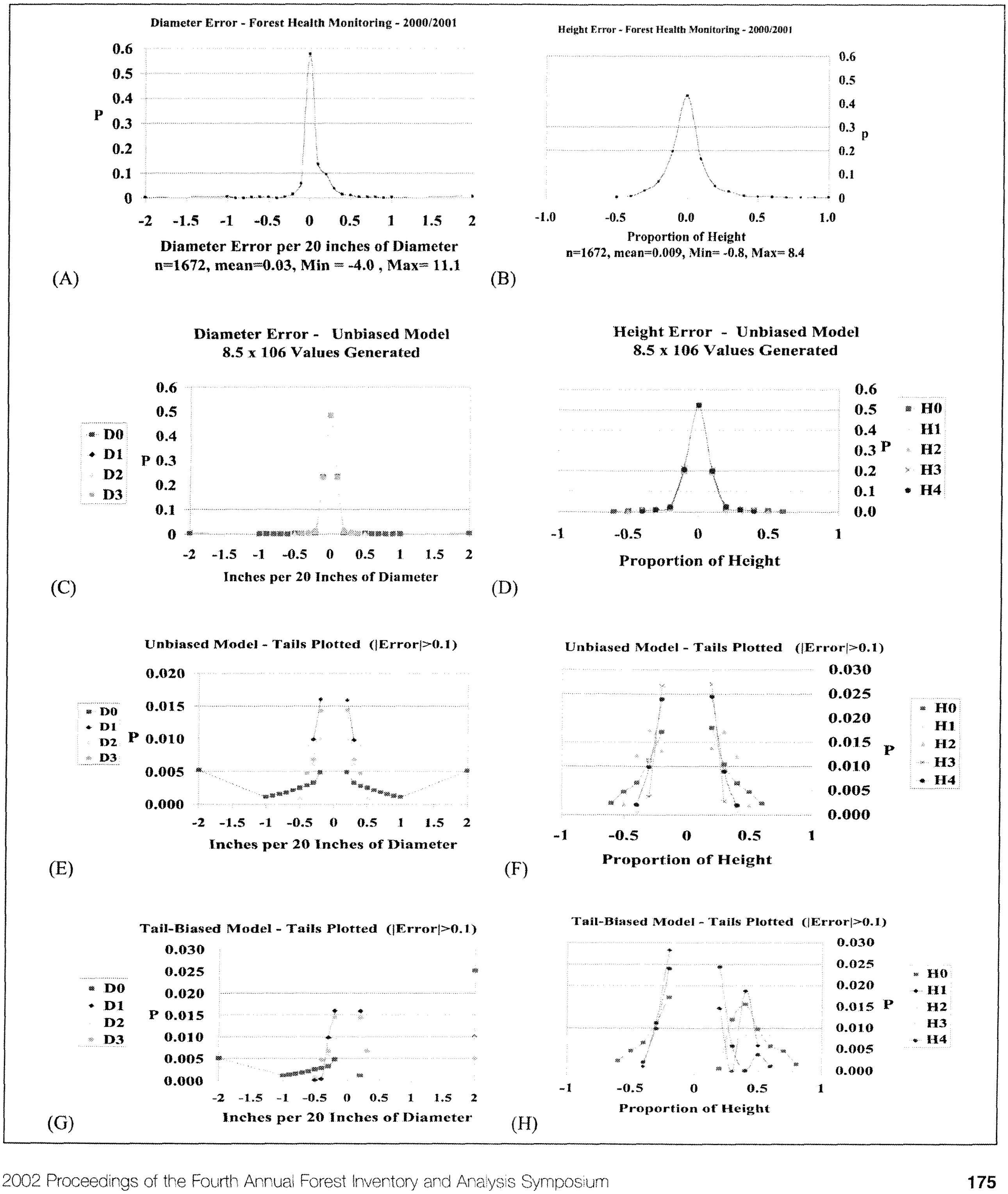
$$MD = \sum_{i=1}^{1000}\left[\frac{1}{C}\sum_{j=1}^{C}\varepsilon_j\right]\Big/1000, \quad MAD = \sum_{i=1}^{1000}\left[\frac{1}{C}\sum_{j=1}^{C}\left|\varepsilon_j\right|\right]\Big/1000,$$

and

$$MSD = \sum_{i=1}^{1000}\left[\frac{1}{C}\sum_{j=1}^{C}\varepsilon_j^2\right]\Big/1000.$$

