

Propagating Uncertainty in Predicting Individuals and Means Illustrated with Foliar Chemistry and Forest Biomass

Ruth D. Yanai,^{1*}  John E. Drake,¹ Hannah L. Buckley,² Bradley S. Case,²
Paul J. Lilly,³ Richard C. Woollons,⁴ and Javier G. P. Gamarra⁵

¹Department of Sustainable Resources Management, State University of New York College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, New York 13210, USA; ²School of Science, Auckland University of Technology, 34 St. Paul Street, Auckland 1010, New Zealand; ³EP Carbon, 2930 Shattuck Ave., Berkeley, California 94705, USA; ⁴School of Forestry, University of Canterbury, Private Bag 4800, Christchurch, New Zealand; ⁵Forestry Division, Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy

ABSTRACT

Quantifying uncertainty is important to establishing the significance of comparisons, to making predictions with known confidence, and to identifying priorities for investment. However, uncertainty can be difficult to quantify correctly. While sampling error is commonly reported based on replicate measurements, the uncertainty in regression models used to estimate forest biomass from tree dimensions is commonly ignored and has sometimes been reported incorrectly, due either to lack of clarity in recommended procedures or to incentives to underestimate uncertainties. Even more rarely are the uncertainty in predicting indi-

viduals and the uncertainty in the mean both recognized for their contributions to overall uncertainty. In this paper, we demonstrate the effect of propagating these two sources of uncertainty using a simple example of calcium concentration of sugar maple foliage, which does not require regression, then the mass of foliage and calcium content of foliage, and finally an entire forest with multiple species and tissue types. The uncertainty due to predicting individuals is greater than the uncertainty in the mean for studies with few trees—up to 30 trees for foliar calcium concentration and 50 trees for foliar mass and calcium content in the data set we analyzed from the Hubbard Brook Experimental Forest. The most correct analysis will take both sources of uncertainty into account, but for practical purposes, country-level reports of uncertainty in carbon stocks can safely ignore the uncertainty in individuals, which becomes negligible with large enough numbers of trees. Ignoring the uncertainty in the mean will result in exaggerated confidence in estimates of forest biomass and carbon and nutrient contents.

Received 28 February 2023; accepted 25 October 2023

Richard C. Woollons: Deceased.

Author Contributions: RDY initiated this study to resolve the debate about accounting for uncertainty in individuals versus uncertainty in the mean. RCW provided early biostatistical guidance. HLB began coding the Monte Carlo analysis in R, with BSC translating input from RDY. JED joined the effort and improved the analysis to account for both sources. PJJ set up the Monte Carlo in Excel. JGPG clarified the mathematical notation and resolved the conflict between the numerical and analytical approaches to error propagation. RDY led the writing with input from RCW, JED, and GPG. We learned a lot, albeit slowly; this project involved > 7 years of intermittent effort and a changing cast of characters.

*Corresponding author; e-mail: rdyana@syr.edu

Published online: 22 January 2024

Key words: allometric models; forest biomass; nutrient budgets; Monte Carlo error propagation; uncertainty in regression; foliar calcium; forest carbon; Hubbard Brook.

HIGHLIGHTS

- Predicting attributes of a single individual is more uncertain than the mean.
- With large numbers of individuals, uncertainty in the mean is more important.
- Both sources are important in small samples, which has not previously been recognized.

INTRODUCTION

In some contexts, it can be important to predict the likelihood of outcomes for individuals, such as risks to human health (Bogardus and others 1999) or failures in equipment (Heng and others 2009). In others, it is important to predict the likely properties of means, such as a population of voters (Wlezien and others 2013) or a portfolio of investments (Zaimovic and others 2021). While the

statistics for reporting uncertainties in either the prediction of individuals or the estimates of means are both well known, methods for computing the combined effect of both sources are not. Importantly, ecosystem science operates at scales in which both sources of uncertainty are commonly relevant.

Establishing statistical confidence in forest budgets is essential to research, management, and policy goals. Forest elemental budgets are needed to understand nutrient limitation, uptake, and harvest removals. At larger scales, forest carbon accounting is increasingly important to climate mitigation efforts (Keith and others 2021). In international carbon finance for climate mitigation, uncertainty in estimates of emission reductions from deforestation is important to determining payments made (Yanai and others 2020).

Long-term monitoring of forest carbon and nutrient budgets is not usually based on destructive harvests, but depends instead on measuring tree attributes such as diameter and height, converting these to biomass using allometric relationships developed from a destructive sample of trees (Box 1), and converting biomass to carbon and nutrient contents based on measured concentrations. There are thus multiple sources of uncer-

Box 1: Uncertainty in regression

The construction of allometric regression models is fundamental to most studies of forest biomass and nutrient content, because harvesting trees to obtain direct measures is destructive. Fortunately, there are consistent relationships between tree biomass and non-destructive measures such as diameter and height, which can be obtained for a sample that is representative of the forest of interest. These relationships are non-linear, but a log-log relationship is often very linear. Thus, we commonly construct simple linear regression models with data from a sample A of n trees used to construct the allometric model, of form:

$$\hat{Y}_{A,i} = \hat{a} + \hat{b}X_{A,i} + e_{A,i}, \quad (1)$$

where $\hat{Y}_{A,i}$ = the estimate of the dependent variable, usually something difficult or expensive to measure (in our example, log-transformed foliar biomass) in tree i of sample A .

$X_{A,i}$ = an independent variable, usually some simple tree dimension (in our example, log-transformed tree diameter), in tree i of the sample A used to construct the allometry,

\hat{a}, \hat{b} = intercept and slope parameter estimates, computed by least squares techniques (Draper & Smith, 1998), and

$e_{A,i}$ = a random error term in the tree i of the sample A , assumed to be normally and independently distributed with a mean of zero and a constant variance (σ^2). Specifically, it defines a random draw from a distribution defined by σZ_0 , where $Z_0 \sim N(0, 1)$ is a standard normal random variable, and σ is the residual standard deviation of the regression:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (Y_{A,i} - \hat{Y}_{A,i})^2}{n - 2}},$$

where $Y_{A,i}$ = the i th value of Y used in constructing Eq. (1),

$\hat{Y}_{A,i}$ = the corresponding predicted value obtained from Eq. (1), and
 n = number of observations used in constructing Eq. (1).

The standard deviation of the regression can also be used to quantify the precision of Eq. (1) in predicting either a mean value or a specific individual value, both of which depend on the value of the independent variable X for that particular individual, which we call X_0 .

Models such as Eq. (1) are usually constructed from data obtained from a sample of the population of interest. Therefore, the model does not perfectly describe the population of interest; it is subject to error. For example, if several random samples are drawn from the population, the parameter estimates will differ (a little) from sample to sample.

The estimated value, \hat{Y}_0 , is the same whether predicting a mean or an individual, but the uncertainty is much larger when predicting an individual. Note that the difference between the uncertainty in prediction of the mean (Eq. 2) and the uncertainty in prediction of an individual (Eq. 4) is the standard deviation of the regression, σ (Draper and Smith 1998).

Uncertainty in prediction of the mean for a particular value of x :

$$\sigma \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X}_A)^2}{\sum_{i=1}^n (X_{A,i} - \bar{X}_A)^2}} \quad (2)$$

Uncertainty in prediction of an individual:

$$\sigma \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X}_A)^2}{\sum_{i=1}^n (X_{A,i} - \bar{X}_A)^2}} \quad (3)$$

The uncertainty in both cases is smallest when X_0 equals the mean of the sample (\bar{X}_A), and increases as X_0 deviates in either direction from \bar{X}_A .

Multiplying Eq. (2) or (3) by a t -statistic with $(n-2)$ degrees of freedom at a specified value of α (commonly 0.05) gives a confidence interval for Eq. (2) or prediction interval for Eq. (3).

To illustrate confidence in Eq. (1), it is helpful to graph the regression line along with the sample data as well as depicting confidence and prediction bounds for a range of X_0 (Figure 3). The area within each of the bounds represents where 95% of the predictions of the mean or of the individual predictions should lie.

