Ecological Monographs, 75(4), 2005, pp. 451–466

GAUSSIAN ERROR PROPAGATION APPLIED TO ECOLOGICAL DATA: POST-ICE-STORM-DOWNED WOODY BIOMASS

ERNEST LO[1]

Groupe de Recherche en Ècologie Forestière inter-universitaire (GREFi), Department of Biology, McGill University, 1205 Docteur Penfield, Montréal, Québec H3A 1B1 Canada

Abstract. Error analysis using Gaussian error propagation (GEP) can be used to analytically determine the error or uncertainty produced by multiple and interacting measurements or variables. The technique is especially useful for studies that involve step-by-step calculations, where measurements taken at a smaller temporal or spatial scale are used to estimate a value at larger scales (e.g., daily total tree-crown carbon assimilation is estimated from carbon assimilation rate per unit leaf area per unit time).

The GEP technique is not well known and rarely used in ecology. The purpose of this paper is to illustrate the concepts and methods of GEP in a manner that is accessible and relevant to students and researchers in ecology. The technique is also extended to calculate the ''error budget'' and ''sensitivity indices'' of error sources. The concept of the ''error structure'' of an experiment or calculation is introduced, and different partitioning methods and optimization strategies for analyzing and reducing error, which further develop the potential usefulness of GEP, are shown. An example of its application to ecological data is demonstrated using the post-ice-storm-downed woody-biomass data set, previously reported by M. C. Hooper, K. Arii, and M. J. Lechowicz. Both the data and the error analysis can be viewed as being representative of and relevant to a general class of step-by-step and scaling-up ecological calculations. Finally the use of GEP reveals that the error structure is a scale-dependent quantity, a result that is relevant to both scaling theory and experimental design.

Key words: biomass estimation; error analysis; experimental design; forest disturbance; Gaussian error propagation; scaling theory; sensitivity analysis; Taylor expansion.

INTRODUCTION

The importance of error propagation (which is the uncertainty produced by the calculations or systems that combine errors from multiple sources) has been recognized in ecological science (e.g., Brewer and Gross 2003). In the author's experience, however, awareness of error-analysis methods to assess error propagation, and the inclination to consider and apply such techniques, seem to be restricted in ecology to modeling ''experts'' or technically minded specialists. This paper presents a concise but thorough description of a powerful error-analysis method, that of Gaussian error propagation (GEP). Basic GEP has been used for error analysis at least since the 1960s, and is taught as a standard data-analysis technique in undergraduate physics. However it has remained largely unknown and inaccessible in the field of ecology. When used in either domain, the method is applied only in its most basic form and its potential as a tool to analyze and partition the structure of error has not been fully realized. Therefore this paper attempts firstly to explain and illustrate the method from an introductory point of view, in order to encourage a more general audience of ecology researchers to consider using GEP in their work. Secondly, the GEP technique is extended both theoretically and in its methods of application. I demonstrate the latter error analysis calculations with the goal of encouraging others to perform similar or related analyses.

Manuscript received 12 January 2005; accepted 12 April 2005; final version received 26 May 2005. Corresponding Editor: S. R. Lele.

[1] Present address: 211 Avenue Mont-Royal Ouest, Apt. 2, Montréal, Québec H2T 2T2 Canada. E-mail: treemake@gmail.com

The context of the GEP method in statistical inference

Statistical inference refers to the process by which hypotheses, ideas, or mechanisms (collectively referred to as ''models'') are formulated and tested against real data. Chatfield (1995) classifies errors in statistical inference into three types: uncertainty in model structure (e.g., whether or not to accept a hypothesis), uncertainty in model parameters (e.g., estimation of true means from finite samples), and the unexplained random variation in observed variables in a model whose structure and parameters are known. The first two types of error arise from procedures commonly termed ''model selection'' and ''parameter estimation.'' The third category of uncertainty represents what is typically described as ''precision error'' (Clifford 1973, Sokal and Rohlf 1981, Taylor 1982, Zar 1996). The identification of this type of error is important since it may be desirable to evaluate or reduce this error with-

out necessarily addressing issues of model or parameter correctness. All applications of GEP in this article pertain to this precision error.

There are presently two main paradigms in statistical inference, that of classical or frequentist statistics, and that of Bayesian inference. The current study falls under the former paradigm in that all the error quantities are interpreted as Gaussian confidence intervals about theoretically known parameter values in a specified model. Bayesian inference, in contrast, estimates posterior probability distributions for models and parameters of interest, using prior distributions and the rules of conditional probability (Ellison 1996). These parameters and models do not in principle have ''true'' values and are considered random variables. Thus Bayesian inference is concerned with the first and second categories of uncertainty listed in the preceding paragraph. Therefore the current study differs from a Bayesian context in that there are no conditional probabilities, parameters and models are interpreted differently, and a different category of statistical uncertainty is being addressed. In general, however, GEP as a calculation tool is not incompatible with Bayesian inference, as I discuss later.

The context of the GEP method in error analysis

Statistical dispersion around known values (from hereon referred to as ''error'') has many sources. These include measurement error, confidence intervals in parameters calculated from statistical analyses, variability produced by stochastic processes (e.g., within computer simulation models), and error that is propagated by the interaction or combination of the above three sources. The term ''error analysis'' itself has been widely used; here it is taken to comprise three aspects–the calculation of the magnitude of error(s), the identification of the major sources of error in an experiment or calculation, and lastly the determination of optimal strategies to reduce the error. Knowledge of the magnitude of error permits the establishment of confidence intervals and the meaningful comparison of ecological quantities. The assurance that the error levels are acceptable is essential to the design of feasible experiments and calculations. Finally, the quantification of error is important, as the component errors in any system often interact or propagate to produce an unexpectedly large error in the final result.

The primary (and more common) method for determining error propagation in the ecology literature has been Monte Carlo type analyses (e.g., Håkanson 1999). The term ''Monte Carlo'' is used in this article to refer to bootstrapping-type analyses to estimate confidence intervals. They are often used to evaluate computer models (e.g., Pacala et al. 1996, Magnussen and Burgass 1997, Deutschman et al. 1999), and entail repeated computations of output variables as a function of random variations in all combinations (or in a case-by-case analysis) of input variables in order to determine propagated confidence intervals. These studies are illustrative of some important aspects of error analysis. They identify the magnitude of error that has been propagated to output variables, and attempt to identify the effect and importance of specific input variables. Disadvantages of this type of analysis are that they are computationally intensive and not amenable to detailed analyses of error structure. The GEP method described here represents a more powerful and flexible method. As will be shown, the formulae for the error produced by GEP can be subjected to detailed analysis to determine the effect of input variables on the error magnitude, to find the conditions for minimum error, and to optimize an experimental design or calculation structure. These analyses extend the application of GEP over previous studies (e.g., Oostrom et al. 1995, Beaudet et al. 2000). An important constraint of the method is that the output variables under consideration must be expressible as an analytical equation (i.e., a mathematical formula) of the input variables. However, this requirement can often be satisfied, especially in the case of step-by-step calculations that are intended to compute ecological quantities at larger scales from measurements made at smaller scales.

The present study applies and extends the GEP method of error analysis using an ecological data set of downed woody biomass in an old-growth forest. After the basic use of GEP is explained, the method is extended and applied in different ways to examine the ''error structure'' of the data. This analysis includes the derivation of error budgets and sensitivity indices, as well as different error-partitioning strategies. Optimal ways in which the experiment could have been modified to minimize the error are identified. Lastly, the effect of spatial scale on the error structure is studied, and the resulting scaling behavior is analyzed.

Theory: Description of the Gaussian Error-Propagation Equations

The simplest form of error propagation occurs when the input variables or observables are statistically independent or uncorrelated. In this case, the error in the input variables that is *propagated* to the output variable is described by

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x_1}\sigma_{x1}\right)^2 + \left(\frac{\partial q}{\partial x_2}\sigma_{x2}\right)^2 + \cdots + \left(\frac{\partial q}{\partial x_n}\sigma_{xn}\right)^2 = \sum_{i=1}^{n}\left(\frac{\partial q}{\partial x_i}\sigma_{xi}\right)^2 \quad (1)$$

where $x_1, x_2, \ldots x_n$ represent the n input variables and $\sigma_{x1}, \sigma_{x2}, \ldots \sigma_{xn}$ are their associated errors (e.g., standard errors). The output variable q is a function of the input variables [i.e., $q = f(x_1, x_2, \ldots x_n)$], and the error in q is denoted by σ_q. Eq. 1 shows that the square of the output error is equal to the sum of the squares of the input variables weighted by a partial derivative of the output function with respect to the corresponding

input variable. Therefore if and once the partial derivatives are known, the output error can then be evaluated from the input errors. The simplest cases of the function q being a sum, product, and quotient, are shown in Appendix A.