Observations in earlier work (Roesch, in review) and in the 2000/2001 FHM QA data set showed that the error distributions for diameter and height were well behaved in the "in-con-

Figure 1.—*The error distributions observed in the 2000/2001 FHM QA data for diameter (A), and height (B), followed by the simulated distributions for the unbiased model for diameter (C), and height (D). The distribution tails of (C) are rescaled for clarity and plotted in (E). Likewise the tails of plot D are rescaled and plotted in (F). The tails of the simulated distributions for the tail-biased model for diameter are plotted in (G), while the corresponding tails for height are plotted in (H).*



trol" region, and poorly behaved in the "out-of-control" region. So this simulation concentrated on secondary criteria applied to the formally uncontrolled areas of the error distributions.

Species Identity

To vary the species identity determination, a random variate u_i was drawn from a uniform distribution ($U(0,1)$) for each tree i. Let p_1 be the proportion of time that the protocol requires identification of the correct species, and p_2 be the proportion of time that the protocol requires identification of the correct genus, $0 < p_1 < p_2 < 1$. The simulated species determination for tree i, $S_i{*}$, was calculated by sampling the following distribution:

$$S_i{*} = \begin{cases} S_i & u_i \le p_1 \\ rs(G_{S_i}) & p_1 < u_i \le p_2 \\ rs(F_{S_i}) & p_2 < u_i \le p_2 + r(1-p_2) \\ rs(A) & p_2 + r(1-p_2) < u_i \le 1 \end{cases}$$

where:

S_i = the true species of tree i,

$rs\ (G_{S_i})$ = a random selection from all observed species of the same genus as tree i, except for the species of tree i, unless S_i is the sole species within the genus,

$rs\ (F_{S_i})$ = a random selection from all species in the species list belonging to the same family as tree i, minus those species in the genus of tree *i*,

$rs\ (A)$ = a random species selection from the entire species list minus those species in the family of tree *i*, and

r = the proportion of time that out of genus errors are assumed to be within the family of tree *i*.

Note that under this distribution, the expected value of a correct species call is actually higher than the protocol requires for sole-species genera. This is necessary to meet the within-genus criterion for sole-species genera. The FHM QA data showed that 70 percent of the time when a species identity error fell outside of the correct genus under the current QA specifications, it fell within the correct family. This proportion was used for r in the straw man distribution (S0 in table 1) that is based on the current MQOs. The alternative MQOs for species identity investigated in this study also appear in table 1 (S1, S2, S3, and S4) and involve increases in p_1, p_2 and $p_3 = p_2 + r(1 - p_2)$.

Table 1.—*Alternative additional measurement quality objectives; the units for tolerance (t) are as follows: (1) for species and genus - deviation from true biological identity, (2) for height - the percent deviation from the true height, and (3) for diameter – inches per 20 inches of true diameter. p is the percentage of observations that are required to be correct. The set of alternatives (S0,H0,D0), derived from the original specifications and the 2000/2001 FHM data, form the assumed baseline in table 2.*