Picard and others (2012) and Breidenbach and others (2016) give a more formal statistical description of these intervals, together with generalized prediction formulae in matrix notation, which can be used for more complicated models.

The use of Eqs. (2) or (3) requires the number of observations in the regression, n ; the mean of the observations, \bar{X}_A ; and the sum of squared deviations of the X , $\sum_{i=1}^n (X_{A,i} - \bar{X}_A)^2$. In the past, these statistics were not commonly reported, but current practices make it more likely that they could be calculated from published data sets (for example Falster and others 2015).

tainty in these estimates (Yanai and others 2012) and many possible ways to make mistakes in accounting for them.

Sampling error, which is due to spatial variation in tree and forest properties across the landscape, is commonly the biggest contributor to uncertainty in forest inventory (for example, Holdaway and others 2014; McRoberts and others 2016) and is easily quantified using replicate plots. Measurement error, for example of tree diameter and height, can also be quantified by replicate measurements, as are commonly made in the quality assurance process (Yanai and others 2022). Natural variation in the concentration of carbon (McRoberts and others

2016) and nutrients (Yang and others 2016) in tree tissues is also readily quantified by replicate measurements. In contrast, the uncertainty in predicting tree biomass based on tree dimensions is more difficult to quantify correctly, because it requires understanding how to propagate uncertainty in regression models.

When a regression model is applied to a number of trees to estimate their biomass, those estimates are affected by uncertainties related to both how far an observation for an individual tree may depart from the regression model prediction and also how accurately the regression model has captured the true relationship between biomass and tree

dimensions (Box 1). Both of these sources of uncertainty can be important, but they are rarely evaluated in tandem. Some investigators have represented the uncertainty of forest estimates by propagating individual-level uncertainty, while others have propagated uncertainty in the mean.

For example, uncertainty in the carbon content of the Hubbard Brook Experimental Forest was based on uncertainty in the prediction of individuals (Fahay and others 2005), while uncertainty in forest nitrogen content at Hubbard Brook was based on the uncertainty in the mean (Yanai and others 2010). In the New Zealand forest inventory, uncertainty in the mean was used for volume, but uncertainty in individuals was used for wood density (Holdaway and others 2014). In a study in Canada, uncertainty in individuals was used to describe plot-level uncertainty (Paré and others 2013), and in another in California, uncertainty in individuals was used in remote-sensing-based carbon assessment (Gonzalez and others 2010). Thus, previous investigators have often ignored one or the other source of allometric uncertainty. A complete uncertainty accounting would propagate both the uncertainty in predicting the properties of an individual and the uncertainty in estimating mean properties.

In this paper, we illustrate how to propagate uncertainty in predicting mean properties, such as those of a forest, and how this differs from the uncertainty in predicting the properties of an individual, such as a tree. We begin with a single dependent variable, namely the calcium concentration of leaves, to illustrate the effect of the number of trees on the importance of accounting for individual prediction. We then extend this analysis to a regression model describing leaf biomass as a function of tree diameter, which is more complex. Our final application is to a forest nutrient budget with multiple species and tissue types. These analyses all show that the uncertainty in predicting individuals is important for small numbers of individuals but that the confidence in the model (or the mean, in the univariate case) is important in all cases and should not be ignored. Understanding this difference is essential to correctly propagating uncertainty in estimates of forest attributes, including carbon storage, at scales from the tree to the globe.

ILLUSTRATION

Uncertainty in the Univariate Case of Nutrient Concentrations

Constructing forest nutrient budgets can require estimating the concentrations of multiple elements

for multiple tissue types (because leaves, bark, and wood differ in concentration) in multiple tree species. We use the example of the concentration of calcium in sugar maple leaves, a topic of concern for sugar maple health (Horsley and others 2000), to illustrate the uncertainty in the population mean and the uncertainty in the prediction of an individual.

Consider an idealized forest composed entirely of sugar maple trees in which each individual tree has a characteristic concentration of Ca in its foliage. In this simplified forest, we ignore the fact that leaf concentrations vary within a tree (sun leaves commonly differ from shade leaves) and, for the mean concentration of the forest, we ignore the fact that some trees have more leaves than others. We ask two questions: “What is the uncertainty in estimating the mean Ca concentration of leaves in the forest?” and “What is the uncertainty in estimating the Ca concentration of the leaves of a particular tree?” We were taught that to answer the former question, which is about the population mean, we should use the standard error of the mean, while for the latter question, which is about predicting an individual, we should use the standard deviation. However, both of these uncertainties can be important, depending on the sample size (Box 2).

To illustrate the difference between a sample mean and the true population mean, we generated Ca concentrations for the trees in an imaginary forest, randomly assigning values from a distribution with a mean of 5 mg/g and a standard deviation of 0.5 mg/g (Figure 1). In nature, we never know the true mean, but in this case, we created the imaginary forest with known concentration. We then randomly selected 12 trees for our sample, from which we took an imaginary sample of leaves and obtained a mean of 5.279 mg/g with a standard deviation of 0.477 and a standard error (SE) of 0.139 mg/g (Figure 1, solid black circle). The mean of a sample does not return the true population mean, which is important to the concept of the uncertainty in the mean. The SE describes the standard deviation of the distribution of estimates of the sample mean over different samples.

The number of trees to which our estimate will be applied is also important. To illustrate the effect of inventory size, we imagined our forest to have a density of 500 trees/ha, such that plots containing 10, 30, 50, 100, 1000, or 10,000 trees could be considered to represent areas of 0.02, 0.06, 0.1, 0.2, 2, or 20 ha. The plot area is not important to our estimates, but it helps convey what might be realistic numbers of trees to characterize for various

Box 2: Analytical solution for the uncertainty in the mean of a small sample: univariate case

We commonly represent uncertainty in the estimated mean value of a population of interest using the standard error of the mean. In some cases, we are also interested in the uncertainty in the prediction of individual values. In this paper, we have asked an unconventional question, but one that should be often relevant: What is the uncertainty in predicting the mean value of an attribute of a small number k of individuals when we are not confident of the true mean of the population? The answer depends in part on the uncertainty in the estimate of the mean of the population sampled, which would be smaller if it had been based on a larger sample n . It will also depend on the number of individuals, k , to which we apply it. In the body of this paper, we show this relationship using Monte Carlo simulation. In this Box, we provide the analytical solution.

We use n to describe the number of observations used to obtain an estimate of a population mean through the study A , in our case, of foliar calcium. The mean \bar{X}_A is $\frac{\sum_{i=1}^n X_{A,i}}{n}$, where $X_{A,i}$ is the concentration of calcium in the leaves of tree i . The sample variance σ^2 is $\frac{\sum_{i=1}^n (X_{A,i} - \hat{X}_{A,i})^2}{n-1}$ and the standard deviation, σ , is $\sqrt{\sigma^2}$. The uncertainty in the mean is $\frac{\sigma}{\sqrt{n}}$, called the standard error of the mean. This describes the spread of the discrepancy between our sample mean and the true population mean.

We use k to describe the number of individuals in the population from the inventory I to which we apply this estimate. What is the uncertainty of the estimated mean foliar calcium concentration of these k individuals? If we were confident of the true population mean (which we are not- we only have a sample, not the true value), the standard error of the mean foliar calcium for k trees would be $\frac{\sigma}{\sqrt{k}}$. To take account of the uncertainty in the estimated population mean, $\frac{\sigma}{\sqrt{n}}$, we combine these two independent terms, using the sum of the variances, $\frac{\sigma^2}{k} + \frac{\sigma^2}{n}$. This is called a normal mixture, meaning that the mean, not just the distribution around the mean, is also normally distributed. Taking the square root to find the standard error of the mean gives $\sigma\sqrt{\frac{1}{n} + \frac{1}{k}}$. This result does not take account of the uncertainty in our estimate of σ , which would require use of Student's t-distribution instead of the normal distribution or a Taylor expansion. Including it would only increase the uncertainty that we are pointing out has been underestimated in the past.