A more general error-propagation equation that is needed when some of the input variables are correlated:

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x_1}\right)^2 \sigma_{x1}^2 + \left(\frac{\partial q}{\partial x_2}\right)^2 \sigma_{x2}^2 + \cdots + 2\rho_{x1x2}\frac{\partial q}{\partial x_1}\frac{\partial q}{\partial x_2}\sigma_{x1x2} + 2\rho_{x1x3}\frac{\partial q}{\partial x_1}\frac{\partial q}{\partial x_3}\sigma_{x1x3} + 2\rho_{x2x3}\frac{\partial q}{\partial x_2}\frac{\partial q}{\partial x_3} + \cdots = \sum_{i=1}^{n}\left(\frac{\partial q}{\partial x_i}\sigma_{xi}\right)^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\rho_{xixj}\frac{\partial q}{\partial x_i}\frac{\partial q}{\partial x_j}\sigma_{xixj}. \quad (2)$$

In this equation, there are additional terms that correspond to each of the pair-wise combinations of the correlated variables. The parameter ρ represents the correlation coefficient between the variables.The propagation-of-error equations have a direct geometrical analogy with the addition of vectors. This analogy provides a helpful physical picture of the error-propagation equations, and is described in Appendix B. Eqs. 1 and 2 are termed the ''error equations''; a more detailed derivation of these equations can be found in Clifford (1973) and Taylor (1982).

The first step in the error-analysis process is therefore to express the step-by-step calculation of the desired output variable as an equation of the input variables. That is, to obtain the function, $q = f(x_1,x_2, \ldots x_n)$. The expression of the ecological calculation as a single mathematical function is a necessary and sometimes lengthy part of the error-analysis method. Arithmetic operations, evaluation of functions, interpolation, and statistical methods must all be written as equations and then combined. It is not always possible to obtain the function $f(x_1,x_2, \ldots x_n)$, in which case the error analysis can be applied separately to different parts of the calculation, or other methods may need to be considered. The analysis on the present data set as well as cases described in the literature (Oostrom et al. 1995, Beaudet et al. 2000) provide some examples of this procedure.

Once the function $q = f(x_1,x_2, \ldots x_n)$ has been written, the partial derivatives of the output variable with respect to each of the input variables can be determined through differentiation. The equations for the partial derivatives are then substituted into Eq. 1 or Eq. 2 to compute the error in the output variable. Error equations can also be written (and the error determined) for any of the intermediate variables that correspond to the individual steps of the calculation as well, as will be demonstrated in the ice-storm data set. There are some limitations and caveats with the Gaussian error propagation method and its application, and these are discussed further in the *Discussion* section, below

Application of GEP to the Ice-Storm Data

Description of the ice-storm data set

The data set used to demonstrate the error-analysis techniques was used to estimate the total downed woody biomass produced by a severe ice-storm in 1998 (Hooper 1999, Hooper et al. 2001). The study site was the Mont-St.-Hilaire mountain in the Gault Nature Reserve near Montreal, Quebec, Canada (Maycock 1961), and consists of approximately 10-km^2 of mainly old-growth deciduous forest dominated by *Acer saccharum*, *Fagus grandifolia*, and *Quercus rubra*. The biomass data consist of the diameter measurements (d_i) for 17 159 branches of the 10 most abundant species, measured from 117 circular plots 6 m in radius and spread over seven line transects (Fig. 1 shows the spatial layout of the data set). Pairs of linear-regression parameters (the slope A, and intercept B) describe a strong linear relationship between the log-transformed branch diameter (measured at the base) and the log-transformed biomass, from sample measurements taken over the entire mountain, for each species (see Hooper 1999). These regression parameters are listed in Table 1.

Fig. 2 illustrates the step-by-step nature of the calculation procedure for the total branch biomass. There are three calculation steps: (a) application of the back-transformed linear-regression relation to calculate branch biomass from branch diameters, (b) the summing of branch biomasses to obtain the total branch biomass by species, and (c) the sum over branch biomasses from all species to obtain the total branch biomass for the mountain. There are also four variable sets: (I) the branch diameter measurements (d_i) and regression parameters (A, B) for each species, (II) the branch biomasses calculated from the regression relations (BM_i), (III) the total branch biomasses for each species (BM_S), and (IV) the total branch biomass for the entire mountain (BM_T). It should be noted that the calculation steps transform measurements taken at a very small spatial scale (that of the individual branch) to a value that pertains to a very large spatial scale (the biomass estimate for the entire mountain). Calculations of this nature are common in ecological science, and therefore this data and error analysis can be viewed as being representative of and relevant to a general class of step-by-step or scaling-up types of ecological calculations. In this particular case, allometry, integration or summation, and a sampling scheme have been used to estimate total branch biomass from a set of diameter measurements.

Previous studies of allometry have typically used similar log-transformed branch data and found strong linear correlation over a wide range of branch sizes (Whittaker and Woodwell 1968; M. Hooper, *personal communication*, K. Arii, *personal communication*, and M. Lechowicz, *personal comunication*). Eq. 3a shows the allometric equation that relates branch biomass to branch diameter (this equation becomes a linear re-

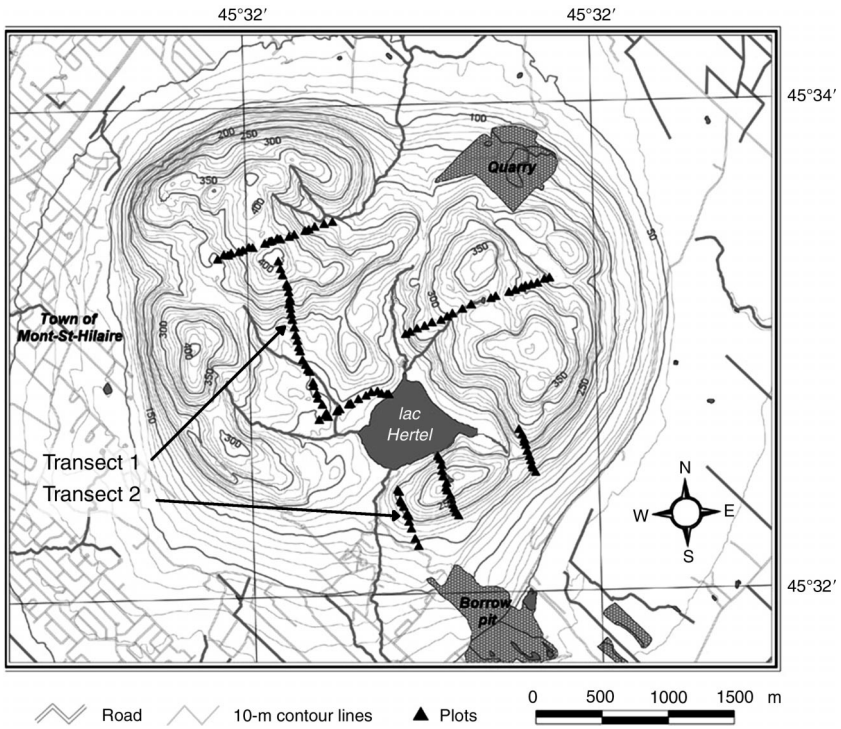

FIG. 1. Spatial layout of the data set on Mont St. Hilaire (near Montreal, Quebec, Canada). The 117 plots are each represented by a solid dark triangle and are arranged into seven linear transects that cover the entire mountain.

gression when put into logarithmic form). $BM_{S,i}$ denotes the biomass for branch i of species S; CF is a constant, species-specific correction factor; A_s and B_s are the species-specific regression coefficients; $d_{s,i}$ is the diameter of a branch at the base. The downed-branch biomasses for a given species were calculated from the measured branch diameters (Eq. 3a). The sum yields the total branch biomass for that species, BM_S, as shown in Eq. 3b. The total downed-branch biomass, BM_T, is equal to the sum of the individual species biomasses, which is shown in Eq. 3c. Eqs. 3a–c represent calculation steps (a), (b), and (c), respectively, in Fig. 2:

$$BM_{S,i} = CF_S(e^{-B_S} d_{S,i}^{A_S}) \quad (3a)$$

$$BM_S = \sum_i BM_{S,i} = CF_S e^{-B_S} \sum_i d_{S,i}^{A_S} \quad (3b)$$

$$BM_T = \sum_S BM_S = \sum_S \left(CF_S e^{-B_S} \sum_i d_{S,i}^{A_S} \right). \quad (3c)$$

Eq. 3c expresses analytically the desired output variable (total branch biomass) in terms of the measured input variables (the diameters and regression parameters). It represents the result of step-by-step calculations that scale the branch diameter measurements to a biomass estimate for the entire mountain. The error and its structure for the total branch biomass can be

TABLE 1. Regression data for the 10 tree species measured in the study, including regression sample sizes (N), R^2, and regression parameter values with errors and correlation coefficients.

Species	N	R^2	Slope, A	Error (%) ($2\sigma_A/A$) × 100%	Intercept, B	Error (%) ($2\sigma_B/B$) × 100%	Correlation coefficient, ρ_{AB}
Acer pensylvanicum	67	0.945	2.660	5.981	−3.016	3.774	−0.8708
A. saccharum	39	0.979	2.771	4.865	−2.942	6.841	−0.9407
Betula alleghaniensis	59	0.972	2.770	7.728	−2.998	8.774	−0.9451
B. papyrifera	21	0.986	2.520	3.115	−2.575	3.142	−0.8124
Fagus grandifolia	66	0.983	2.657	4.420	−2.744	5.051	−0.9078
Fraxinus americana	58	0.972	2.607	3.320	−2.644	4.652	−0.9239
Ostrya virginiana	39	0.987	2.709	3.763	−2.799	5.069	−0.9282
Populus grandidentata	22	0.968	2.921	8.161	−3.325	8.674	−0.9505
Quercus rubra	63	0.973	2.795	4.242	−3.023	5.703	−0.9604
Tilia americana	65	0.937	2.928	6.520	−3.768	5.793	−0.9346

Note: The N, R^2, A, and B data are from Hooper (1999).

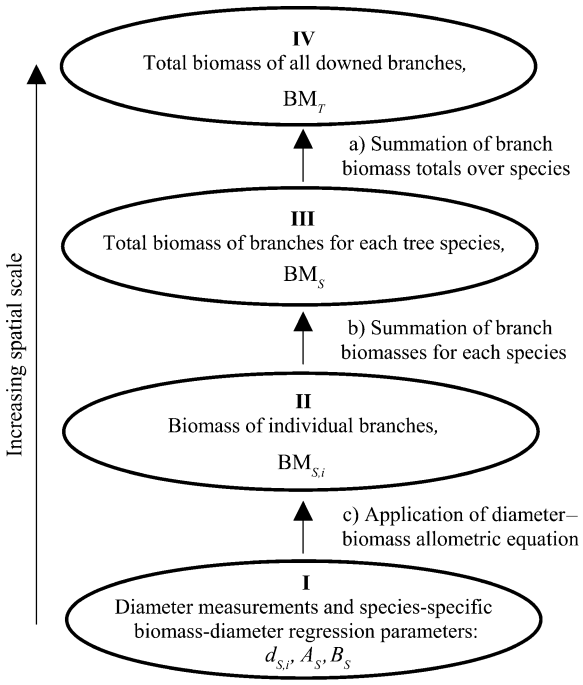

FIG. 2. Illustration of the step-by-step nature of the total downed branch biomass calculation. The four variable sets (I, II, III, IV) are listed in the ovals; $d_{S,i}$, A_S, and B_S represent the species-specific (species = S) branch diameters and regression parameters; $BM_{S,i}$, branch biomasses by species; BM_S, total estimated biomasses for each species; and BM_T total estimated branch biomass for all species. The calculation steps (a, b, c) used to compute variable sets at higher scales from those at adjacent lower scales are indicated between the corresponding variable sets. Errors in measurement and from regression analysis occur in the input variables (level I) and are propagated upward to the scale of the entire mountain (level IV).

determined by applying the error-propagation equations (Eqs. 1 and 2) to Eq. 3c. However, as already mentioned as well, error analysis can also be applied to the intermediate variables described by Eq. 3a and 3b. These equations together represent the model used to measure and calculate downed-branch biomass that was employed by Hooper et al. (2001) and Hooper (1999).

Calculation of the error magnitude in the total and species-specific biomasses

The application of Eqs. 1 and 2 to determine the error magnitude will now be demonstrated. The species-specific biomasses will be considered first. Eq. 3b represents an analytical expression for the species-specific biomass, BM_S, as a function of input variables. These input variables are the regression parameters A and B, and all of the diameter measurements that were made specific to the species S. The regression parameters each have a variance (from which the percentage errors are determined) and are also correlated. The variances and correlation coefficient (σ_A, σ_B, and ρ_{AB}) can be obtained from the covariance matrix that is produced when the regression parameters are calculated (e.g., from procedure PROC REG in SAS version 8 [SAS Institute 1999]); the values are also listed in Table 1. These variances reflect (and are produced by) the scatter or residual error (ε) between the data and linear regression model ($BM_{S,i} = [B \pm \sigma_B] + [A \pm \sigma_A]d_{S,i} \sim B + Ad_{S,i} + \varepsilon_i$) (Zar 1996: Ch. 16). The variance in the diameter measurements, σ_{di}, was estimated at 5% of the value of the diameter (cf. Appendix E).

The partial derivatives are determined from a simple differentiation of the biomass function described by Eq. 3b (the results are listed in Appendix C). Finally, the error in the species-specific biomass is calculated. Eq. 1, rewritten in terms of the species-specific biomass, takes on the form shown below:

$$\sigma^2_{BM_S} = \left(\frac{\partial BM_S}{\partial A}\right)^2 \sigma^2_A + \left(\frac{\partial BM_S}{\partial B}\right)^2 \sigma^2_B + \sum_i \left(\frac{\partial BM_S}{\partial d_i}\right)^2 \sigma^2_{di} + 2\rho_{AB}\sigma_A\sigma_B\left(\frac{\partial BM_S}{\partial A}\right)\left(\frac{\partial BM_S}{\partial B}\right). \quad (4)$$

Each diameter measurement represents an independent variable and so each one contributes to the sum of error terms. Substitution of the values of the input variables and their errors and correlation coefficients into Eq. 4 yields a value for the propagated error in the species-specific biomass. The results are tabulated for each species in Table 2.

The error in the total biomass is calculated in a similar way by applying the error-propagation equation to Eq. 3c. The total biomass, BM_T now becomes the output variable, and the input variables are the 10 species-specific biomasses. As the BM_S are completely independent of each other, they are each treated as independent input variables. Their variances have been determined from the preceding analysis. As the input variables are uncorrelated, Eq. 1 will be used. Differentiation of Eq. 3c shows that the partial derivatives are all equal to 1 (see Appendix C). Therefore, Eq. 1 rewritten in terms of the total biomass takes on the simple form of a sum of squares, as shown below (for species-code subscripts, see Table 3):

$$\sigma^2_{BM_T} = \sum_S \sigma^2_S = \sigma^2_{AP} + \sigma^2_{AS} + \sigma^2_{BA} + \cdots + \sigma^2_{TA}. \quad (5)$$

In this way the error magnitude in the total biomass estimate is determined. The result is shown at the bottom of the last column in Table 2.

The application of the error-propagation equations (Eqs. 1 and 2) to calculate the error magnitude has been demonstrated. The calculation has been performed in two steps in order to also determine the error magnitude in the intermediate variables (the species-specific biomasses). Note that the intermediate variables and their errors form the input variables for the next calculation stage.

TABLE 2. Error calculation results for the total and species-specific biomasses.

Species	No. branches†	Biomass (kg)†	Error (kg), $2\sigma_{BM_S}$	Error (%), $2\sigma_{BM_S}/BM_S \times 100\%$
Acer pensylvanicum	608	414.5	69.4	16.74
A. saccharum	8557	10 954.9	1028.5	9.39
Betula alleghaniensis	97	161.7	16.6	10.29
B. papyrifera	175	79.9	7.1	8.89
Fagus grandifolia	3362	6064.2	529.0	8.72
Fraxinus americana	1045	781.3	54.8	7.01
Ostrya virginiana	148	269.0	35.5	13.21
Populus grandidentata	32	65.4	11.2	17.21
Quercus rubra	3010	7424.1	1180.4	15.90
Tilia americana	125	128.5	25.6	19.92
Total	17 159	26 343.5	1655.7	6.28

† Data are from Hooper (1999).

Extension of the error-analysis method and definition of the ''error structure''

The calculated error can be partitioned or grouped according to its different sources. The grouping can be done by variable, by object, or spatially. The process is somewhat analogous to traditional ANOVA-type analyses, where measurements are also grouped in crossed or nested structures and their variances determined. Different ways of partitioning the error yield insights into its causes and nature, and the resulting equations can be used to find the precise conditions that minimize the error, either globally or from a single contributing source. Optimization or cost–benefit analyses can be done to achieve minimum error at minimum cost. It is the analytical nature of this particular error-analysis method (i.e., using the error-propagation equations) that permits these kinds of quantitative analyses of the error and its structure. These analyses refine our understanding of the nature of the error and allow for the development of specific experimental guidelines on its management and minimization.

The contribution of a source of error is measured in two ways. The calculation of the relative magnitudes of the various sources is termed the ''error budget.'' The error budget thus identifies the principle contributors to the total error. The sensitivity index, on the other hand, measures the effect of a change in the output error caused by a change in the input error and indicates the efficiency with which the output error can be reduced. Both the error budget and sensitivity indices can be calculated from the error equation and they collectively indicate which error sources can be most effectively reduced in order to reduce the magnitude of the output error. The set of error budgets and sensitivities for the various partitioning configurations of the output error is denoted the ''error structure'' of the experiment or calculation.

Calculation of the error budget

Eq. 6 is used to calculate the error budget for uncorrelated variables. It is derived by normalizing Eq. 1 by the square of the total error (i.e., the left-hand-side of the equation):

TABLE 3. Error budgets for the total biomass (by species and by variable) and the species-specific biomasses (by variable), for two cases of diameter error.

Species	Species code	Instrumental error only (%)†		5% diameter error (%)‡		Species error budget (%)
		Diameter	Regression	Diameter	Regression	
Acer pensylvanicum	AP	0.39	99.6	6.5	93.5	0.17
A. saccharum	AS	0.08	99.9	2.7	97.3	35.71
Betula alleghaniensis	BA	3.45	96.5	56.6	43.4	0.02
B. papyrifera	BP	6.50	93.5	25.3	74.7	0.00
Fagus grandifolia	FG	0.18	99.8	8.8	91.2	10.07
Fraxinus americana	FA	1.37	98.6	17.9	82.1	0.12
Ostrya virginiana	OV	2.48	97.5	66.4	33.6	0.12
Populus grandidentata	PG	3.45	96.6	30.0	70.0	0.01
Quercus rubra	QR	0.08	99.9	14.9	85.1	53.76
Tilia americana	TA	1.28	98.7	34.2	65.8	0.03
Total		0.10	99.9	10.0	90.0	

† Instrument-measurement error in the experiment was a maximum of 0.5%; see Appendix E for details.

‡ Diameter error refers to the natural variability of branches (e.g., non-circularity) and amounts to ~5% proportional error; see Appendix E for details.

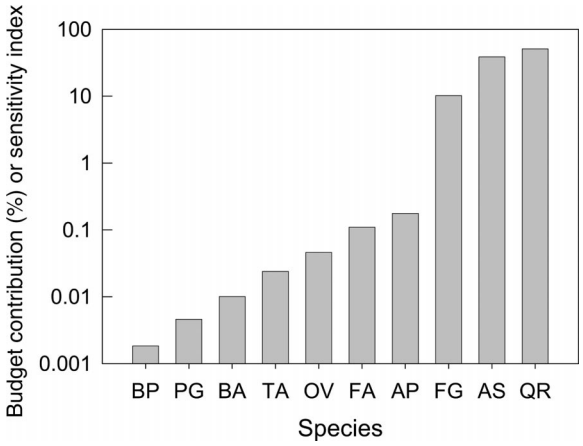

FIG. 3. Error budget of total biomass error in terms of species biomass errors, or equivalently, the sensitivity indices of the total biomass error to the biomass of each species. For species codes, see Table 3. Note the *y*-axis log scale.

$$1 = \left(\frac{\partial q}{\partial x}\right)^2 \frac{\sigma_{x1}^2}{\sigma_q^2} + \left(\frac{\partial q}{\partial x_2}\right)^2 \frac{\sigma_{x2}^2}{\sigma_q^2} + \cdots + \left(\frac{\partial q}{\partial x_n}\right)^2 \frac{\sigma_{xn}^2}{\sigma_q^2}. \quad (6)$$

Each term in Eq. 6 represents the percentage contribution of an input variable. The sum of all contributions equals to 1 (100%). The error budget also has a geometrical interpretation (described in Appendix B). Fig. 3 shows the percentage contributions of all 10 species to the total error as determined by the error budget calculation; the corresponding numerical values are listed in Table 3: Species error-budget column. The actual error-budget equations for the total biomass are shown in Appendix D. Each error-budget equation was obtained by normalizing Eq. 5 to the square of the total biomass error. Term 1 represents the percentage contribution of species 1 to the total error, term 2 that of species 2, and so on.

It is evident that the dominant sources of error in the total biomass, when grouped by species, occurs with AS, FG, and QR. These three species represent 35.7%, 10.1%, and 53.8% of the total error, respectively, for a combined contribution of 99.5%. Note that the percentage errors in the biomasses of other species may be larger (as shown in Table 2: Error (%) column). However the significantly larger biomass values of FG, AS, and QR result in larger absolute values for their uncertainties, and hence larger contributions to the total error. This behavior is expected since the total biomass is a linear sum of the species-specific biomasses. Eq. D.1 (in Appendix D) and Fig. 3 collectively represent a ''partitioning'' of the error ''by species.''

The error budget may be determined for the species-specific biomasses by applying the same normalization procedure to Eq. 3b. The resulting error budget represents a partitioning of the species biomass error ''by variable,'' and the results are shown in Fig. 4 for each species. The actual species-specific biomass error-budget equation is depicted in Eq. D.2 (in Appendix D), which shows the percentage contributions of the regression variables and the diameter measurements to the species-specific biomass error. The terms have been grouped into a regression group and a diameter group as each represents a different measurement type.

For most species, the major contribution to the error stems from the regression parameters. Examination of the terms in Eq. 6 and Eq. D.2 for the error in species-specific biomass indicates that the error terms that stem from the regression parameters vary quadratically with the number of branches (N^2), whereas the term that represents the diameter error increases only linearly (N). Thus for most species the error contribution of the regression parameter dominates, but for those species for which relatively few branches were measured, the contribution of diameter error becomes significant.

Therefore the error partitioning by species and by variable suggest that reduction of error is best accomplished by improving the regression relation in the primary contributors to the total error, AS, FG, and QR.

Calculation of sensitivity indices

The sensitivity index (e.g., Liu et al. 1995, Mayaka et al. 2004) represents the percentage rate of change in an output variable ($\Delta q/q$) produced per unit percentage change in its input variable ($\Delta x/x$), where the output variable $q(x)$ is a function of the input variable x. Eq. 7 shows the percentage change in the output variable as a function of the input variable; the sensitivity index is represented by the term in the square brackets and defined as the multiplier that scales the change in input variable to the change in the output variable:

$$\frac{\Delta q}{q} = \left(\frac{\partial q}{\partial x} \times \frac{x}{q}\right)\frac{\Delta x}{x}. \quad (7)$$

The same analysis can then be applied to the error

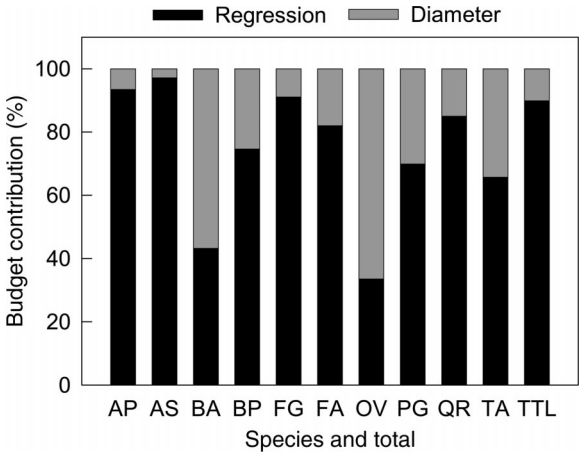

FIG. 4. The error budgets partitioned by variable (diameter and regression) for the 10 species-specific biomasses and the total (TTL) biomass. Note that for the total biomass and the three most abundant species (AS, FG, and QR) the error budget is dominated by the regression error.

equation in order to determine the sensitivity of the output error as a function of input errors. In this case σ_q becomes the output variable and σ_x becomes the input variable in Eq. 7. Eq. 8a describes the percentage variation produced in the output error by a variation in the input error of a given input variable; when applied to the total biomass as a function of species-specific biomasses, we find that the sensitivity index for the propagated error is in fact the same as its error-budget contributions, which is shown in Eq. 8b:

$$\frac{\Delta\sigma_q}{\sigma_q} = \left(\frac{\partial q}{\partial x}\right)^2 \left(\frac{\sigma_x}{\sigma_q}\right)^2 \left(\frac{\Delta\sigma_x}{\sigma_x}\right) \tag{8a}$$

$$\text{sensitivity index for total biomass} = \left(\frac{\sigma_S}{\sigma_{\text{BM}_\text{T}}}\right)^2. \tag{8b}$$

In practice, therefore, the process of error reduction is greatly simplified for the present analysis of total biomass error: reduction of the largest contributors to the error (i.e., for the species AS, FG, and QR) also result in the greatest reduction of the total error. In addition, a single calculation will provide both the relative magnitudes of the contributions to error (the error budget) as well as the sensitivities of the total error to its sources (the sensitivity indices).

Each percentage reduction in σ_{AS}, σ_{FG}, and σ_{QR} will result in a reduction in the total error, $\sigma_{\text{BM}_\text{T}}$, by 0.36%, 0.1%, and 0.54% respectively. As with the budget contributions, the sensitivities of these three species is far greater than the others (see Table 3: Species error-budget column). The strategy for the reduction of error up to this stage of the error analysis, as indicated by the by-variable partitioning, would be to reduce the error in the biomasses of QR, AS, and FG by reducing the error in the regression parameters for these species. In practice, however, as the correlation between data and regression relations are already quite strong (cf. Table 1), it is unclear if and how significant reductions in regression-parameter error can be achieved. Other methods of partitioning the error structure (as described below) should be investigated and will represent alternative and more effective possibilities.

Partitioning of the total biomass by variable and redundancy in the diameter error

The ''by-variable'' partitioning will be pursued further and it will be shown that error analysis not only identifies the major sources of error that should be reduced, but also reveals those that are redundant. Both of these aspects are important for experimental design: the former indicates areas where the design should be improved, while the latter indicates areas where the design is adequate or overly adequate, and where costs can possibly be reduced.

It has been shown that the process of error partitioning consists basically of grouping terms in the error-budget equation. This process is illustrated more clearly with the following analysis of the diameter measurement error contribution. In the previous budget analysis, the error in total biomass was partitioned into the contributions from the biomass errors of each species, and the species-specific errors were then each further partitioned into the error contributions of each variable. However, as all of the error equations are sums of terms, it is also possible to partition the total biomass error by variable. This is done by substituting Eq. 4 into each term in Eq. 5 and then grouping together the regression terms and the diameter terms over all species:

$$\begin{aligned}\sigma^2_{\text{BM}_\text{T}} &= \sum_{\text{species}} \left(\sum_{\text{variables}} \Delta^2_{\text{var,sp}} \right) \\ &= \sum_{\text{species}} \Delta^2_{\text{regression}} + \sum_{\text{species}} \Delta^2_{\text{diameter}}\end{aligned} \tag{9}$$

where the symbol ''Δ'' in Eq. 9 represents the error terms (e.g., partial derivative $\times$ standard error). The result is described schematically by Eq. 9 and amounts to a rearrangement and then regrouping of terms. Normalization of this equation to the total biomass error will then yield the error budget for the total biomass error partitioned by variable (the normalized equation is detailed in Appendix D). The results are shown in the bottom row of Table 3: diameter-error columns; they show that the diameter measurements collectively contribute ~10.0% of the error, while the regression parameters collectively contribute ~90.0%.

The diameter error consists of two components (cf. Appendix E for a detailed discussion). The main component is a ~5% proportional error that represents natural variability. The second component is a fixed and much smaller error that stems from the measurement precision in the instruments. The effect of this measurement precision can be isolated. The budget can be recalculated using Eq. 9 and Eq. 4, with only the measurement-precision error applied to the diameter measurements. The result is shown in Table 3: instrument-error columns. It can be seen that when only the measurement-precision error in the diameter is considered, that the error contribution of the diameter measurements to the total biomass error is negligible (~0.1%) as well as to most of the species-specific biomass errors (maximum ~6.5%, with most <1%).

In fact a decrease in the instrumental precision by a factor of 10 does not result in an appreciable impact on the total biomass error. When the budget calculation above was repeated, increasing the value of the diameter error (σ_{di}) from $\pm$0.1 cm to $\pm$1.0 cm, the error in total biomass only increased from 6.28% to 6.58%. Therefore it is concluded that implementing a less precise diameter-measurement method would save significant time and work in the field. For example the vernier and tape measurements could be made more quickly, or a simpler tool such as a centimeter ruler or measuring fork could be used. The insensitivity of the biomass error to even large values of diameter-measurement

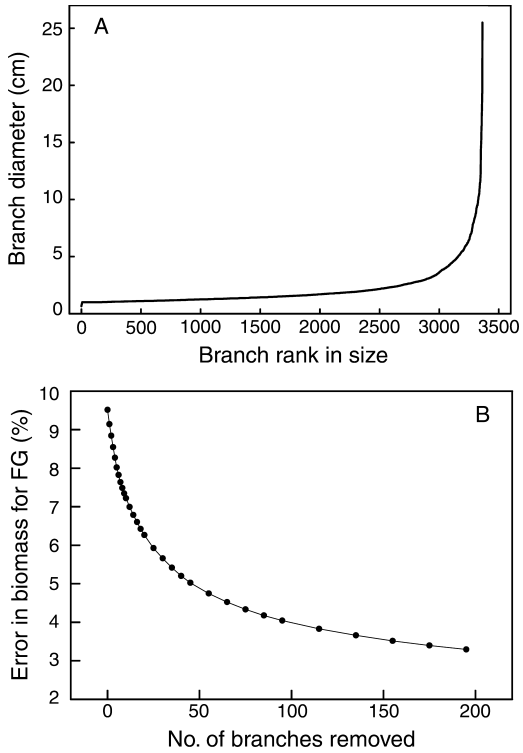

FIG. 5. Results of the branch-removal study, an example of error partitioning by object. (A) Branch rank in size for all of the branches of *Fagus grandifolia* (FG). Note the large number of relatively small branches and the small number of relatively large branches. (B) The effect on the species-specific biomass error ($2\sigma_{FG}/BM_{FG} \times 100\%$) of successive removal of the largest branches.

precision is an unexpected result of the error analysis that has significant implications for efficient sampling methodology.

Partitioning of the error ''by object''

A partitioning of the error structure by ecological object is next demonstrated. In this case, the objects are the branches. It is also demonstrated how the analytical nature of the error equation can be used to evaluate different measurement methods and to perform cost–benefit analyses with regards to minimizing error and optimizing experimental design.

It is hypothesized that the largest branches contribute disproportionately to the biomass error; since the biomass scales as a power law with the diameter (see Eqs. 3a–c), the error in the regression parameter (A) is expected to be greatly magnified for the largest diameter branches. As shown in Fig. 5a, which depicts the ranked branch distribution for a representative species, FG, there are relatively few extremely large branches whose biomasses are significantly greater (by orders of magnitude) than that of the median or average branch biomass. This suggests that these branches could be readily removed and that this removal would be an efficient means to reduce the error magnitude. Thus a method is needed to partition the error by branch in order to evaluate the contributions of the largest branches and to investigate if the results can be used to effectively reduce the total error.

Partitioning of the error by branch is done by arranging terms in the species-specific biomass equation (Eq. 3b) rather than the error equation that was used in previous analyses. (Only the three species AS, FG, and QR will be analyzed here for simplicity, and also because these were identified by the preceding error-budget calculation as being the primary sources for error in total biomass). The largest branches in the sum over branch biomasses in Eq. 3b are removed successively. Eq. 3b rewritten to show the n largest branches as distinct terms is represented in Eq. 10:

$$\begin{aligned} BM &= \sum_{i}^{N} BM_i \\ &= CFe^{-B} \sum_{i}^{n} d_i^A + BM_{N-n} + BM_{N-n+1} + \cdots \\ &\quad + BM_{N-1} + BM_N. \end{aligned} \tag{10}$$

The error is then recalculated using Eq. 4 after each removal. Note that the largest branches are retained in the total biomass value, even when they are removed from the error analysis. This mathematical method corresponds to an experimental method in which the largest branches are weighed directly using a scale. The measurement error in a direct weighing (including mass lost due to possible cutting into segments) is estimated to be $<1\%$ (i.e., $<\sim 50$ g for a 100-kg branch), which is much smaller than the error in individual branch biomasses produced by the regression method. Therefore the largest branches are removed from the error calculation but they are included in the total biomass estimate. The results are shown in Fig. 5b: as hypothesized, the species-specific biomass error drops very rapidly as the largest branches are removed.

Fig. 5b represents a curve of diminishing return that shows a trade-off between the additional labor of manually measuring branch biomasses, and a continual reduction in error. For example, the manual measurement of biomass for 1% of branches results in 56%, 63%, and 89% reductions in species-specific biomass error for AS, FG, and QR, respectively. These percentages would entail the removal of approximately 86, 34, and 30 branches, respectively. Although it would take significant time to weigh this number of branches it would not represent a significantly greater time cost over the manual diameter measurements that were already done. One could also use Fig. 5b to compute the number of branches that need to be removed in order to obtain a reduction in the species-specific biomass error by 50%. Analysis of these curves determines that this objective is obtained through removal of 55, 15, and 2 branches for AS, FG, and QR respectively.

TABLE 4. Results of the branch-removal error analysis.

Treatment†	Total biomass, BM_T				
	Error ($2\sigma_{BM_T}$)	Error (%) $2\sigma_{BM_T}/BM_T \times 100\%$	Derivative (%err/branch)‡	Reduction in total error (%)	No. branches removed
None	1655.7	6.28		0	0
50% error reduction					
QR	1302.4	4.94	−2.770	−21.34	2
AS	1395.7	5.30	−0.025	−15.70	55
FG	1591.0	6.04	−0.096	−3.90	15
All	832.4	3.16		−49.72	72
1% branch removal					
QR	1168.2	4.43	−0.044	−29.44	30
AS	1374.1	5.22	−0.014	−17.00	85
FG	1581.0	6.00	−0.043	−4.51	34
All	519.7	1.97		−68.61	149

Note: For species codes, see Table 3.

† Branch removal is an efficient means to reduce error magnitude (see *Application of GEP to the Ice-Storm Data: Partitioning of the error ''by object''* for explanation).

‡ The derivatives (slopes) of the branch-removal curves (e.g., Fig. 5) signify the change (decrease) in the biomass error percentage with each branch removed.

The above results are summarized in Table 4. The first row shows the error and % error with no branch removal. The next four rows show the effect on the total error of a 50% reduction in the species-specific errors of AS, FG, and QR separately and then together. The bottom four rows of the table show the effect on the total error of the removal of 1% of branches in each species separately and then together. Columns 2 and 3 show the resultant error and % error in total biomass. Column 5 lists the percentage change (reduction), and Column 6 indicates the number of branches that need to be removed and weighed separately for each of the treatments.

The derivatives of the branch removal curves signify the change (decrease) in percent error in biomass with each branch removed. They indicate the exact benefit-per-cost ratios and are calculated at the 50% error points and the 1% branch points. The results are listed in column 4 of Table 4. For example, after the error has been reduced by 50% in QR, each additional branch removed at this point results in an additional 2.8% reduction in error. In another example, after 1% of the branches of AS have been removed, additional branch removal results in a 0.014% reduction in error, which may not be cost effective. At this point one may wish to consider other strategies (i.e., other partitioning methods or the improvement of error in other variables) to reduce error.

In conclusion, the partitioning-by-branch analysis reveals a highly efficient approach to reducing error that is likely superior to the previously identified strategy of reducing the error in the regression parameters. With a little additional effort (direct weighing vs. diameter measurement of the largest branches) the total error is reduced by more than 50%. In fact if direct weighing of larger branches were actually done during the collection of field data, these branches could also then be added to the regression database to further improve the precision of the regression parameter estimates as well.

Spatial partitioning of the error

Partitioning by location.—Spatial partitioning can be done by plot location. The total biomass can be written as the sum of branch biomass contributions from the different plots each of which has a unique spatial coordinate (Fig. 1). Therefore, the total biomass error can also be thought of as the sum of the error contributions stemming from different plots.

It is possible that certain spatial locations have more error than others. The plots differ in their branch size distributions, numbers of branches, variabilities in branch form, etc. These differences could in turn reflect differences in biotic and abiotic factors such as tree species composition, soil moisture and chemistry, elevation, etc. Thus variations in the biomass error structure from plot to plot could correlate with and reveal structure in the ecologically relevant characteristics of the sites. With regard to experimental design, different experimental techniques could then be developed to handle particular biotic or abiotic site conditions that have been identified to produce large amounts of error.

To perform the by-location analysis, Eq. 3c is applied to each plot to calculate its downed branch biomass. An error analysis is then performed on the total biomass from each plot. The analysis of the spatial distribution of errors for two transects (Transects 1 and 2 as indicated in Fig. 1) was done. Each represents a significant ecological gradient and they were selected to maximize the probability that any correlations between error structure and site conditions would be revealed. The results of the analysis from Transect 1 are shown in

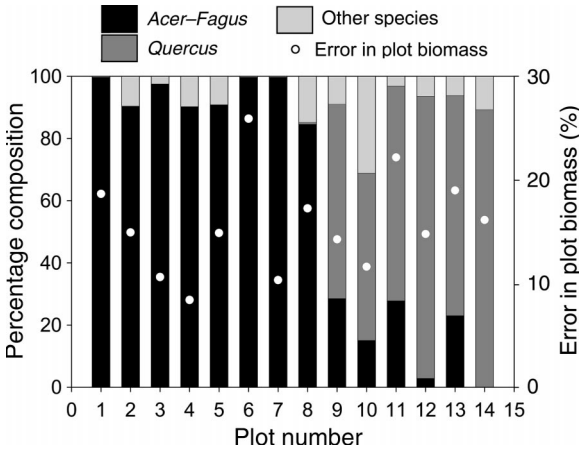

FIG. 6. The percentage error in species biomass ($[2\sigma_{BM}/BM] \times 100\%$), represented by the open dots and the *y*-axis on the right-hand side, for a sequence of plots that extend over an ecological gradient. Species composition is shown in the underlying shaded bars and the *y*-axis on the left-hand side.

Fig. 6. This transect consisted of a beech–maple forest that changed abruptly into an oak–maple forest, as the transect ascended and then descended over a ridge. Fig. 6 shows the species composition of the plots as well as the variation in biomass percentage error by plot.

No significant spatial trends in the error structure were detected; it was most likely masked by natural and stochastic variability in the data. Despite the lack of a conclusive pattern in the results, however, error partitioning by spatial location represents a valuable and interesting method both to analyze error and to identify or corroborate ecological structure. Ecological data are readily organized spatially and there are often spatial patterns or trends. Other data sets may have less statistical noise and a greater correlation between error and site condition. A positive correlation between error structure and site condition, and the identification of its mechanism, would be of great interest.

Partitioning by-scale.—There are four distinct levels of spatial scale that naturally occur in this experiment: the branch biomass is calculated at the scales of single branches, plots, transects, and over the entire mountain (i.e., all transects). The ''by-scale'' analysis is similar to the by-location analysis in that Eq. 3c is now applied to branch biomasses grouped by scale. An error analysis is then performed on the total biomass at each level or subset of branches (i.e., to single branches, to the biomass of branches from plots, to the branch biomass of transects, and finally to the entire data set.)

The variation of error magnitude with scale is shown in Fig. 7. In this paper, for simplicity only the results for *Acer saccharum* (AS) are shown. Fig. 7A shows that the percentage biomass error increases with decreasing scale. The decreasing levels of scale from mountain to individual branch have been assigned integer values from 1 to 4. When the biomass estimation technique is applied to a single branch, an error of ~28–33% results; at the level of the entire mountain, the technique yields an error of only ~9%. Thus the effectiveness of the experimental method is shown to depend on spatial scale. Fig. 7B shows the error-budget contribution of the diameter error plotted against scale. At the level of the whole mountain, the diameter error accounts for only ~3% of the total biomass error (for AS). This increases steadily with decreasing scale so that for individual branches, the diameter error contributes 60–95%. Therefore while the precision in diameter measurements can be reduced in the interest of speed and ease of measurement in the estimation of whole-mountain downed biomass, they in fact must be performed more carefully at smaller scales.

A description of the characteristics of each of the four levels of scale is provided in Table 5. It can be seen that the scale level roughly corresponds to several other variables–spatial surface area, number of branches(N), and branch biomass. Note also that there is often significant overlap between the characteristics of different levels of scale (for example in the number of branches and biomass between the plot scale level = 3) and transect (scale level = 2) scales. The concept of ''scale'' is often quantitatively ambiguous. For example, when referring to surface area, body mass, or population size it describes these quantities only to within an order of magnitude. In general, the level of scale in ecology is usually correlated with several ecological and geographical variables that vary in different ways as the scale varies. The effect of scale on a quantity or phenomenon of interest actually stems from the collective change and interaction of these correlated variables.

To illustrate the above concept and to clarify the effect of scale on the error structure, the branch number is chosen as a proxy variable for scale; branch number has a more precise ecological and mechanistic meaning. Fig. 7C and D shows the error structure charted as a function of branch number rather than level of scale. Fig. 7C shows a clearer decrease in error magnitude with scale when the proxy variable is used. In Fig. 7D the error-budget data now fall very tightly on a continuous curve. The plot and transect scales overlap in these graphs, indicating the ambiguity in the scale variable as mentioned. It is highly interesting to note a distinct threshold at approximately 100 branches, above which the diameter contribution to the total error is negligible and the error budget is then dominated by the regression parameters. This finding has significant implications for experimental design.

In the branch biomass estimation method, the error is determined by an interaction between two effects, each of which dominates at different scales. A higher number of branches leads to the relative dominance of error by the regression parameters whose contribution varies as N^2. As branch number is reduced, the relative sizes of the error terms shifts, and the biomass error

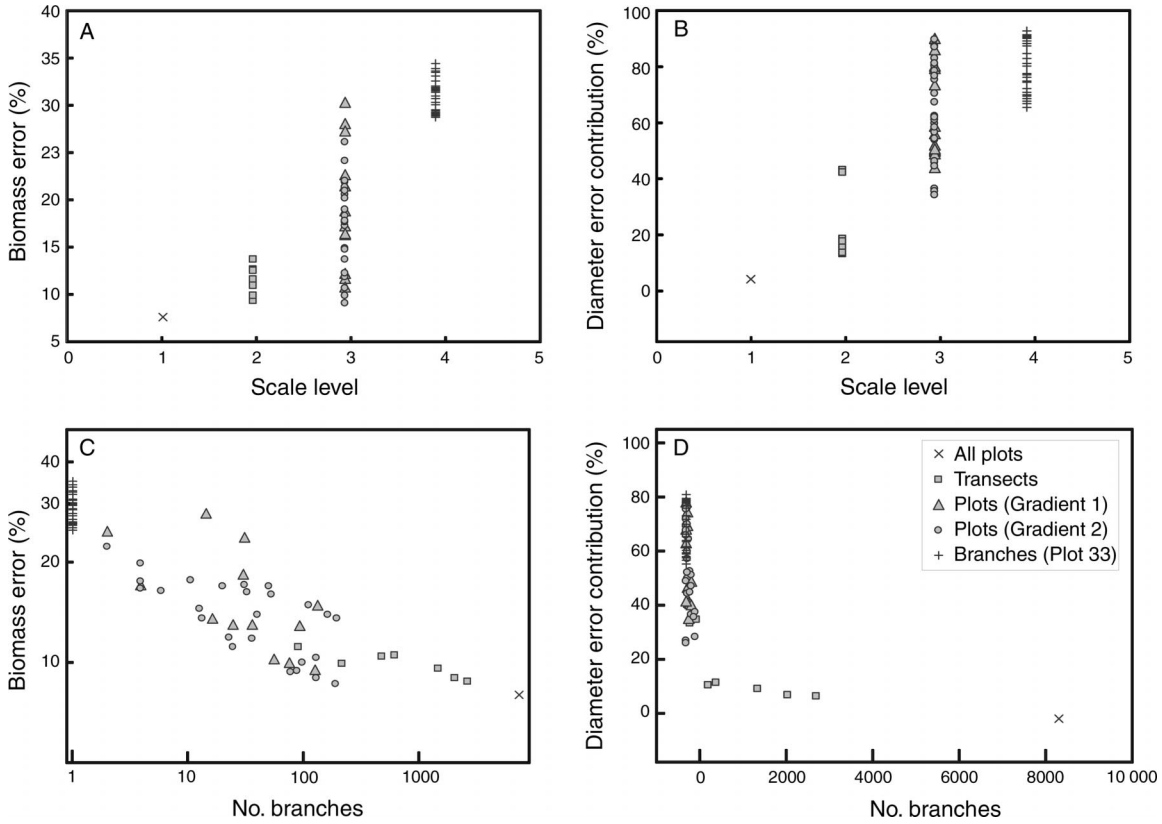

FIG. 7. Variation in error structure with scale: (A, B) the error magnitude ($2\sigma_{BM}/BM \times 100\%$) and the diameter contribution to the biomass-error budget plotted against scale level (C, D); the same quantities plotted against the scale proxy variable, the number of branches.

becomes dominated by the contribution of the diameter error, which varies as *N*. The interplay between the two contributions has been discussed already in the preceding analyses. The observed and abrupt threshold at 100 branches likely represents a mathematical effect that can be observed in a function of the form $f(x) = ax + bx^2$, where the quadratic term becomes abruptly larger than the linear term as x increases.

The scaling behavior of this downed-branch-biomass system may be applicable to other systems as well. Thresholds in scale that separate two quite-different regimes of behavior are not uncommon. It is quite possible that they are produced by the transition between two competing effects, as was the case in this error analysis. The identification of underlying variables that change with scale, and the choice of an appropriate proxy variable to clarify the mechanisms for variation with scale, as was done in this analysis, may be relevant to other scaling studies as well.

The by-scale partitioning of error for the case of *Acer saccharum* shows that the method used for branch biomass estimation becomes less appropriate below a scale of about 100 branches, due to an increase in the magnitude of error. Five of the species (BA, BP, OV, PG, TA) have branch sample sizes that are close to 100. It is also conceivable that plot-by-plot estimations of the downed biomass could be needed from the data set; many plots have 100 branches or less. The error analysis thus shows that the error in biomass for these cases would be much higher (and caused by the diameter

TABLE 5. The four levels of scale and their error structures.

Scale description	Scale level	No. plots	Area (m^2)	No. of branches	Biomass (kg)	Error, $(2\sigma_{BM}/BM) \times 100\%$	Contribution (%)	
							Diameter error	Regression
Mountain	1	117	13 232	8557	10 955	9.4	2.7	97.3
Transect	2	11–28	1244.1–3116.8	98–3005	254–2756	9.3–13.2	10.7–42.0	58.0–89.3
Plot	3	1	113.1	2–212	0.22–456.2	9.3–30.3	33.7–90.9	9.1–66.3
Branch	4	<1	<113.1	1	0.054–44.96	31.7–34.4	65.5–93.8	6.2–34.5

error), than the error estimated for the total biomass. Direct weighing of branches would be recommended as a superior method for these smaller species-specific and plot-specific data sets. Finally, the analysis shows that, in general, an experimental design should not only consider the error structure of the system, but also its possible variation with scale. A design or calculation method that is appropriate at one scale may be inappropriate or exhibit different behavior at another, in light of its error structure.

Overall Discussion

Conclusions regarding the error analysis of the downed-biomass calculation

Knowledge of the error magnitude gives statistical credibility to comparisons made between the downed biomass amounts produced by the 1998 ice storm with other forms of natural disturbance (Hooper 1999). It also gives greater statistical confidence in quantitative assessments of the impact of the ice storm on succession, environmental stress, ecology and dynamics, nutrient cycling, and other ecosystem-level processes (Hooper 1999). A key result of the present study is that the experimental and calculation methods used in the ice-storm-downed biomass study (of Hooper [1999]) have been verified to be acceptable; the error magnitude in the total biomass is an order of magnitude lower than the biomass value itself.

Several findings from this study could well impact the way future downed biomass assessments of this kind are done. It was found that branch diameter measurements can be made to an uncertainty of up to ± 1 cm without a significant effect on the total error. This result could greatly reduce the amount of time and work required in the experiment. In addition, the most effective method of reducing error was identified as the removal of the heaviest branches for the three species, AS, FG, and QR (species codes given in Table 3). A calculation was done that shows the incremental improvement in error with each branch removed, as a function of the total number of branches removed. This analysis allows the experimenter to judge (based on the cost and effort required) quantitatively how many branches should be removed before the incremental improvement in error becomes too small to be worthwhile. Finally, the branch biomass estimation method will yield significantly higher error (up to 33%) when the number of branches in a sample drops below 100. In these cases, (e.g., for smaller plots, or rarer species), the use of alternative methods, such as direct weighing, is recommended.

Scope and applicability of the method

Whether or not ''true'' or ''known'' models (the biomass calculation method is treated as such in the Gaussian error propagation [GEP] analysis) actually exist and can be identified is debatable (Chatfield 1995, Hilborn and Mangel 1997). Some researchers feel that the only realistic option is to compare the degree of likelihood between different models fitted to observed data. The existence of 'true' models is less of an issue in the physical sciences and engineering (from where the GEP method is taken) than in ecology. In the former disciplines, laboratory experiments can be prepared and scientific phenomena isolated to remove almost all external factors, which makes model identification less ambiguous. This is not usually possible in ecology research. However, even in ecological studies a calculation method or experimental design must in the end be specified in order to determine or measure desired ecological variables. This chosen method or design can be thought of as a specified, known model through which error in measured input variables propagates to produce a certain variability in the output variable(s). This error can be thought of as the fundamental lower limit of the error, which is produced by statistical dispersion (i.e., precision error). The calculation results can then be used for the purposes of identifying sources of error within the model structure, and consequently evaluating and improving the chosen calculation or experimental method (as opposed to assessing model correctness). In this sense, the GEP method and its intended application as described in this article have a comparatively more practical meaning and application than other forms of error analysis.

Although the GEP method has been clearly defined in terms of its context in statistical inference and contrasted with Monte Carlo methods, it is not incompatible with analyses of other kinds of uncertainty or with other methods. For example, Monte Carlo could be used in conjunction with GEP, or it could be used to verify the accuracy and applicability of GEP estimates of error. The use of basic GEP under the name ''delta method'' in Bayesian inference is an example where a different interpretation (i.e., that of a posterior probability) can be given to the results of an equivalent analysis (Sorensen and Gianola 2002). In general, GEP can likely be used for purposes beyond those described in this study; exploration of these possibilities represents future work.

The error propagation equations (Eqs. 1 and 2) themselves are derived from a first-order Taylor expansion of the Gaussian distribution. The Taylor expansion is described in most standard texts on calculus (e.g., Berkey 1994), and an excellent overview is given in Tinker and Lambourne (2000). Thus the error magnitudes involved are assumed to be an order-of-magnitude smaller than their associated variables (in order to satisfy a first-order approximation). This assumption is usually true, and is satisfied in the present study with the regression parameters and diameter measurements. In many real systems, the Taylor expansion has been found to be a surprisingly robust approximation even when applied with large deviations. Oostrom et al. (1995) compared the error (i.e., statistical dispersion)

predicted by GEP with that obtained from repeated measurements in a gamma-radiation soil-measurement system. An excellent numerical match was found between prediction and measurement. The accuracy of the Taylor expansion used in GEP can be readily verified. Taylor's theorem (Berkey 1994: 694) provides an estimate for the upper bound for the error in the Taylor expansion approximation. This upper bound or remainder is readily calculated (Appendix F). Alternatively, if repeated measurements of the desired output variable are available, their distribution can be checked for normality. A skewed distribution of values could indicate that the first-order approximation has broken down (i.e., the second-order term has become significant in magnitude) and that the GEP method should not be used.

It should also be noted that the expression of the desired output variable in terms of a single mathematical equation of the input variables (as shown by Eq. 3c in this study), is an essential requirement for the application of this method. This first and necessary step can be time-consuming, may require mathematical expertise, and may not always be possible. However it is suggested that with persistence and mathematical creativity, a surprisingly large number of calculation processes can be expressed as equations. Processes such as summation, integration, regression, interpolation, as well as other equations and many statistical processes all have an analytic form. At the very least, if some of the calculation steps cannot be expressed mathematically, then the error structure could be determined for selected parts of the system. The method could then be combined with Monte Carlo-type methods to estimate total error. The resultant aggregation of many intermediate calculation steps can sometimes result in very large error equations (for example see Beaudet et al. [2000]: Appendix 3). In these cases, the computation of error magnitudes and the error structure can also be time-consuming but in the end will reveal the effect of all the calculation steps and error interactions.

The analytical nature of the method is also one of its greatest strengths. Since the error is expressed as a mathematical equation, the precise nature of its variation with the input variables, system parameters, and input sources of error can be charted. Not only can the error magnitude be quantitatively minimized, but the optimal methods of doing so can be found, as was shown with the sample data set. It should be noted here that the partitioning methods described in this article (that involve the grouping of terms in the error equation and the biomass equation) are generally applicable and not dependent on the particular equations being analyzed. Thus this type of method gives much greater understanding of, as well as control over, the nature of the error and its structure. In contrast, Monte Carlo-type methods require exhaustive simulations of combinations of possibilities in order to identify even the most basic elements of the error behavior. Or they often represent only one-time estimations of error for a single set of parameters.

It should also be noted that the error analysis done for the downed-biomass data set was done in hindsight, that is, after the experiment had already been completed. In practice, in cases where error is a concern, experiments should be designed and carried through in parallel with error analysis. The situation would be similar to the in situ adjustments made by the experimenter to account for statistical issues in data or adverse field conditions. Estimates of the nature of the data and how it might be distributed, in size or space for example, could be made from previous studies or from existing knowledge, which would then permit a priori estimates of the error structure. It is recommended in general that the error structure should be an issue that a researcher always has in mind when designing and conducting an experiment. Thus large sources of error, the amplification of error through propagation, required levels of measurement precision, and the identification of optimal methods to reduce error can all be integrated into the experimental-design process. In the same way that collected data may be useless if it is unsuitable for the intended numerical analyses (cf. Legendre and Legendre 1998:4–5) it may be likewise unusable if its error magnitudes are unacceptably large.

The relevance of error analysis to the issue of scale in ecology

Step-by-step calculations that scale up small-scale observations or measurements to the level of entire ecosystems or landscapes are ubiquitous in ecological science. This is due to the important need for determining ecosystem-level quantities in concert with the practical difficulty in making direct, large-scale measurements. There has been an extensive and ongoing discussion in ecology on issues related to scale (e.g., Levin 1992, Peterson and Parker 1998). It has been recognized that the mechanisms that drive ecological phenomena often change with scale, producing corresponding changes in observed structure and patterns. However the actual calculation methods and techniques that can be used to estimate ecological variables across scales has received less attention. Scaling-up calculations are invariably multi-step in nature and involve the aggregation of data at smaller scales and equations that extrapolate these data to the desired larger scale. Some examples include the estimation of whole-lake mercury content (Håkanson 1999), whole-tree carbon gain (Beaudet et al. 2000), net primary productivity of ecosystems (Wiegert and Evans 1964, Linthurst and Reimold 1978), and forest economic yield (Greene et al. 2002). As already mentioned, these types of calculations are highly susceptible to unforeseen interactions in different error sources and the possible am-

plification of error in the output variables. Therefore even when a scaling process is known, its calculation may not be feasible due to error propagation.

The error-analysis method presented here is ideally suited to determining the error in multi-step, scaling-up ecological calculations. The quantitative knowledge of confidence intervals that results is also highly important to the statistical and ecological significance of the large-scale indices (such as ecosystem biomass and productivity), which are determined from those calculations. Moreover, if the error structure is also known, then the major sources for error in the system and optimal methods to reduce the error (if required), can be identified.

The present scaling analysis for the downed-biomass system has also illustrated several possibly general aspects of scaling calculations. The analysis has exhibited a fairly classic scaling behavior in that there are two distinct regimes of behavior at smaller and larger scales, separated by a very distinct threshold (of $\sim$100 branches). The causes for this behavior have been identified as a tension between two effects (that of regression parameter and diameter-error propagation) that vary at different rates with respect to the number of branches. The threshold is determined by the point at which a term proportional to N^2 becomes larger than one that is proportional to N. It has been shown how the identification of a proxy variable for scale can greatly clarify the effects and mechanisms for scaling behavior. Finally, the present analysis has shown that the error and the corresponding feasibility of an experimental design can change with scale.

Conclusions

This work extends the theory and applicability of the standard Gaussian error-propagation method. The concept of the error structure of an experiment or calculation has been described as the set of error budgets and sensitivity indices over different partitioning methods. The error structure offers a comprehensive framework in which the error distribution of the system can be understood, and ways to reduce error most effectively identified.

The method of determining the error structure has been explained and demonstrated on a sample data set, and many useful and interesting results have been found. It is hoped that this paper will serve as a teaching guide and reference to students and scientists who wish to perform similar error analyses on their own data. In particular, this method is highly applicable to ecological studies (i.e., calculations and experiments) that have multiple calculation steps or that involve scaling-up calculations used to estimate large-scale values or indices.

Acknowledgments

I thank Andrew Park, Frederic Guichard, Justine Karst, Benjamin Gilbert, Ken Arii, Jon Shik, Timo Kuulnvainen, and Martin Lechowicz for their encouragement and assistance in reviewing earlier drafts of the manuscript, and Michael Hooper for access to the ice-storm-downed biomass data. The thoughtful and detailed comments, and inquiries made by Mark L. Taper and an anonymous reviewer are also greatly appreciated.

Literature Cited

Beaudet, M., C. Messier, D. W. Hilbert, E. Lo, Z. M. Wang, and M. J. Lechowicz. 2000. Leaf- and plant-level carbon gain in yellow birch, sugar maple, and beech seedlings from contrasting forest light environments. Canadian Journal of Forest Research **30**:390–404.

Berkey, D. D. 1994. Applied calculus. Saunders College Publishing, Fort Worth, Texas, USA.

Brewer, C. A., and L. J. Gross. 2003. Training ecologists to think with uncertainty in mind. Ecology **84**:1412–1414.

Chatfield, C. 1995. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society Series A **158**:419–466.

Clifford, A. A. 1973. Multivariate error analysis. Applied Science Publishers, Essex, UK.

Deutschman, D. H., S. A. Levin, and S. W. Pacala. 1999. Error propagation in a forest succession model: the role of fine-scale heterogeneity in light. Ecology **80**:1927–1943.

Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications **6**:1036–1046.

Greene, D. F., D. D. Kneeshaw, C. Messier, V. Lieffers, D. Cormier, R. Doucet, C. Calogeropoulos, K. D. Coates, D. Cormier, V. Lieffers, and G. Grover. 2002. Modelling silvicultural alternatives for conifor regeneration in boreal mixedwood stands. Forestry Chronicle **78**:281–295.

Håkanson, L. 1999. Error propagation in step-by-step predictions: examples for environmental management using regression models for lake ecosystems. Environmental Modelling and Software **14**:49–58.

Hilborn, R., and M. Mangel. 1997. The ecological detective, confronting models with data. Princeton University Press, Princeton, New Jersey, USA.

Hooper, M. C. 1999. Impacts of a catastrophic ice storm on an old-growth, hardwood forest Thesis, Department of Biology, McGill University, Montreal, Quebec, Canada.

Hooper, M. C., K. Arii, and M. J. Lechowicz. 2001. Impact of a major ice storm on an old-growth hardwood forest. Canadian Journal of Botany **79**:70–75.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Elsevier, Amsterdam, The Netherlands.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology **73**:1943–1967.

Linthurst, R. A., and R. J. Reimold. 1978. An evaluation of methods for estimating the net aerial primary productivity of estuarine angiosperms. Journal of Applied Ecology **15**: 919–931.

Liu, J., J. B. Dunning, Jr., and H. R. Pulliam. 1995. Potential effects of a forest management plan on Bachman's sparrows (Aimophila aestivalis): linking a spatially explicit model with GIS. Conservation Biology **9**(1):62–75.

Magnussen, S., and D. Burgess. 1997. Stochastic resampling techniques for quantifying error propagations in forest field experiments. Canadian Journal of Forest Research **27**:630–637.

Mayaka, T. B., J. D. Stigter, I. M. A. Heitkönig, and H. H. T. Prins. 2004. A population dynamics model for the management of Buffon's kob (Kobus kob kob) in the Bénoué National Park Complex, Cameroon. Ecological Modelling **176**(1–2):135–153.

Maycock, P. 1961. Botanical studies on Mont St. Hilaire, Rouville County, Quebec. General description and a botanical survey. Canadian Journal of Botany **39**:1293–1325.

Oostrom, M., J. H. Dane, B. C. Missildine, and R. J. Lenhard. 1995. Error analysis of dual-energy gamma radiation measurements. Soil Science **160**(1):28–42.

Pacala, S., C. Canham, J. Saponara, J. Silander, R. Kobe, and E. Ribbens. 1996. Forest models defined by field measurements: estimation, error analysis, and dynamics. Ecological Monographs **66**:1–43.

Peterson, D. L., and V. T. Parker, editors. 1998. Ecological scale: theory and applications. Columbia University Press, New York, New York, USA.

SAS Institute. 1999. SAS for Windows, Version 8. SAS Institute, Cary, North Carolina USA.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry: the principles and practice of statistics in biological research. Second edition. W.H. Freeman and Company, San Francisco, California, USA.

Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian and MCMC methods in quantitative genetics. Springer-Verlag.

Taylor, J. R. 1982. An introduction to error analysis. University Science Books, Mill Valley, California, USA.

Tinker, M., and R. Lambourne. 2000. Further mathematics for the physical sciences. John Wiley and Sons, New York, New York, USA.

Wiegert, R. G., and F. C. Evans. 1964. Primary production and the disappearance of dead vegetation on an old field in southeastern Michigan. Ecology **45**:49–63.

Whittaker, R. H., and G. H. Woodwell. 1968. Dimension and production relations of trees and shrubs in the Brookhaven Forest, New York. Journal of Ecology **56**:1–25.

Zar, J. H. 1996. Biostatistical analysis. Third edition. Prentice Hall, Upper Saddle River, New Jersey, USA.