Alternative	Specification	*t*	*P*
Current criteria	Species, genus	0, 0	0.95, 0.99
	Height	0.1	0.90
	Diameter	0.1	0.95
S0	Species, genus, family	0, 0, 0	0.95, 0.99, 0.997
S1	Species, genus, family	0, 0, 0	0.975, 0.99, 0.997
S2	Species, genus, family	0, 0, 0	0.975, 0.995, 0.9985
S3	Species, genus, family	0, 0, 0	0.975, 0.995, 0.999
S4	Species, genus, family	0, 0, 0	0.975, 0.995, 0.9995
H0	Height	0.1, 0.5	0.90, 0.99
H1	Height	0.1, 0.2	0.90, 0.95
H2	Height	0.1, 0.3	0.90, 0.95
H3	Height	0.1, 0.2	0.90, 0.98
H4	Height	0.1, 0.3	0.90, 0.99
D0	Diameter	0.1, 1.0	0.95, 0.99
D1	Diameter	0.1, 0.2	0.95, 0.98
D2	Diameter	0.1, 0.3	0.95, 0.98
D3	Diameter	0.1, 0.3	0.95, 0.99

Diameter

Recall that the current specifications require that diameter at breast height is measured to within +/- 0.1 inch per 20 inches of diameter 95 percent of the time. Four alternative specifications (D1, D2, D3, and D4 in table 1) are compared to two straw man distributions based on the current MQOs (D0 in table 1). The first straw man distribution for diameter error is created by splining overlapping unbiased normal distributions, scaled by p_1, t_1, p_2, t_2 and d_{cat}. The unbiased straw man distribution assumes that the small bias observed in the 2000/2001 FHM data is an anomaly of that particular data set and is ignorable. An assumption that diameter measurements are unbiased is supported by the findings of Pollard and Smith (2001). We will use this FHM QA data (plotted in figure 1(A)) to classify the tails of the distribution. The error distributions of this data set might differ from the true underlying distribution because the data are weighted toward inexperienced observers, and they measure observer-to-observer error rather than observer-to-truth error. Let:

z_i = a random variate from a *($N(0,1)$)* for tree *i*,

p_1= proportion of time the measurement must be within a tolerance t_1, of true diameter (d_i), in tenths of an inch per 20 inches of diameter, $0 < p_1 < p_2 < 1$,

p_2 = 0.99, the proportion of time the 2000/2001 FHM data fell within 1.0" (t_2) of the true diameter.

$$d_{cat} = \left[trunc(d_i/20.0) + 1 \right]$$

$$\alpha_1 = 0.5(1 - p_1)$$

$$\alpha_2 = 0.5(1 - p_2)$$

$$e_i = \begin{cases} z_i\left(d_{cat}t_1 + 0.0499\right)/z_{\alpha_1} & if \quad |z_i| \le z_{\alpha_1} \\ -d_{cat}t_1 - 0.05 + \left(d_{cat}\left(t_2 - t_1\right)\frac{(z_i + z_{\alpha_1})}{(z_{\alpha_2} - z_{\alpha_1})}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \le z_{\alpha_2}\right) \cup \left(z_i \le 0\right) \\ d_{cat}t_1 + 0.05 + \left(d_{cat}\left(t_2 - t_1\right)\frac{(z_i - z_{\alpha_1})}{(z_{\alpha_2} - z_{\alpha_1})}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \le z_{\alpha_2}\right) \cup \left(z_i > 0\right) \\ -0.05 + d_{cat}t_2\frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i \le 0\right) \\ 0.05 + d_{cat}t_2\frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i > 0\right) \end{cases}$$

The distributions we used to compare increased tolerance specifications to the straw man differ from the straw man only in the definitions of p_2 and t_2 and inferences about them. Here t_2 is a required tolerance to be met a proportion p_2 of the time, rather than an assumed parameter of the underlying distribution. That is, we are enforcing a second tier of control, which is more restrictive than the underlying error distribution that arises from the original level of control. Therefore, the error distributions for the alternative distributions are the same as the straw man error distribution, save for the definition and interpretation of p_2 and t_2. Now:

p_2 = the proportion of time the measurement must be within a tolerance t_2 , of true diameter (d_i), in tenths of an inch per 20 inches of diameter, $0 < p_1 < p_2 < 1$.

The error distributions for diameter resulting from applying the unbiased error distribution 8.5x10[6] times under the various sets of MQOs for diameter are seen in figure 1(C), while the tails of the distributions are rescaled for clarity in figure 1(E). Note that the long tails observed in the FHM QA data have been retained in the straw man distribution, but they are reduced as control due to the MQOs increases.