This formula reduces to $\frac{\sigma}{\sqrt{n}}$, the standard error of the mean, as k approaches infinity. For an individual ($k = 1$), it differs from the standard deviation σ : the uncertainty is $\sigma\sqrt{\frac{1}{n} + 1}$, which approaches σ with increasing n . When either n or k are small, as is often the case in forestry applications, it is important to use $\sigma\sqrt{\frac{1}{n} + \frac{1}{k}}$ to characterize the uncertainty rather than the standard error, $\frac{\sigma}{\sqrt{n}}$, or the standard deviation, σ .

purposes. We used the Monte Carlo approach (Figure 2, Box 3) to determine the uncertainty of the estimates. R code demonstrating these analyses is available (Drake and others 2023).

Using the estimated mean and standard deviation of our imaginary sample of 12 trees, we randomly sampled possible values of foliar Ca concentrations in trees for each of these various plot sizes, and we did this repeatedly to illustrate the uncertainty related to the prediction of individuals (left column of panels in Figures 3 and 4), the uncertainty related to the estimate of the mean (middle column, Figures 3 and 4), and the com-

bined uncertainty due to both the mean and individuals (right column, Figures 3 and 4).

When we predict the values of individual trees and average them within each iteration, there is considerable variation among iterations for small plot sizes (left column, 10 trees, Figure 3). The coefficient of variation (standard deviation divided by the mean) across the iterations is about 3%. As the plot size increases, however, the variation among iterations declines and eventually converges on our estimate of the mean (left column, 10,000 trees, Figure 3). Recall, however, that our estimate of the mean is not the true mean (Figure 1). This approach exaggerates our confidence in the esti-

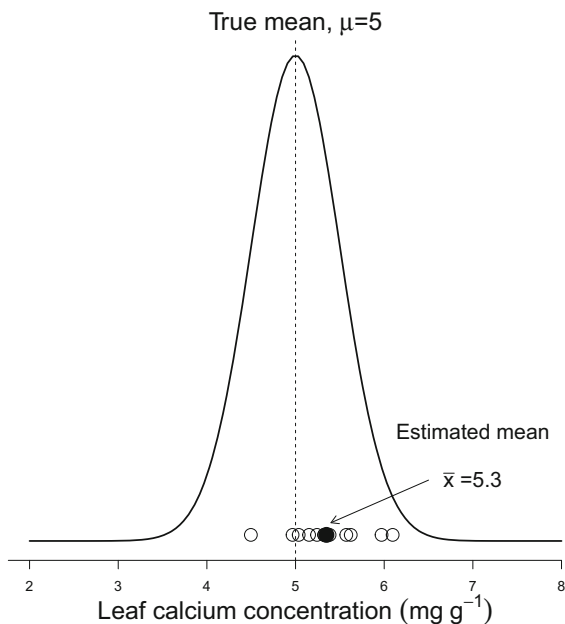


Figure 1. Illustration of sampling from a true underlying distribution. Suppose that there is a true underlying distribution of leaf calcium concentrations with a mean of 5 mg g^{-1} (solid black distribution with vertical dashed line illustrating the mean). This distribution is not observed directly. Rather, we collect samples from that distribution (open circles), which can be averaged to produce an estimate of the mean (solid black circle). In this case the estimated mean of 5.3 mg g^{-1} is higher than the true mean.

mate, as it ignores the uncertainty we have in the mean. We know that the average of the trees in the sample (5.279) was a poor estimate of the population mean, because we created the sample from an imaginary forest with a true mean concentration of 5.000.

Alternatively, we can represent uncertainty in the mean Ca concentration of the trees on a plot using the uncertainty in our estimate of the mean, described by the standard error of the mean. Here, we ignore variability among individuals; all the trees on a plot are assigned the same concentration at each iteration of the Monte Carlo, chosen randomly from a distribution defined by the mean and SE of our imaginary sample. Because individuals are not assigned different concentrations, the variation in the 10,000 iterations of the Monte Carlo is the same regardless of the number of trees in a plot (center column of panels in Figures 3 and 4). The uncertainty due to this source is about 3% of the mean, regardless of the number of trees. The uncertainties shown by the histograms in the figures are summarized using coefficients of variation in Table 1.

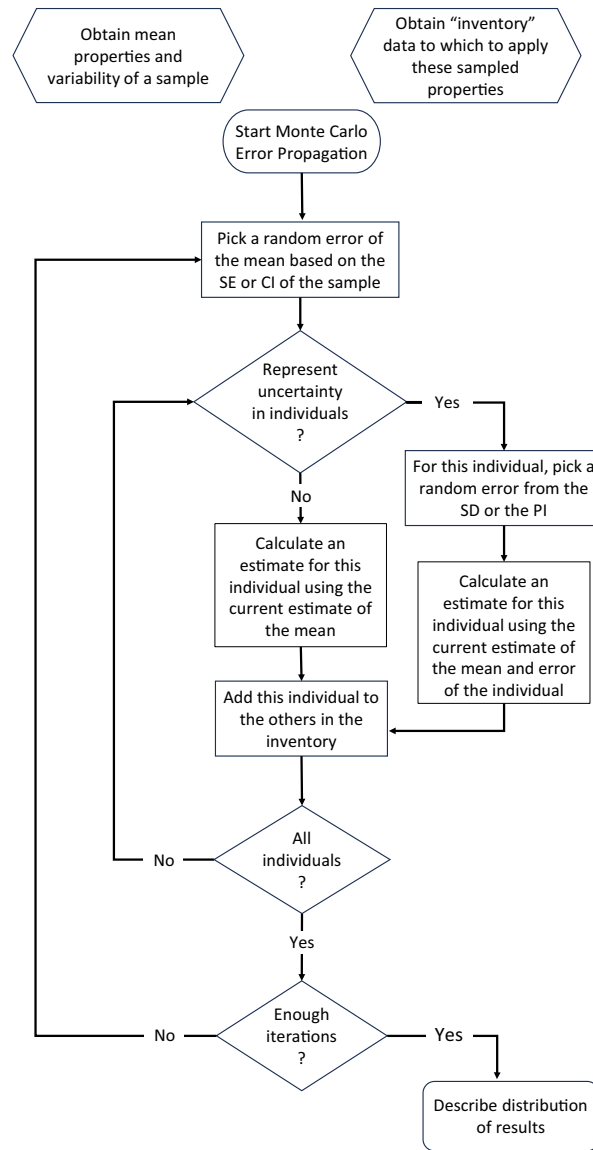


Figure 2. Flowchart of the steps in the Monte Carlo calculation of uncertainty in means, with or without accounting for uncertainty in individuals. The uncertainty in the mean is defined by the SE of the mean, in the univariate case, or, in the case of regression, the confidence interval (CI). The additional uncertainty in the prediction of an individual is defined by the SD of the mean or the prediction interval (PI).

Finally, we illustrate the uncertainty in the mean foliar Ca of trees on a plot as a function of plot size when both sources of uncertainty are accounted for. In each iteration of the Monte Carlo simulation, a random error in the mean is selected based on the SE of the sample, which applies to all the trees in the plot for that iteration, and an additional error is randomly sampled for each tree, based on the SD of the sample (right column of panels in

Box 3: A numerical approach: Monte Carlo Simulation

There are analytical solutions to simple cases of error propagation (Boxes 2 and 4). For example, the variance of a sum is the sum of the variances of the individual terms, if the terms can be assumed to be independent. In the case of forest biomass, however, the calculations can be too complex to be solved analytically. For example, a mixed species forest may be characterized using allometric models specific to each of several species. Tree tissues vary dramatically in concentration, and thus estimates of forest nutrient content are obtained using allometric models for the mass of each tissue type, multiplied by estimates of nutrient concentration for each tissue type. If a nutrient budget depends on 25 log–log regressions describing 5 tissue types (leaves, branches, stem wood, stem bark, and roots) and 5 species (for example, Whittaker and others 1979), a different approach to error propagation is needed (Yanai and others 2010).