An alternative Straw Man distribution would arise if we thought that the previously ignored bias in the 2000/2001 FHM data indicated the true underlying distribution. To model the slight bias and skewness, we would alter our original straw man by applying all of the observed bias to the right tail. Therefore, the second straw man distribution for diameter error is identical to the first except that the bias observed in the 2000/2001 FHM data is added to the right tail of the distribution to approximate both the observed bias and skewness. Let:

$b_i = 0.031 d_{cat}\,(2/(1 - p_1))$

Then:

$$e_i = \begin{cases} z_i\left(d_{cat}t_1 + 0.0499\right)/z_{\alpha_1} & if \quad |z_i| \le z_{\alpha_1} \\ -d_{cat}t_1 - 0.05 + \left(d_{cat}\left(t_2 - t_1\right)\frac{(z_i + z_{\alpha_1})}{(z_{\alpha_2} - z_{\alpha_1})}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \le z_{\alpha_2}\right) \cup \left(z_i \le 0\right) \\ b_i + d_{cat}t_1 + 0.05 + \left(d_{cat}\left(t_2 - t_1\right)\frac{(z_i - z_{\alpha_1})}{(z_{\alpha_2} - z_{\alpha_1})}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \le z_{\alpha_2}\right) \cup \left(z_i > 0\right) \\ -0.05 + d_{cat}t_2\frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i \le 0\right) \\ b_i + 0.05 + d_{cat}t_2\frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i > 0\right) \end{cases}$$

For the alternative distributions, we assume that bias can be eliminated from the " in-control" region of the distributions when the second level of control is applied. Therefore, the alternative error distributions are identical to the biased straw man error distribution shown above except that they do not include the bias term in the third line on the right hand side of the equation.

The tails of the error distributions for diameter resulting from sampling the biased error distribution 8.5x10[6] times under the various sets of MQO's for diameter are seen in figure 1(G). That graph shows that the bias, skewness, and influence of the tails are all reduced as the MQOs are increased.

Diameter entries, regardless of error are always interpreted as recorded to the measurement interval of 0.1 inch, and are never negative. Therefore, let *round()* be an operation that rounds to whole integers. Then all of the diameter error distributions are made discrete to the measurement interval, and negative diameters at set equal to zero:

$$d_i^* = \begin{cases} \left(round\left(10.0 * \left(d_i + e_i\right)\right)\right)/10.0 & if\left(d_i + e_i\right) \geq 0 \\ 0 & otherwise \end{cases}.$$

The bias to *gcv* estimates added by truncation of diameters at zero is extremely small, since the *gcv* of trees less than 5.0 inches in diameter is zero. Therefore, we will note but otherwise ignore this small amount of bias added to the perturbed diameters.

Height

Because the proportion of height error data from the FHM QA data set showed roughly the same properties as the diameter error data, I used the same approach to simulating height error as I did for diameter error. First, I assumed that the observed bias and skewness are simply anomalies found in that data set rather than indicating the true underlying distribution for FIA proportion of height error. Then, in a second straw man distribution I assumed that the observed bias and skewness truly indicates the underlying distribution. The unbiased model was formed the same way as the unbiased model for diameter. That is by using a spline of overlapping normal distributions, the first scaled by p_1, t_1, and true height and the second scaled by p_2, t_2 and true height. Let:

z_i = a random variate from a $(N(0,1))$ for tree i,

p_1 = the proportion of time the measurement must be within a tolerance t_1, of true height, $0 < p_1 < p_2 < 1$,

p_2 = 0.99, the proportion of time the FHM data fell with in 0.5 t_2 of the true height.