A Monte Carlo simulation can be used to characterize uncertainty in a complex result by simple repetition, thanks to modern computing capability. An early application of the approach was the approximation of the value of π , estimated by dropping needles to find the fraction that cross a grid (Siniksaran 2008). More complex questions faced the physicists developing nuclear weapons at Los Alamos in 1946. Stanislaw Ulam was out sick, playing solitaire, and thinking about how to calculate the probability of winning. This would be extremely difficult to solve analytically, but it is not difficult, if one is home sick, to estimate by playing many games of solitaire following random shuffles of the deck. “Monte Carlo” was the secret code used to describe this approach during the development of the atomic bomb, which was a reference to Ulam’s uncle who liked to gamble at the Monte Carlo Casino in Monaco (Metropolis 1987).

The Monte Carlo approach to error propagation involves making a calculation many times, each time with a different random sample of the input values that reflects their uncertainty (the distribution of possible values). After many iterations, the distribution of the many results is used to characterize the uncertainty of the result due to the uncertainty in the inputs (for example Figs. 3–4). The Monte Carlo approach is easy to implement, but it is also easy to make mistakes in representing the uncertainty in the inputs and in deciding at what level to randomly select values. For example, if a root:shoot ratio is used across multiple forest types, a single value should be dealt to all the forest types at each iteration, or uncertainties in forest biomass will be underestimated (Yanai and others 2020). In the case addressed in this paper, uncertainty in individual predictions should be applied independently for each individual but uncertainty in the mean, or the confidence in a regression, should be applied simultaneously for all individuals at each iteration (each deal of the inputs).

Figures 3 and 4). With small numbers of trees, the uncertainty in the individual predictions contributes to the overall uncertainty: For 10 trees, ignoring either source of uncertainty gives a coefficient of variation of 3%, where the correct combined uncertainty is 4% (Table 1). The same result could be obtained by summing in quadrature: $3^2 + 3^2 = 4.2^2$ (the variance of a sum is the sum of the variances, if the variances are independent, and the variance is the square of the standard deviation). With large numbers of trees, uncertainty in the individuals is not important, and the estimates based on the uncertainty in the mean approach the correct value of $\sim 3\%$ (Figure 4, Table 1).

Uncertainty in Regression: The Mass of Leaves

The mass of trees and of tree tissues are usually predicted by allometric models, because measuring tree mass directly at the scale of a plot or a forest is impractical and destructive. Instead, tree diameters are measured and used to predict the mass of leaves, branches, bark, roots, and stem wood using allometric models, commonly based on a linear regression of log-transformed diameter and mass (Box 1). The predictions of these allometric models are not perfect, of course, and have uncertainty. To illustrate uncertainty in predictions of mass obtained by this method, we used data from 14 sugar maple trees that were cut down and weighed at the Hubbard Brook Experimental Forest, USA (Whittaker and others 1974). We fit a regression model predicting the logarithm of foliar mass from the

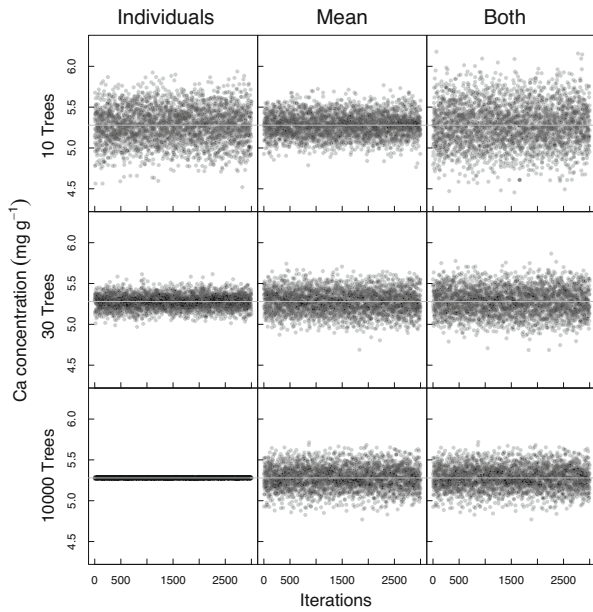


Figure 3. Monte Carlo simulation of uncertainty for a simple univariate example of leaf Ca concentration for 5, 30, or 10,000 trees. Monte Carlo sampling can be used to quantify the uncertainty in the prediction of an individual by randomly assigning a value to each individual based on the SD of the mean, the uncertainty in the estimate of the mean by randomly selecting a mean at each iteration of the Monte Carlo based on the SE of the mean, or both randomly selecting a mean at each iteration based on the SE and randomly selecting individual departures from the mean based on the SD.

logarithm of tree diameter (Figure 5) and obtained the same parameter estimates reported by Whittaker and others (1974). This model is analogous to estimating the mean in the case of calcium concentration (Figure 1) in that the parameter values of the regression model are estimates based on a sample. We used this regression model to predict the mass of leaves in plots with different numbers of trees:

$$\hat{Y}_{I,i} = 1.09 + 1.993 \cdot X_{I,i}$$

where $\hat{Y}_{I,i}$ is the estimate of \log_{10} (leaf biomass, in kg) and $X_{I,i}$ is \log_{10} (diameter, in cm) of tree i of inventory I . Summing the leaves on the plot requires back-transformation of logarithmic units, which incurs a bias (Baskerville 1972). For simplicity, we ignore this bias in this illustration. Another way to avoid bias is to characterize the relationship without the log transformation using a nonlinear model.

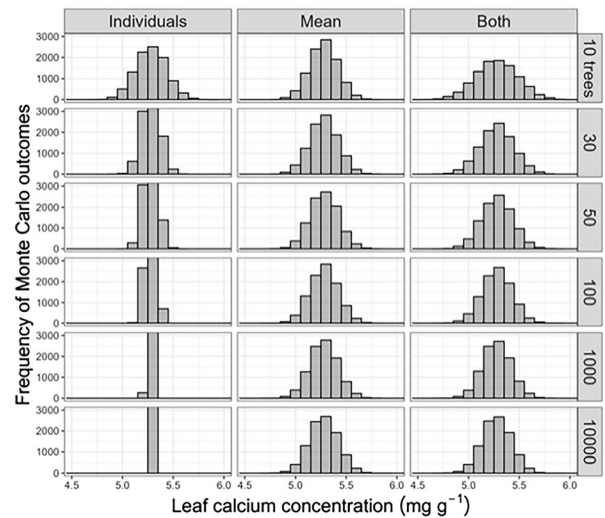


Figure 4. Estimates of uncertainty in the calcium concentration of the leaves of sugar maple trees in plot sizes ranging from 0.02 to 20 ha (10–10,000 trees). Each distribution is the result of 10,000 Monte Carlo iterations. For all the Monte Carlo simulations, we sampled from a distribution defined by a sample of 12 trees that had a mean foliar Ca concentration of 5.28, a SD of 0.48, and a SE of 0.14 mg/g. For the uncertainty in predicting an individual, we assigned each tree a concentration randomly selected from a distribution defined by the mean and the SD. For the uncertainty in predicting the mean, at each iteration we assigned all the trees in the plot the same Ca concentration randomly selected from a distribution defined by the mean and the SE. For the full uncertainty analysis, we simulated both sources of error: We randomly selected a mean for all the trees in the plot at each iteration, and we also randomly selected from the additional variability due to individual samples to predict the Ca concentration of each tree.

We illustrate the uncertainty in predicting individuals and uncertainty in the mean (referred to as the regression “model fit”) using Monte Carlo error propagation, just as we did for uncertainty in concentration. An analytical approach to combining these two sources of uncertainty is provided in Box 4. We created imaginary inventory data for plots containing 10, 30, 50, 100, or 1000 trees. We wanted each imaginary plot to have the same distribution of tree sizes, to avoid having different leaf masses per unit area for different plot sizes in our simulated results. So we selected 10 of the sugar maple trees in the Whittaker data set and used them 1, 3, 5, 10, or 100 times each.