$$\alpha_1 = .5(1 - p_1)$$
$$\alpha_2 = .5(1 - p_2)$$

$$e_i = \begin{cases} z_i\left(h_i t_1 + 0.499\right)/z_{\alpha_1} & if & |z_i| \leq z_{\alpha_1} \\ -h_i t_1 - 0.5 + \left(h_i\left(t_2 - t_1\right)\frac{\left(z_i + z_{\alpha_1}\right)}{\left(z_{\alpha_2} - z_{\alpha_1}\right)}\right) & if & \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \leq z_{\alpha_2}\right) \cup \left(z_i \leq 0\right) \\ h_i t_1 + 0.5 + \left(h_i\left(t_2 - t_1\right)\frac{\left(z_i - z_{\alpha_1}\right)}{\left(z_{\alpha_2} - z_{\alpha_1}\right)}\right) & if & \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \leq z_{\alpha_2}\right) \cup \left(z_i > 0\right) \\ -0.5 + h_i t_2 \frac{z_i}{z_{\alpha_2}} & if & \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i \leq 0\right) \\ 0.5 + h_i t_2 \frac{z_i}{z_{\alpha_2}} & if & \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i > 0\right) \end{cases}$$

As in the case of diameter, the distribution that we used to compare increased tolerance specifications to the straw man differ from the distribution for the straw man only in the definitions of p_2 and t_2 and inferences about them. Here t_2 is a required tolerance to be met a proportion p_2 of the time, rather than an assumed parameter of the underlying distribution. That is, in the alternative MQO specifications (H1, H2, H3, and H4) we are enforcing a second tier of control that is more restrictive than the underlying error distribution that arises from the original level of control (H0). Therefore, the error distributions assuming no bias are the same as the straw man error distribution, save for the definition and interpretation of p_2 and t_2:

p_2 = the proportion of time the measurement must be within a tolerance $t_2 h_i$, of true height (h_i), $0 < p_1 < p_2 < 1$.

The error distributions for proportion of height resulting from sampling the unbiased error distribution 8.5x10[6] times under the various sets of MQOs for height are seen in figure 1(D). The tails of that graph are rescaled for clarity and plotted in figure 1(F). As with the unbiased diameter error distributions, the tails are drawn toward the center as MQOs are increased.

Again, an alternative straw man distribution would arise if we thought that the previously ignored bias and skewness in the 2000/2001 FHM data was somewhat indicative of the true underlying distribution. The second straw man distribution for proportion of height error is identical to the first, except that the small amount of bias observed in the 2000/2001 FHM data is added to the right tail of the distribution to approximate both the bias and skewness seen in that data:

$$b_i = 0.00901 h_i \left(2/\left(1 - p_i\right)\right)$$

$$e_i = \begin{cases} z_i\left(h_i t_1 + 0.499\right)/z_{\alpha_1} & if & |z_i| \leq z_{\alpha_1} \\ -h_i t_1 - 0.5 + \left(h_i\left(t_2 - t_1\right)\frac{\left(z_i + z_{\alpha_1}\right)}{\left(z_{\alpha_2} - z_{\alpha_1}\right)}\right) & if & \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \leq z_{\alpha_2}\right) \cup \left(z_i \leq 0\right) \\ b_i + h_i t_1 + 0.5 + \left(h_i\left(t_2 - t_1\right)\frac{\left(z_i - z_{\alpha_1}\right)}{\left(z_{\alpha_2} - z_{\alpha_1}\right)}\right) & if & \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \leq z_{\alpha_2}\right) \cup \left(z_i > 0\right) \\ -0.5 + h_i t_2 \frac{z_i}{z_{\alpha_2}} & if & \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i \leq 0\right) \\ b_i + 0.5 + h_i t_2 \frac{z_i}{z_{\alpha_2}} & if & \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i > 0\right) \end{cases}$$

As with diameter, we assume that bias can be eliminated from the "in-control" region of the distributions when the second level of control is applied. Therefore, the distributions arising under the alternative MQO specifications (H1, H2, H3, and H4) are identical to the error distribution above, except that they do not include the bias term in the third line on the right hand side of the equation.