For the uncertainty in the model fit, we randomly sampled values of an error term defined by Eq. (3) in Box 1. The same error term was applied to all the trees, until the next iteration of the Monte

Table 1. Uncertainty in the Calcium Concentration of Sugar Maple Foliage (Conc), the Mass of Leaves (Mass), and the Calcium Content of Leaves (Content) in Plot Sizes Ranging from 1 to 10,000 Trees (0.002–2 ha) Estimated by Three Different Approaches (Uncertainty in Individuals, Uncertainty in the Mean, and Both), as Indicated by the Coefficient of Variation (Standard Deviation Divided by the Mean) of the Monte Carlo Simulations Depicted in Figures 4 (Concentration), 6 (Mass), and 7 (Content)

# of trees	Uncertainty in individuals			Uncertainty in the mean			Both		
	Conc	Mass	Content	Conc	Mass	Content	Conc	Mass	Content
10	3	16	17	3	12	12	4	20	21
30	2	9	10	3	11	12	3	15	15
50	1	7	8	3	12	12	3	14	14
100	1	5	5	3	12	12	3	13	13
1000	0	2	2	3	12	12	3	12	12
10,000	0	1	1	3	12	12	3	12	12

We used 10,000 Monte Carlo iterations because there was visible variation due to random sampling when we used only 1000.

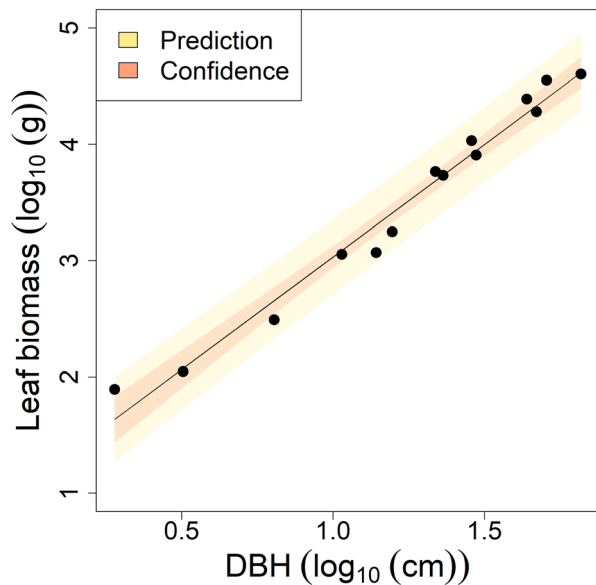


Figure 5. The diameter of a tree is a good predictor of the mass of its leaves. The log of leaf mass is a linear function of the log of tree diameter at breast height (DBH). The orange envelope shows the confidence interval, and the yellow shows the prediction interval (see Box 1).

Carlo, when a new error term was selected (Figure 2). This single random sample was retained for all trees within an iteration; if the allometric equation was biased high or low relative to the underlying true value, that bias would affect the estimates for all trees in the inventory. This procedure allowed us to quantify the uncertainty in model fit.

The uncertainty in the prediction of individuals is evaluated independently for each tree; thus, as the

number of trees on the plot increases, the uncertainty in the mean decreases (Figure 6), as was the case for the foliar concentration example (Figure 4). With a large number of trees, the uncertainty in the regression is underestimated, because each iteration of the Monte Carlo returns a similar result. In other words, all the estimates agree on the best-fit prediction based on the allometric sample of 14 trees, although the 14 trees do not perfectly characterize the population they represent. Obviously, this approach does not correctly describe the uncertainty in the result.

To include both sources of uncertainty, we added to the estimates in the Monte Carlo for the model fit a random sample of the standard error of the regression (Eq. 2 in Box 1). The results regarding the uncertainties of leaf mass (Figure 6) are visually similar to the results regarding leaf Ca concentration (Figure 4), but the uncertainties are larger (Table 1). For the smallest plot size (10 trees), the uncertainty of predicting individuals is the largest component, at 16%. At all inventory sizes, the uncertainty of predicting means is about 12%. The combined effect of the two sources is 20%, consistent with summing in quadrature ($16^2 + 12^2 = 20^2$). Propagating both uncertainties is worthwhile up to about 1000 trees, after which the uncertainty of predicting individuals is $< 1\%$ of the mean (Table 1).

Uncertainty in Nutrient Contents: Concentration Times Mass

Finally, we illustrate the Monte Carlo propagation of uncertainty in nutrient contents, which requires multiplying concentration and mass. For the calcium content of leaves on a plot, we used the foliar

Box 4: Analytical Solution for the Uncertainty in the Mean of a Small Sample: Regression

As in the case of a mean (Box 2), we can derive an analytical solution for the combined uncertainty in individuals and uncertainty in regression. Suppose we want to apply the regression in Eq. (1) to estimate the total biomass, or equivalently, the mean biomass per tree, in an inventory I of k trees. Let $X_{I,i}$ represent the measured log(diameter) of the i th tree in the sample of k trees in the area I , and $Y_{I,i}$ the true (as opposed to predicted) log(biomass) of the same tree. Note that these k trees differ from the n trees used in the A allometric study (Box 1). Then we have,

$$Y_{I,i} = \hat{a} + \hat{b}X_{I,i} + \sigma Z_{I,i}, \quad i = 1, 2, \dots, k \quad (4)$$

Here the random standard normal variates $Z_{I,i}$ may be assumed independent from one tree to the next. Averaging both sides of Eq. (4) over i from 1 to k , we obtain

$$\bar{Y}_I = \hat{a} + \hat{b}\bar{X}_I + \sigma Z_0/\sqrt{k}, \quad (5)$$

where Z_0 represents the standard normal variate that arises after averaging the $Z_{I,i}$ in Eq. (5).

Let μ be the mean population parameter of log-transformed DBH over the entire area, a quantity that is not observed directly, but which is estimated by \bar{X} . Indeed, we may again posit an exact probabilistic relationship of the form

$$\bar{X}_I = \mu + \sigma_X Z_1/\sqrt{k}, \quad (6)$$

where Z_1 is another standard normal, independent of Z_0 , and σ_X is the standard deviation of the population of X_I values.

By substituting Eq. (6) into Eq. (5) we obtain:

$$\bar{Y}_I = \hat{a} + \hat{b}\mu + \hat{b}\sigma_X Z_1/\sqrt{k} + \sigma Z_0/\sqrt{k}$$

We want to quantify the uncertainty when \bar{Y}_I is used as an estimate of $E(\bar{Y}_I) = a + b\mu$, the true average log-biomass per tree in the area. To do that, we use a development based on mixture distributions. Let (X, Y) be jointly distributed random variables. Then $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$, and

$$\text{Var}(\hat{a} + \hat{b}\bar{X}_I) = E\left(\text{Var}(\hat{a} + \hat{b}\bar{X}_I|\bar{X}_I)\right) + \text{Var}\left(E(\hat{a} + \hat{b}\bar{X}_I|\bar{X}_I)\right) \quad (7)$$

Since \hat{a} and \hat{b} are unbiased and independent of \bar{X}_I , $E(\hat{a} + \hat{b}\bar{X}_I|\bar{X}_I) = E(\hat{a}) + E(\hat{b})\bar{X}_I = a + b\bar{X}_I$. Thus, the *second term* in Eq. (7) becomes $\text{Var}(a + b\bar{X}_I) = b^2 \text{Var}(\bar{X}_I) = b^2 \sigma_X^2/k$.

Using standard regression formulas for variance and covariance of the regression coefficients requires some care, since the \bar{X} in those formulas is different from the average of the sample in the inventoried plot. Defining the sum of squares $SS(X_A) = \sum_{i=1}^n (X_{A,i} - \bar{X}_A)^2$. Thus, the *first term* in Eq. (7), defining the variance of the *estimator*, can be derived through Eq. (1): $\text{Var}(\hat{a} + \hat{b}\bar{X}_I|\bar{X}_I) = \sigma^2 \left(\frac{1}{n} + \frac{(\bar{X}_I - \bar{X}_A)^2}{SS(X_A)} \right)$.

Since $\frac{E(\bar{X}_I - \bar{X}_A)^2}{SS(X_A)} = \frac{E(\bar{X}_I)^2 + 2\bar{X}_A E(\bar{X}_I) + \bar{X}_A^2}{SS(X_A)} = \frac{\sigma_X^2}{kSS(X_A)} + \frac{(\mu - \bar{X}_A)^2}{SS(X_A)}$, the first term of Eq. (7) is $\frac{\sigma^2}{n} + \frac{\sigma^2 \sigma_X^2}{kSS(X_A)} + \frac{\sigma^2 (\mu - \bar{X}_A)^2}{SS(X_A)}$, and the uncertainty of the *estimator*, according to Eq. (7), is $\text{Var}(\hat{a} + \hat{b}\bar{X}_I) = \frac{\sigma^2}{n} + \frac{\sigma^2 \sigma_X^2}{kSS(X_A)} + \frac{\sigma_X^2 b^2}{k} + \frac{\sigma^2 (\mu - \bar{X}_A)^2}{SS(X_A)}$. Using Eq. (5), the uncertainty of the *estimation* results in $E(\bar{Y}_I - a - b\mu)^2 = \text{Var}(\bar{Y}_I) = \text{Var}(\hat{a} + \hat{b}\bar{X}_I) + \text{Var}(\sigma Z_0/\sqrt{k})$, and substituting,

$$E(\bar{Y}_I - a - b\mu)^2 = \frac{\sigma^2}{n} + \frac{\sigma^2}{k} + \frac{\sigma^2 \sigma_X^2}{kSS(X_A)} + \frac{\sigma_X^2 b^2}{k} + \frac{\sigma^2 (\mu - \bar{X}_A)^2}{SS(X_A)} \quad (8)$$

The uncertainty of the estimator $\hat{a} + \hat{b}\bar{X}_I$, i.e., its variance, should be distinguished from the uncertainty of the *estimation*. The former is the same as Eq. (7), but without the second term. Only the first two terms appear in the non-regression form of this problem considered in Box 2. In use, one must replace the (unknown) parameter values σ, σ_X, μ , and b with their estimators. Also, one would normally take the square root of Eq. (8) as the final estimate of the uncertainty.

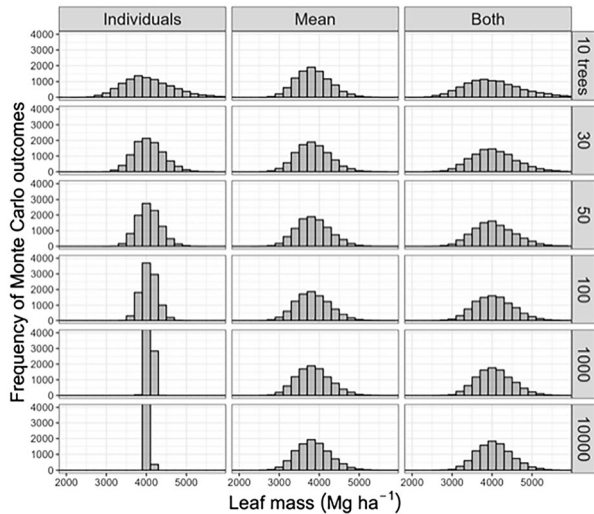


Figure 6. Estimates of uncertainty in the mass of sugar maple leaves in plot sizes ranging from 0.02 to 20 ha (10–10,000 trees). Each distribution is the result of 10,000 Monte Carlo iterations. For all the Monte Carlo simulations, we sampled from the uncertainty in a regression of leaf mass as a function of stem diameter. For the uncertainty in predicting an individual, each tree was assigned a random prediction error (Eq. 2). For the uncertainty in predicting the mean, all the trees in the plot were assigned the same random sample of the error in the regression (Eq. 3) for each iteration of the Monte Carlo. For the full uncertainty analysis, we simulated both sources of error: We randomly selected the error in the regression for all the trees in the plot at each iteration, and we also randomly selected a residual error to determine the foliar mass of each tree. The y axis is scaled to best show the distributions; the most narrow are cut off at the top.

calcium concentrations (Figure 4) and multiplied them by the foliar masses of each tree (Figure 6), running through all the trees on the plot in each of 10,000 Monte Carlo iterations, to obtain the uncertainty in estimates of plot-level foliar calcium content (Figure 7). Again, we see that ignoring uncertainty in the mean gives incorrectly small uncertainties, especially for large inventories (Figure 7). The uncertainties for calcium content are numerically very similar to those for mass (Table 1)

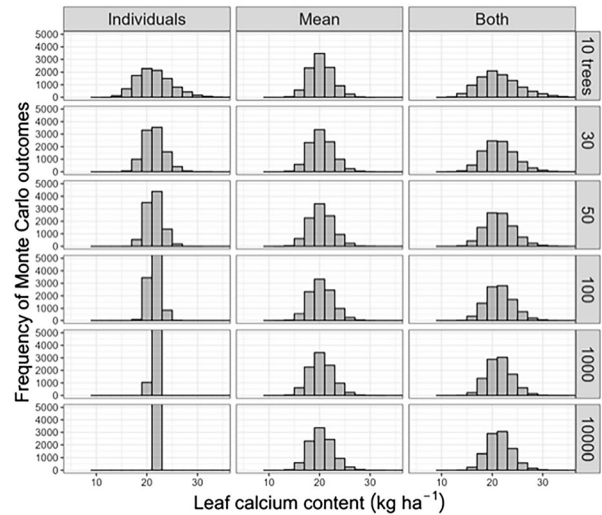


Figure 7. Estimates of uncertainty in the calcium content of sugar maple leaves in plot sizes ranging from 0.02 to 20 ha (10–10,000 trees). Each distribution is the result of 10,000 Monte Carlo iterations. The calcium content of leaves on the plot is the sum of the calcium contents of all the trees, which is obtained by multiplying the foliar mass (Figure 4) times the calcium concentration (Figure 2). For the uncertainty in predicting an individual, we ignored uncertainty in the mean. For the uncertainty in predicting the mean, we ignored uncertainty in the individuals. For the full uncertainty analysis, we included both sources of error. The y axis is scaled to best show the distributions; the most narrow are cut off at the top.

because the contribution of uncertainty in concentration was relatively small.

The approach illustrated here can be adapted to quite complex calculations. To estimate the calcium contents of trees in a mixed species forest requires estimates of concentration and biomass of multiple tissue types (leaves, branches, bark, wood, and roots), across multiple species. We did this calculation for the reference watershed at Hubbard Brook using allometric models (Whittaker and others 1974) and concentrations of calcium (Likens and Bormann, 1970) for 7 tissue types of 6 species. The 13-ha watershed was divided into 208 0.0625-ha plots, and in a 0.01-ha subplot of each plot,

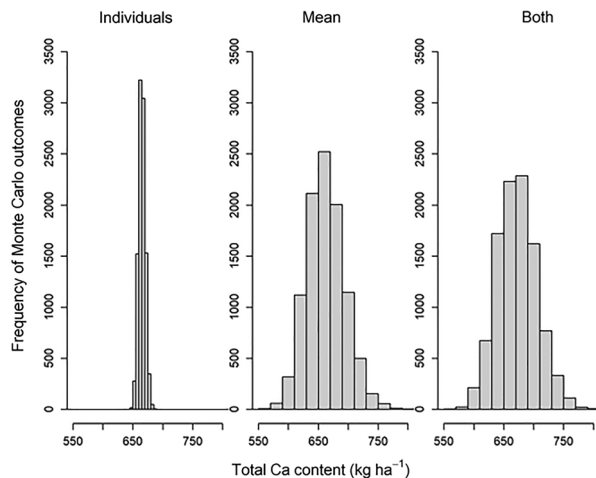


Figure 8. Estimates of uncertainty in the calcium content of trees in the reference watershed at the Hubbard Brook Experimental Forest, based on allometry and calcium concentrations of leaves, branches, bark, wood, and roots of 6 tree species and an inventory of 3990 trees.

species and diameter of all trees > 2 cm dbh were recorded (Whittaker and others 1974). Fifteen species were tallied, and those not included in the allometric and chemical data sets were represented by species thought to be similar. These calculations are available as an Excel workbook (Lilly and others 2023). In this case, with a total of 3990 trees, the uncertainty in forest calcium stocks associated with prediction of individuals was 0.8%, the uncertainty in the mean was 4.8%, and including both resulted in an uncertainty of 4.9% (Figure 8). Thus, in this case with an extensive inventory of many trees, the uncertainty in the mean was nearly equivalent to the uncertainty of both sources together. Although the uncertainty of both sources together is always higher, it would be only infinitesimally higher with an infinite number of trees. Thus, instances with very large inventories can likely ignore the uncertainties of individuals.

DISCUSSION

It is common to describe uncertainty of forest-scale estimates using the SE of the mean (or of the regression) and to describe the distribution of individual observations using the SD (of the residuals, in the case of regression). It is less common to recognize situations in which both sources of uncertainty are important. Here we have shown that uncertainty in individuals is important, in addition to uncertainty in mean properties, when the number of individuals is small. Thus, when

experimental treatments involve small numbers of trees, it would be wise to include uncertainty in individuals in error propagation. At the other extreme, when thousands of trees are involved, uncertainty is grossly underestimated if uncertainty in the mean is omitted from error propagation. An example of this from the remote sensing field resulted in an estimate of forest carbon with $< 1\%$ uncertainty, despite using an allometric model with considerable uncertainty (Gonzalez and others 2010). In reporting carbon emissions or emission reductions for climate mitigation at the scale of entire countries, uncertainty in individuals can safely be ignored.

Whether uncertainty in individuals is likely to be negligible depends on the specifics of the case and the number of trees in the inventory. Four contrasting forest types were evaluated for allometric uncertainty in estimates of forest biomass (Lin and others 2023), and the four case studies differed in the relative importance of uncertainty in predicting individuals. The greatest uncertainty in predicting individuals was in a semi-arid site with multi-stemmed trees, where the model fit was poor. Small uncertainties were observed where model fit was good, as was the case in a monoculture plantation and in a subtropical jungle with hundreds of trees contributing to the allometric model. In the example we developed in this paper, based on data from the Hubbard Brook Experimental Forest, the number of trees needed for uncertainty in the prediction of individuals to be smaller than uncertainty in the mean was less for calcium concentration (about 10 trees) than for foliar mass or calcium contents (closer to 30 trees) (Table 1). The uncertainty in predicting individuals was less than 1% of the mean with only 50 trees for foliar concentration but with 10,000 trees for foliar mass or calcium content (Table 1). It will always be most correct, but sometimes by a very small margin, to include both sources of uncertainty.

There are other ways to represent uncertainty in regression models than the approach represented here, which is based on Monte Carlo sampling (Box 3) of uncertainty derived from parametric statistics (Box 1). Bootstrapping is an approach that involves refitting the model to random samples of the data (with the same sample size). Another approach is to randomly sample values of the model parameters (the slope and intercept), accounting for the covariance between them. Bayesian approaches estimate the uncertainty in model parameters using probability distributions. All four approaches give similar results (Lin and others 2023), except that bootstrapping may result in

greater uncertainty if the allometric sample size is small and includes outliers. Thus, the choice of approach can be made on practical considerations such as user familiarity. Our final example, which was the most complex, was conducted in Excel, with the aid of macros to attain 10,000 iterations (Lilly and others 2023). The others were coded in R (Drake and others 2023).

Analytical approaches to error propagation (Boxes 2 and 4) are easier to implement than numerical approaches when the calculations are simple. When they are complex, as is often the case for ecosystem budgets and country-level accounting of carbon emissions, a Monte Carlo approach is more attractive. Importantly, the Monte Carlo approach does not require any assumptions about the distributions of the inputs, whereas the analytical solution depends on the inputs being normally distributed. Our Monte Carlo results agreed with results of the analytical approach in the case of calcium concentrations, which we sampled from a normal distribution, but not in the case of leaf biomass, for which we used 10 trees from the Whittaker data set. The disagreement is greatest when the number of trees is small and their variability is high (σ_X , Box 4).

There are many other sources of uncertainty in estimating carbon and nutrient storage in forests besides the uncertainty in allometric models. For deforestation, forest degradation, and forest growth, the greatest source of uncertainty is the estimation of the area mapped as forest, when these are based on remote sensing (Esteban and others 2020; Neeff 2021). In plot-based national forest inventories, sampling error is the most important source, which reflects spatial variation. This source of uncertainty, characterized by the SE of the estimate, depends on the variability across sample plots and the number of sample plots, which can be designed to attain a target confidence. Lesser sources of uncertainty include the root-to-shoot ratio, when belowground biomass is estimated from aboveground biomass, the wood density, when allometric models provide volume, and the carbon fraction of biomass (McRoberts and others 2016). The uncertainty in allometric models may be among the more important of these lesser sources.

The uncertainty in allometric models is not limited to the uncertainty in the model: In most cases, there are a variety of possible models to select, each of which would give a different answer (Melson and others 2011; Picard and others 2015). Thus, model selection error is a source of uncertainty in forest budgets. In addition, the selection of trees for

allometric models may induce a bias: Trees may be selected for good form, omitting those with damaged crowns, forks, or stem rot, and thus the models are not based on a representative sample of the population to which they will be applied. These sources of error, in which the model does not accurately describe the trees to which it is applied, are more difficult to quantify than the error in the model, which is the source addressed in this paper.

Reporting uncertainty is important, not only in forest accounting, but in all endeavors in which uncertainty is high. In environmental sciences, uncertainty is not reported as often as it should be. Based on a random sample of 139 papers published in 2019 (Yanai and others 2021), fewer than half of eligible sources were reported, with sampling error the most often reported (for example, in 84% of vegetation studies). Only four papers in the sample used biomass models; none of them reported model uncertainty (Yanai and others 2021). In country-level carbon accounting, rates of uncertainty reporting are improving. Since 2018, at least 50% of the national reference levels reported to the United Nations Framework Convention on Climate Change have propagated uncertainty in estimates of forest carbon emissions, whereas from 2014 to 2017, rates ranged from 0 to 40% (Yanai and others 2020). Whether these uncertainties are correctly quantified is another matter. Since payments for reducing emissions from deforestation and forest degradation (REDD) depend on the reported uncertainties in emission reductions, there are financial incentives to underestimate them. We hope that this paper will help increase the accuracy of uncertainty reporting in forest accounting, for purposes ranging from research and forest management to carbon finance for climate mitigation.

ACKNOWLEDGEMENTS

Terry McConnell provided remedial statistics instruction and the derivations presented in Boxes 2 and 4. Joe Nash adapted the Excel model of forest nitrogen at Hubbard Brook for calcium and our three scenarios of uncertainty in concentration and biomass. Ron McRoberts provided useful criticism of an earlier draft of this paper. This publication is a product of QUEST (Quantifying Uncertainty in Ecosystem Studies), a working group dedicated to advancing uncertainty analysis in ecosystem studies (www.quantifyinguncertainty.org) and QUERCA (Quantifying Uncertainty Estimates and Risk for Carbon Accounting), which is funded by the US Department of State and US Agency for International Development. Please visit our website at www.QUEST.org

www.quantifyinguncertainty.org for papers, sample code, presentations, tutorials, and discussion.

FUNDING

This research was funded by grants from the National Science Foundation for a Research Coordination Network (DEB-1257906) and the U.S. Department of State (20-DG-11132762-304).

DATA AVAILABILITY

R code demonstrating these analyses is available at https://github.com/jedrake/Uncertainty_individuals_means and an Excel file is available at <https://doi.org/10.6084/m9.figshare.21937235.v1>.

Declarations

Conflict of interest The authors declare no conflict of interest.

OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Baskerville GL. 1972. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* 2(1):49–53. <https://doi.org/10.1139/x72-009>.
- Bogardus ST Jr, Holmboe E, Jekel JF. 1999. Perils, pitfalls, and possibilities in talking about medical risk. *J. Am. Med. Assoc.* 281(11):1037–41.
- Breidenbach J, McRoberts RE, Astrup R. 2016. Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. *Remote Sens. Environ.* 173:274–81.

- Drake, J.E, H. Buckley, B. Case, R. Yanai. 2023. Github repository regarding the uncertainty of individuals, means, and both combined. https://github.com/jedrake/Uncertainty_individuals_means
- Draper NR, Smith H. 1998. Applied regression analysis. New York: Wiley.
- Esteban J, McRoberts RE, Fernández-Landa A, Tomé JL, Marchamalo M. 2020. A model-based volume estimator that accounts for both land cover misclassification and model prediction uncertainty. *Remote Sens.* 12(20):3360. <https://doi.org/10.3390/rs12203360>.
- Fahey TJ, Siccama TG, Driscoll CT, Likens GE, Campbell J, Johnson CE, Battles JJ, Aber JD, Cole JJ, Fisk MC, Groffman PM. 2005. The biogeochemistry of carbon at Hubbard Brook. *Biogeochemistry* 75:109–76. <https://doi.org/10.1007/s10533-004-6321-y>.
- Falster DS, Duursma RA, Ishihara MI, Barneche DR, FitzJohn RG, Vårhammar A, Aiba M, Ando M, Anten N, Aspinwall MJ, Gargaglione VB. 2015. BAAD: A biomass and allometry database for woody plants. *Ecol Soc Am.* <https://doi.org/10.1890/14-1889.1>.
- Gonzalez P, Asner GP, Battles JJ, Lefsky MA, Waring KM, Palace M. 2010. Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. *Remote Sens. Environ.* 114(7):1561–75. <https://doi.org/10.1016/j.rse.2010.02.011>.
- Heng A, Zhang S, Tan AC, Mathew J. 2009. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mech. Syst. Signal Process.* 23(3):724–739.
- Holdaway RJ, McNeill SJ, Mason NW, Carswell FE. 2014. Propagating uncertainty in plot-based estimates of forest carbon stock and carbon stock change. *Ecosystems* 17:627–40. <https://doi.org/10.1007/s10021-014-9749-5>.
- Horsley SB, Long RP, Bailey SW, Hallett RA, Hall TJ. 2000. Factors associated with the decline disease of sugar maple on the Allegheny Plateau. *Can. J. For. Res.* 30(9):1365–78. <https://doi.org/10.1139/x00-057>.
- Keith H, Vardon M, Obst C, Young V, Houghton RA, Mackey B. 2021. Evaluating nature-based solutions for climate mitigation and conservation requires comprehensive carbon accounting. *Sci. Total Environ.* 769:144341. <https://doi.org/10.1016/j.scitotenv.2020.144341>.
- Likens GE, Bormann FH. 1970. Chemical analyses of plant tissues from the Hubbard Brook ecosystem in New Hampshire.
- Lilly PJ, Nash JM, Drake JE, Yanai RD. 2023. S1_Allometric uncertainty HBEF.xlsm. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.21937235.v1>
- Lin J, Gamarra JGP, Drake JE, Cuchietti A, Yanai RD. 2023. Scaling up uncertainties in allometric models: How to see the forest, not the trees. *For. Ecol. Manag.* 537:120943. <https://doi.org/10.1016/j.foreco.2023.120943>.
- McRoberts RE, Chen Q, Domke GM, Ståhl G, Saarela S, Westfall JA. 2016. Hybrid estimators for mean aboveground carbon per unit area. *For. Ecol. Manag.* 378:44–56. <https://doi.org/10.1016/j.foreco.2016.07.007>.
- Melson SL, Harmon ME, Fried JS, Domingo JB. 2011. Estimates of live-tree carbon stores in the Pacific Northwest are sensitive to model selection. *Carbon Bal. Manag.* 6:1–6. <https://doi.org/10.1186/1750-0680-6-2>.
- Metropolis N. 1987. The beginning of the Monte Carlo method. *Los Alamos Science. Los Alamos Sci. Special Issue* 15:125–30.

- Neeff T. 2021. What is the risk of overestimating emission reductions from forests—and What can be done about it? *Climat. Change* 166(1–2):26. <https://doi.org/10.1007/s10584-021-03079-z>.
- Paré D, Bernier P, Lafleur B, Titus BD, Thiffault E, Maynard DG, Guo X. 2013. Estimating stand-scale biomass, nutrient contents, and associated uncertainties for tree species of Canadian forests. *Can. J. For. Res.* 43(7):599–608.
- Picard N, Boyemba Bosela F, Rossi V. 2015. Reducing the error in biomass estimates strongly depends on model selection. *Ann. For. Sci.* 72:811–23. <https://doi.org/10.1007/s13595-014-0434-9>.
- Picard N, Saint-André L, Henry M. 2012. Manual for building tree volume and biomass allometric equations: From field measurement to prediction. Food and Agricultural Organization of the United Nations, Rome, and *Centre de Coopération Internationale en Recherche Agronomique pour le Développement*, Montpellier, p. 215.
- Siniksaran E. 2008. Throwing Buffon's needle with Mathematica. *Math J* 11(1):71–90. <https://doi.org/10.1017/mag.2020.117>.
- Whittaker RH, Bormann FH, Likens GE, Siccama TG. 1974. The Hubbard Brook ecosystem study: Forest biomass and production. *Ecol. Monogr.* 44(2):233–54. <https://doi.org/10.2307/1942313>.
- Whittaker RH, Likens GE, Bormann FH, Easton JS, Siccama TG. 1979. The Hubbard Brook ecosystem study: Forest nutrient cycling and element behavior. *Ecology* 60(1):203–20.
- Wlezien C, Jennings W, Fisher S, Ford R, Pickup M. 2013. Polls and the vote in Britain. *Polit. Stud.* 61:66–91.
- Yanai RD, Battles JJ, Richardson AD, Blodgett CA, Wood DM, Rastetter EB. 2010. Estimating uncertainty in ecosystem budget calculations. *Ecosystems* 13:239–48. <https://doi.org/10.1007/s10021-010-9315-8>.
- Yanai RD, Levine CR, Green MB, Campbell JL. 2012. Quantifying uncertainty in forest nutrient budgets. *J. For.* 110(8):448–56. <https://doi.org/10.5849/jof.11-087>.
- Yanai RD, Wayson C, Lee D, Espejo AB, Campbell JL, Green MB, Zuskwert JM, Yoffe SB, Aukema JE, Lister AJ, Kirchner JW. 2020. Improving uncertainty in forest carbon accounting for REDD+ mitigation efforts. *Environ. Res. Lett.* 15(12):124002. <https://doi.org/10.1088/1748-9326/abb96f>.
- Yanai RD, Mann TA, Hong SD, Pu G, Zuskwert JM. 2021. The current state of uncertainty reporting in ecosystem studies: A systematic evaluation of peer-reviewed literature. *Ecosphere* 12(6):e03535. <https://doi.org/10.1002/ecs2.3535>.
- Yanai RD, Young AR, Campbell JL, Westfall JA, Barnett CJ, Dillon GA, Green MB, Woodall CW. 2022. Measurement uncertainty in a national forest inventory: Results from the Northern Region of the USA. *Can J For Res.* <https://doi.org/10.1139/cjfr-2022-006>.
- Yang Y, Yanai RD, Fatemi FR, Levine CR, Lilly PJ, Briggs RD. 2016. Sources of variability in tissue chemistry in northern hardwood species. *Can. J. For. Res.* 46(3):285–96. <https://doi.org/10.1139/cjfr-2015-0302>.
- Zaimovic A, Omanovic A, Arnaut-Berilo A. 2021. How many stocks are sufficient for equity portfolio diversification? A review of the literature. *J. Risk Financ. Manag.* 14(11):551.