Table 2.—*The mean difference (MD), mean absolute difference (MAD), and mean squared difference (MSD) for each alternative MQO specification in table 1, after 1,000 iterations, under the assumptions of the unbiased and biased straw man error distributions. MD and MAD are in ft^3/acre. MSD is in $(ft^3/acre)^2$*

Errors in county estimates of gross cubic-foot volume per acre.						
MQO	Unbiased model			Tail-biased model		
alternative	MD	MAD	MSD	MD	MAD	MSD
S0,H0,D0	1.414	8.221	116.404	30.170	30.305	1124.661
S1	0.091	8.063	112.607	28.833	28.997	1043.066
S2	0.444	8.057	112.363	29.204	29.357	1065.213
S3	0.471	8.070	112.606	29.223	29.374	1066.557
S4	0.498	8.063	112.459	29.264	29.414	1068.781
H1	1.374	6.384	69.311	21.958	22.109	610.693
H2	1.406	8.001	109.687	21.965	22.391	659.834
H3	1.328	5.782	56.696	17.014	17.235	381.091
H4	1.394	6.351	68.007	15.427	15.859	336.589
D1	1.237	8.011	110.849	22.229	22.636	676.282
D2	1.222	8.020	110.970	22.270	22.673	678.043
D3	1.242	8.022	111.054	19.881	20.466	571.266

The tails of the error distributions for proportion of height resulting from applying the biased error distribution 8.5×10^6 times under the various sets of MQOs for height are seen in figure 1(H). Again, the bias, skewness, and influence of the tails are all reduced as the MQOs are increased. We assume that height errors are also discrete and not negative. Therefore our simulated heights are calculated as:

$$h_i^* = \begin{cases} round(h_i + e_i) & if\ (h_i + e_i) \geq 0 \\ 0 & otherwise \end{cases}.$$

Results and Conclusions

Table 2 gives the mean difference (MD), the mean absolute difference (MAD), and the mean squared difference (MSD) for each MQO specification in table 1, after 1,000 iterations. The results show that assuming unbiased errors may lead to a different control strategy than assuming bias in the "out-of-control" region. Strengthening the MQO for diameter will help reduce the overall variance of volume estimates if diameter errors are slightly biased in this out-of-control region. Height errors responded favorably to increased control in the current out-of-control region under both the biased and unbiased models.

Volume estimates at the county level are somewhat robust to the MQOs for species identity. However, more accurate species identity did play a more important role when the underlying distributions for diameter and height were assumed to be biased than when they were assumed to be unbiased.

Simulation is useful for investigating the effect of MQOs for independent variables on aggregated dependent variable estimates if reasonable error models can be postulated for the measurement errors of the independent variables. In this case, a small amount of QA data was available that is most likely drawn from a population different from the population of interest. Rather than defining a single distribution for height and diameter errors, intended to represent the underlying population of interest, we defined two for each of these variables that are intended to represent the extremes of the true underlying distributions. Any conclusions that could be drawn from both straw man distributions for a particular variable could be considered robust. However, any conclusion that would only be drawn under one of the straw man distributions should probably be applied more cautiously.

The extension of the methodology used in this paper to other measured and derived variables is straightforward. One simply needs to posit reasonable error models for the measured variables and then simulate attribute variance with those models while observing the effect upon the summary statistics of the derived variables. As applicable quality assurance data becomes more available, they will support a decision to maintain, replace, or refine the error models and MQOs.

Literature Cited

Pollard, J.E.; Smith, W.D. 2001. Forest health monitoring 1999 plot component quality assurance report. Research Triangle Park, NC: U.S. Department of Agriculture, Forest Service, National Forest Health Monitoring Program.

Roesch, F.A. In Review. Testing the sensitivity of county-level volume estimates to the measurement quality objectives for independent variables in individual tree volume equations.

U.S. Department of Agriculture. 2001. SRS field guide version 1.5 (update Jan. 19, 2001). Asheville, NC: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Southern Research Station. 311 p.

U.S. Department of Agriculture. 2002. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.6. Washington, DC: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis.