

RESEARCH

Open Access



The Chapman-Richards Distribution and its Relationship to the Generalized Beta

Jeffrey H. Gove^{1*}, Thomas B. Lynch² and Mark J. Ducey³

Abstract

Background: The Chapman-Richards distribution is developed as a special case of the equilibrium solution to the McKendrick-Von Foerster equation. The Chapman-Richards distribution incorporates the vital rate assumptions of the Chapman-Richards growth function, constant mortality and recruitment into the mathematical form of the distribution. Therefore, unlike ‘assumed’ distribution models, it is intrinsically linked with the underlying vital rates for the forest area under consideration.

Methods: It is shown that the Chapman-Richards distribution can be recast as a subset of the generalized beta distribution of the first kind, a rich family of assumed probability distribution models with known properties. These known properties for the generalized beta are then immediately available for the Chapman-Richards distribution, such as the form of the compatible basal area-size distribution. A simple two-stage procedure is proposed for the estimation of the model parameters and simulation experiments are conducted to validate the procedure for four different possible distribution shapes.

Results: The simulations explore the efficacy of the two-stage estimation procedure; these cover the estimation of the growth equation and mortality—recruitment derives from the equilibrium assumption. The parameter estimates are shown to depend on both the sample size and the amount of noise imparted to the synthetic measurements. The results vary somewhat by distribution shape, with the smaller, noisier samples providing less reliable estimates of the vital rates and final distribution forms.

Conclusions: The Chapman-Richards distribution in its original form, or recast as a generalized beta form, presents a potentially useful model integrating vital rates and stand diameters into a flexible family of resultant distributions shapes. The data requirements are modest, and parameter estimation is straightforward provided the minimal recommended sample sizes are obtained.

Keywords: Diameter distributions, Chapman-Richards growth, Generalized beta distribution of the first kind, Maximum likelihood, McKendrick-Von Foerster equation, Physiologically structured population model, Size-structured distributions

Background

Diameter distribution models have a long history of use in forestry. Perhaps the first, and arguably the best-known published model was from a study in silver fir (*Abies alba* Mill.) by de Liocourt (1898) and recently reviewed by Kerr (2014). de Liocourt (1898) developed a reverse J-shaped model to fit the distribution of diameters in a “normal selection forest.” As noted by Kerr (2014),

de Liocourt (1898) did not fit a negative exponential, or so-called q -distribution to these data, which is a common misunderstanding. Indeed, he did not fit a probability distribution at all. Rather, he used a mathematical method that was undoubtedly common at the time, and which is now referred to as a “difference table” approach (Gove 2015). The details are not important, but the fact that this appears to be one of the first known applications of a mathematical model being fitted to a diameter distribution in forestry is certainly of interest. A second interesting historical application of a general mathematical method useful for the fitting of diameter distributions

*Correspondence: jgove@fs.fed.us

¹ USDA Forest Service, Northern Research Station, 271 Mast Road, Durham 03824, NH, USA

Full list of author information is available at the end of the article

was the use of Fourier Series methods, which appeared in the German literature in the 1920's (p. 14, (Meyer 1930)), and was introduced to the more general American audience by Anderson (1937). In addition to Fourier series, Meyer (1930) discusses several graphical methods for fitting "frequency curves" dating back to the 1880's; he also mentions other mathematical methods which include the normal distribution. In his own application, Meyer (1930) used the Charlier system ([now known as the Gram-Charlier Series, e.g., p. 222], (Stuart and Ord 1987)) for modeling diameters and notes that Cajanus (1914) was evidently the first to apply this model to diameter distributions in Finland. Clearly there has been a long history of foresters who have recognized the need to characterize diameter distribution in some mathematical (including graphical) form for use in management.

In more recent decades, foresters have moved from the methods such as those mentioned above to the fitting of probability distributions, or probability density functions (PDFs), to the diameter data. Examples include the lognormal (Bliss and Reinker 1964), negative exponential (Meyer 1952; Leak 1964), Weibull (Bailey and Dell 1973), Johnson's S_B (Hafley and Schreuder 1977; Zhang et al. 2003; Rennolls and Wang 2005) and Burr distributions (Lindsay et al. 1996; Gove et al. 2008), to name but a few. Many of these distributions have several forms which have been explored in the forestry literature; also, they have been set in more complex applications such as finite mixture models in order to fit, e.g., multimodal distribution shapes (Liu et al. 2001; Zhang et al. 2001). These models are often called 'assumed' distribution models, because there is nothing in the mathematics of the individual model development that in any way links it to the underlying diameter data. While some of these models were originally derived with specific applications in mind (e.g., the Weibull model), in forestry they are applied to diameter distributions for their flexibility and, at least early on, the ease with which parameters were able to be estimated. In other words, in the application of such models, there is no intrinsic link to the underlying population processes that gave rise to the observed diameter distribution for which a model is desired: these models indiscriminately fit astronomical data, for example, as well as they do tree diameters. Thus, one might perceive that a weakness to the so-called assumed distribution approach is that these are, with some exceptions, just mathematical constructs that have no inherent link to the processes involved in the generation of the observed data. On the other hand, a perceived strength to this approach is perhaps the opposite: lack of any process or mechanistic theory results in simple models with modest data requirements that can fit a wide variety of observed data.

Two extensions to the above assumed models were developed that begin to address the issue of the data-

or application-agnostic approach. The first, parameter prediction (PP), was introduced by Clutter and Bennett (1965), whose models were subsequently independently evaluated by Burkhart (1971). Many other PP studies followed in the coming years (e.g., (McGee and Della-Bianca 1967; Rennolls et al. 1985)), some with interesting alternatives and extensions as part of the least squares estimation process for the prediction equations (e.g., (Cao 2004; Mehtatalo 2005; Mehtatalo et al. 2011; Poudel and Cao 2013; Robinson 2004; Siipilehto et al. 2007)). These authors demonstrated how distributional parameters could be predicted or estimated based on stand attributes like age, number of trees, basal area, and site index. In the PP method, simple regression relationships were developed from individual plot data where an assumed diameter distribution had been fitted to the observed plot distribution. The data were combined over many plots with a range of stand conditions to form a representative data set, and regression equations calibrated to predict the total number of trees in a stand along with the parameters of an assumed distribution model. The idea was that the regression predictions would tie the estimation of a stand's distribution parameters (and thus the distribution itself) to the stand attributes, thus providing an empirical connection between the two.

Hyink and Moser (1983) generalized the PP model approach and introduced a second related method, which they called the parameter recovery (PR) method (see also, (Hyink 1980)). Similarly to PP methods, PR methods used stand attributes such as basal area and mean stand diameter in a method of moments-type estimation setting. Stand averages or totals could be predicted from relevant growth equations within the system, and then the stand diameter distribution 'recovered' from these predicted values providing there were as many moment equations as parameters in the distribution to be estimated (e.g., (Matney and Sullivan 1982; Burk and Newberry 1984; Lynch and Moser 1986; Murphy and Farrar 1988)). As an alternative or variant to using moments, various percentages of the distribution have also been employed (e.g., (Baldwin Jr and Feduccia 1987; Bailey et al. 1989; Brooks et al. 1992; Knowe 1992)). The moment-based concept of parameter recovery means that the method does not require the prediction of stand variables through equations, it can be applied in a simple estimation scheme matching empirical stand quantities from an inventory to theoretical stand-based moment equations for parameter estimation. These studies in the form of PP and PR models were a step towards the synthesis of stand-based attributes and the associated form of the assumed distribution for diameters, and may be thought of largely as pure stand-based approaches.

The methods of PP and PR might be considered intermediate methods as they are based on models that rely

wholly on stand attributes rather than individual tree attributes. An alternative approach is to couple the stand diameter distribution with the individual tree demographic equations that underlie the dynamics that gave rise to it. An early approach to the solution of this problem was given by Bailey (1980). He presented a method that used the regenerative property of distributions to characterize the diameter increment required in order to ‘regenerate’ the distribution family at time t_2 based on the form assumed at time t_1 (where $t_1 < t_2$), and applied it to several different families of distributions that included the Weibull and beta models. Bailey (1980)’s method was innovative and insightful in producing a method that mathematically linked the diameter distribution to individual tree diameter growth equations. However, it was incomplete in the sense that it lacked a companion theory for births (regeneration or ingrowth) and deaths within the system. The stated assumption was that either there was no mortality, or that it was proportionately distributed over the diameter classes. This limitation was recognized by Cao (1997), who later incorporated a mortality component into Bailey (1980)’s approach. Of further note, Matney and Sullivan (1982) developed a compatible diameter projection model based on Bailey (1980)’s work that was constrained by their stand-based PR projection model. More recently, Matney and Schultz (2008) developed a method for deducing the transformation equations that link a distribution at t_2 from both the initial distribution at t_1 and the distribution of trees surviving from t_1 to t_2 . Clearly, there appears to be an interest in linking the underlying vital rates to distributions of diameter at one or more points in time based on these and other studies.

A general and arguably more flexible approach was taken in population biology where such individual-based models are often referred to as “physiologically structured” population models, replacing the concept of chronological age with “physiological age” (VanSickle 1977; de Roos 1997); here, physiological age refers to some size attribute of an individual. These models are also referred to as size-structured distribution (SSD) models (e.g., (Botsford et al. 1994)). Perhaps the best known example of a SSD, used by these authors and others, arises from the partial differential equation

$$\frac{\partial n(d, t)}{\partial t} + \frac{\partial [g(d, t)n(d, t)]}{\partial d} = -m(d, t)n(d, t) \quad (1)$$

which has units $\text{number} \cdot \text{cm}^{-1} \cdot \text{yr}^{-1}$. This equation states that the rate of change of numbers, $\frac{\partial n(d, t)}{\partial t}$, in a very small diameter (at breast height, DBH) class $d + \Delta d$ is balanced with the rate of growth of individuals through the class, $\frac{\partial [g(d, t)n(d, t)]}{\partial d}$, and the decline in individuals through mortality, $[m(d, t)n(d, t)]$, where t is time. Here the numbers

density, $n(d, t)$ ($\text{number} \cdot \text{cm}^{-1}$), rather than a probability density, is evolved through time. The connection to a PDF, however, is simple since $N_t = \int_{d_0}^{d_\infty} n(d, t) dd$, and therefore $n(d, t)$ is simply a scaled version of a PDF, with minimum and maximum diameters respectively $d_0 = 0$ and $d_\infty = \infty$ in general. The annual diameter growth rate is given in the model by $g(d, t) \text{ cm} \cdot \text{yr}^{-1}$, and the per capita mortality rate is $m(d, t) \text{ yr}^{-1}$. When a boundary condition is included for “births” into the population, this model intrinsically links the three vital rates of recruitment, growth, and mortality into the dynamics of the numbers density. These SSD models are not assumed models, or even hybrid, they are “inherent” diameter distribution models because of the implicit link with the vital rates in (1).

Each of the above three classes of diameter distribution model has different data requirements. Assumed models normally require only the empirical diameter distribution to be fitted. The hybrid PR and PP models potentially require more plot data along with stand attributes for prediction model calibration. Size-structured models are the most data-intense, requiring growth data for calibration of $g(d, t)$, or an existing growth model; similarly for $m(d, t)$ and the recruitment rate. However, sometimes the mortality and recruitment parameters can be estimated without such measurements as in the case of constant mortality under the negative exponential model (Gove 2017).

In some cases—i.e., under appropriate vital rate form assumptions—the steady state SSD model can be shown to analytically conform to an individual, or potentially a family of, assumed distributional forms. In forestry, Muller-Landau et al. (2006) was evidently the first to have made such a connection, demonstrating some specific vital rate models that led to negative exponential, and Weibull distributions. Later, Gove (2017) expanded upon the vital rate model forms that would give rise to a negative exponential. Among the benefits of showing how an implicit distribution relates to some individual or family of assumed distributions, without losing the link to the underlying vital rate parameters, is the inclusion of all the theoretical work related to that family of distributions, though on a different set of distributional parameters. The purpose of this paper is to show how the steady state “Chapman-Richards” distribution (CRD) is actually a subset of the generalized beta distribution of the first kind (GB₁) (McDonald 1984). In addition, a parameter estimation scheme that preserves the essence of the SSD, while simultaneously also fitting the GB₁ distribution is described. It will be demonstrated that the change to the GB₁ form actually presents a simpler equation for the estimation of mortality rate than that arising from the CRD. This simplification, as it were, also lends itself to a much more tractable version of the size-biased form:

specifically, the basal area-size distribution (Gove and Patil 1998). Note that under this scheme, the GB₁ is not an assumed distribution, it is simply a different mathematical representation of the CRD and thus shares all of the inherent coupling of vital rates as the CRD form.

Methods

The Chapman-Richards Distribution

The steady state or equilibrium solution of the general SSD model in (1) is one in which the distribution remains constant through time (found by setting $\frac{\partial n(d,t)}{\partial t} = 0$) and is given by...

$$n(d) = \frac{R}{g(d)} \exp\left(-\int_{d_0}^d \frac{m(d')}{g(d')} dd'\right) \quad d > d_0 \quad (2)$$

$$= \frac{R}{g(d)} S(d)$$

where the exponential term is the survival function, $S(d)$, and R is the recruitment rate (note that the time index is no longer required as time-invariant dynamics are assumed). To find the associated PDF, simply divide the numbers density by the total population size, N .

The assumptions for the vital rates underlying the Chapman-Richards distribution are that mortality is constant, $m(d) \equiv M$, like recruitment. In addition, annual diameter growth follows the Chapman-Richards growth function (Pienaar and Turnbull 1973)

$$g(d) = \frac{dd}{dt} = \eta d^m - \gamma d \quad 0 \leq d < d_\infty \quad (3)$$

with

$$d_\infty = \left(\frac{\gamma}{\eta}\right)^{\frac{1}{m-1}} \quad (4)$$

where m , η , and γ are model parameters to be estimated from the data. The maximum DBH (d_∞) is found by setting the derivative of (3) to zero and solving. It is straightforward to show (Additional file 1: Section S.2) that substitution of these vital rate forms into (2) yields the equilibrium distribution given by

$$n(d) = \frac{R}{\eta d^m - \gamma d} \left(\frac{d^{m-1} (\eta d_0^{m-1} - \gamma)}{d_0^{m-1} (\eta d^{m-1} - \gamma)} \right)^{\frac{M}{\gamma(m-1)}} \quad (5)$$

$$= \frac{R}{\eta d^m - \gamma d} S'(d) \quad 0 \leq d < d_\infty$$

A similar form of this distribution is due originally to Zavala et al. (2007), who used a slightly different parameterization for (3). It was subsequently rederived in the current form by Gove (2017). Neither of these studies named the distribution, which has been remedied here. It is straightforward to verify that the survival function follows immediately via comparison to (2) such that $S(d) \equiv S'(d)$ under growth relation (3) with constant mortality

(see Additional file 1: Section S.2 for the full derivation). And, of course, the CR version of the survival function is a much simpler form because the integration has already been accomplished.

Note that many other forms of mortality models may be coupled with the CR growth equation, so the current form may turn out to be just one form in a larger, more complex family. However, more complex examples are not guaranteed to have closed-form solutions like (5), and can require numerical integration of the survival function for evaluation. The closed-form solution in (5) makes the CRD easier to work with, and lends it to further manipulation—even simplification—as demonstrated below. Note that the range in both parameters and diameters is inherited from the CR growth equation; specifically, the distribution will have finite d_∞ .

The relationship of the CRD to the GB₁

Ducey and Gove (2015) have recently demonstrated that the GB₁ distribution and its associated size-biased form may have some potential applications to diameter distribution modeling in forestry. The GB₁ PDF is given by

$$f(d; a, b, p, q) = \frac{|a| d^{ap-1} \left(1 - \left(\frac{d}{b}\right)^a\right)^{q-1}}{b^{ap} B(p, q)} \quad b, p, q > 0 \quad (6)$$

with $0 < d^a < b^a$, and $B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$ is the beta function.

A subset of the full GB₁ distribution can be derived from the CRD (Additional file 1: Section S.3), which yields the numbers density

$$n(d) = \tilde{N} \frac{|a| d^{ap-1} \left(1 - \left(\frac{d}{b}\right)^a\right)^{q-1}}{b^{ap} B(p, q)} \quad (7)$$

where

$$\tilde{N} = R \left[\eta - \frac{\gamma}{d_0^{m-1}} \right]^{\frac{M}{\gamma(m-1)}} \times \eta^{\frac{M}{\gamma(1-m)}-1} \times \frac{b^{ap} B(p, q)}{|a|} \quad (8)$$

and because the numbers density must integrate to N , we have that

$$N = \tilde{N} \int_{d_0}^{d_\infty} \frac{|a| d^{ap-1} \left(1 - \left(\frac{d}{b}\right)^a\right)^{q-1}}{b^{ap} B(p, q)} dd$$

The relationship of the parameters in (7) to those of the CRD (and growth equation) are: $b = d_\infty$, $a = 1 - m$, $p = 1$ and $q = \frac{M}{\gamma(1-m)}$. The fact that GB₁ parameter p is unity makes (7) a subset of the full generalized beta given in (6); thus, we will refer to the reduced form given in (7)

as the Chapman-Richards GB₁ (CR-GB₁). Note that \tilde{N} is a normalizing constant that makes the right hand side in (7) a numbers density rather than a PDF. Since the remainder of the right-hand side is a PDF, this means that the constant \tilde{N} has a physical interpretation as an estimate of the number of individuals in the population, N . It is worth noting that the normalizing factors for most PDFs have no such physical meaning, therefore this is a somewhat unique model.

The CR-GB₁ contains only a subset of the fully recognizable distributional forms of the GB₁ (McDonald 1984; Ducey and Gove 2015) due to the constraint that $p = 1$. Nevertheless it still has a range of flexibility that includes both positively and negatively skewed, concave, mildly rotated-sigmoid, reverse *J*-shaped, uniform and U-shaped forms as noted by Gove (2017). These forms are all achieved based on the full set of parameters inherited from the CR growth model and CRD itself, since p is not related to any of these parameters. In addition, the two-parameter Weibull and the negative exponential distributions can be derived through limiting arguments as in McDonald (1984), though as we will see later, these forms are of limited value as derived through the CR-GB₁ size-structured setting.

Basal Area-Size Distribution

One useful consequence of the CRD and CR-GB₁ equality is that quantities that might be tedious to derive from (5) may already exist for the CR-GB₁ form. Notably, any result that has been derived for the GB₁ distribution (including all of the special case distribution forms, moments, Gini coefficient, etc.) is applicable to the CR-GB₁ with the above parameterization. One recent application of this is in Ducey and Gove (2015), who derived the size-biased version of the GB₁. A convenient relationship useful in forestry is that the distribution of basal area over DBH—termed the basal area-size distribution (BASD) (Gove and Patil 1998)—is the size-biased distribution of order $\alpha = 2$. In general, if $D \sim f(d; \theta)$ (where θ can be vector-valued) then the size-biased distribution of order α is $D_\alpha^* \sim f_\alpha^*(d; \theta)$ where

$$f_\alpha^*(d; \theta) = \frac{d^\alpha f(d; \theta)}{\int d^\alpha f(d; \theta) dd} \quad (9)$$

Thus, if $D \sim \text{GB}_1(d; a, b, p, q)$, then it follows that $D_2^* \sim \text{GB}_1(d; a, b, p + \frac{2}{a}, q)$ (eq. 9, (Ducey and Gove 2015)). It is therefore straightforward to show that the BASD form of the CRD is given by

$$b(d) = B \frac{|a| d^{ap'-1} \left(1 - \left(\frac{d}{b}\right)^a\right)^{q-1}}{b^{ap'} B(p', q)} \quad (10)$$

where $p' = 1 + \frac{2}{a}$, B is total stand basal area, and $b(d)$ is the basal area density at diameter d , analogous to the numbers

density $n(d)$ (of course, both $b(d)$ and $n(d)$ conforming to the normal continuous density interpretation). It is worth noting here that the PDF component of (10) is a full GB₁ form, since the parameter restriction on p has been lifted through its dependence on the CR growth parameter m . This will result in a richer set of distributional shapes for basal area in general.

Notice in (10) that the total basal area, B , is required. Normally an inventory will yield an estimate of this figure from the sum of the individual tree basal areas per unit area. However, an inventory may not yield a compatible estimate for the entire basal area distribution over full support of the CRD (d_0, d_∞) matching the GB₁ estimate of population size, \tilde{N} , but, the size-biased CRD relationship does indeed provide such an estimate. Recall the simple relationship of numbers and basal area to the quadratic mean stand diameter (\bar{D}_q), (e.g., (Gove and Patil 1998)); viz.,

$$\tilde{B} = \kappa \bar{D}_q^2 \tilde{N} \quad (11)$$

where $\bar{D}_q^2 \equiv \mu'_2$ such that

$$\mu'_\alpha = \frac{b^\alpha B(p + \frac{\alpha}{a}, q)}{B(p, q)} \quad (12)$$

is the α th raw moment of the GB₁ density (Ducey and Gove 2015); and κ the conversion factor from diameter-squared to basal area. Thus substituting \tilde{B} in for B in (10) yields the theoretical basal area density that will exactly match that of the theoretical numbers density.

Parameter estimation

In order to fit the CR growth model, the diameter increment data, $\frac{dd}{dt}$, must be available from increment cores, dendrometer bands, permanent remeasured plots or the like. In the latter case, mortality data would also be available. In this case the procedure would be to fit the growth model using the methods described below, and estimate the mortality rate, M , from the mean of the mortality data. If ingrowth data are available, this would similarly provide an estimate of the recruitment rate, R . If ingrowth data are not available, but mortality rate and the number of trees, N , are known (i.e., estimated) for the population in question, then the recruitment rate can be estimated from the equilibrium relation

$$R = MN \quad (13)$$

When only the diameter increment and DBH data are available, we must resort to slightly more sophisticated methods for estimating recruitment and mortality. A two-step estimation process is proposed where (i) the CR growth equation is estimated from the diameter increment data, possibly on a data set that is independent or a subset of the population for which the final numbers density is desired; and (ii) the mortality rate is estimated using

the empirical diameter distribution for the population of interest. These steps are straightforward and are detailed in the following sections.

Estimating the CR growth curve

Two basic approaches are available for the estimation of parameters in (3): unconstrained and constrained nonlinear least squares (NLS). The unconstrained problem is straightforward, and is often reasonable for general growth modeling applications. However, there is a potential problem with this approach given the desire to fully estimate the CRD (5) (or, equivalently, the CR-GB₁ (7)). The distribution inherits the upper bound on diameter, d_∞ , from the growth curve, and fitting the curve with simple NLS can produce a maximum diameter that is either higher or lower than what is found in the stand distribution under consideration. If the estimated d_∞ is larger than the observed maximum diameter, d_{\max} , then there will be extra density above the largest tree in the population. This is not such a bad situation if the overall density is small in this diameter range, as the growth curve will still predict diameters encompassing the largest tree in the population. More problematic is when the estimated d_∞ is less than the largest tree in the population. The growth function does not exist for trees larger than the estimated d_∞ , nor does the density, both of which are clearly present in the population. This latter case may occur when the larger trees in the population were not represented in the sample of growth trees used to fit the regression. The first case can occur regardless, when the growth on the larger trees show large variation.

An elegant method was proposed by Shifley and Brand (1984) to constrain the CR growth model so that it terminated at some observed or target maximum diameter. These authors used relation (4) to re-write (3) in terms of d_∞ rather than γ , in the form

$$\frac{dd}{dt} = \eta \left(d^m - d_\infty^{(m-1)} d \right) \quad 0 \leq d < d_\infty \quad (14)$$

This equation involves only two unknown parameters to be estimated, η and m , since d_∞ is specified as a constraint and is assumed known from measurements. After this is fitted, one may easily obtain an estimate for γ , to get the third desired CR parameter by simple re-arrangement of (4) as $\gamma = \eta d_\infty^{m-1}$. This constrained nonlinear least squares (CNLS) method was adopted in this study to fit the CR function in the first phase of estimation. As noted above, this will fix the target d_∞ , which will then be inherited by the CR-GB₁, ensuring the appropriate coverage in the numbers density for the range of diameters in the target population.

Estimating mortality

There are many methods that could be used to estimate mortality. One such method, parameter recovery,

has been discussed earlier. In this case a simple estimating equation based on basal area similar to that presented in Gove and Patil (1998) and Gove (2004) can be developed and solved as a simple sums-of-squares minimization problem. However, an even simpler approach is to apply the principles of maximum likelihood (ML) to either (5) or (7) to develop an estimator for mortality. While both approaches produce closed-form solutions, it turns out that the CR-GB₁ form, (7), is easier to work with, and produces a simpler estimator equation than (5). The ML estimator for mortality based on (7) is (Additional file 1: Section S.3.1)

$$\hat{M} = n\gamma(m-1) \left\{ \sum_{i=1}^n \ln \left(1 - \left(\frac{d_i}{d_\infty} \right)^{1-m} \right) \right\}^{-1} \quad (15)$$

where $d_i, i = 1, \dots, n$ is a sample of diameters representing an empirical diameter distribution for the population of interest. Note that this is written in terms of the estimated CR growth parameters γ, m , and d_∞ , though it can be easily re-expressed in terms of η rather than γ (e.g., Additional file 1: equation (S.9)). As a final step, if no ingrowth estimates are available, then recruitment can be estimated using relationship (13).

Simulations

A set of simulations was designed to test the convergence of the simple two-step parameter estimation scheme outlined above. The simulations look at both the effect of sample size and magnitude of added noise to the growth increments on the convergence of the estimated growth equations and the overall distribution. In the simulations it is assumed that n diameters are drawn from the respective population distribution and that growth increment is determined on each tree in the sample. Furthermore, the growth is computed using (3) with the population parameters and additive Gaussian noise; viz.,

$$g(d) = \eta d^m - \gamma d + \varepsilon \quad 0 \leq d < d_\infty$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Two different levels of Gaussian noise were added to the growth increments with standard deviations σ_1 and $\sigma_2 \text{ cm} \cdot \text{yr}^{-1}$ depending on the distribution shape.

The diameters were drawn from four different distributional shapes that roughly characterize the flexibility of the CR-GB₁ distribution. These shapes are listed in Table 1 along with the population parameters used in the simulations; in addition to the growth and vital rate parameters, the populations size \tilde{N} based on (13) and the total stand basal area, \tilde{B} from (11), are also given. Sample sizes of $n = 25, 50, 100$ and 200 trees were drawn in a Monte Carlo experiment for each of the four shape and two noise level combinations, each with $\mathfrak{N} = 500$ Monte Carlo replicates. In each replicate, the maximum diameter

Table 1 The population parameter values for four different distributional shapes arising from the CRD

Population parameter	Concave	Positively Skewed	Rotated-Sigmoid	U-Shaped (Bathtub)
m	-0.180	-0.280	0.260	0.540
η	0.840	0.540	0.330	0.480
γ	0.012	0.004	0.019	0.088
R	20.000	20.000	20.000	21.000
M	0.019	0.019	0.019	0.026
d_{∞}	36.614	46.166	47.353	39.962
\tilde{N}	1052.632	1052.632	1052.632	807.692
\tilde{B}	32.474	19.939	35.858	31.689
σ_1	0.080	0.080	0.100	0.080
σ_2	0.150	0.120	0.180	0.160

was fixed at the population value of d_{∞} for estimation of the growth curve via CNLS. This produced a simple factorial design with 32 factor levels. Convergence in growth equations and distribution is demonstrated graphically in the results, while a set of Kolmogorov-Smirnov (KS) tests ([p. 344], (Conover 1980)) is also presented as an index of fit for the distributional results.

Results

Growth Estimation

There is little new *per se* in our results regarding the estimation of the CR growth function by nonlinear least squares, though the extensive use of the constrained approach in a Monte Carlo setting, which was well-behaved overall, may be somewhat novel. Here, however, the focus is not on the efficacy of the CNLS approach in general, rather the intent is to demonstrate the effects of the assumed noise levels on the estimation of the CR growth parameters by comparing the perturbed results to the population. Moreover, the ultimate goal is to illustrate how the variability in the growth increment data—through its effect on the CNLS fit results—affects the final estimation of mortality and recruitment in the second phase of estimation.

Effects of noise

Figure 1 illustrates the differences in the growth increments due to the noise parameter, σ , using $n = 1,000$ diameters drawn from the Rotated-Sigmoid distribution (Table 1). The set of actual diameters drawn are the same in Fig. 1a,b, only the growth increments differ due to the difference in noise standard deviation. Note that noise values that would cause negative increments were discarded (and a new value re-drawn) in Fig. 1 and the simulations below, resulting in a mildly truncated normal distribution for some values of the individual noise, ε_i . This is minor for both levels, but more pronounced for the higher noise

standard deviation, σ_2 . The increments in Fig. 1a show a scatter that might be reasonably associated with the fit of this model. The figure clearly shows that the CNLS fit for these data is quite close to the population curve from which it was generated, and that d_{∞} is equivalent for both these curves. The unconstrained NLS fit is also presented, showing how the scatter can result in a poorer fit to these data, with larger estimated value for d_{∞} from the associated coefficient estimates. Figure 1b illustrates the result of the larger noise associated with the increments. The fact that the underlying model is shown by the population curve in this case is less convincing due to the high degree of scatter. The resultant fit for the CNLS curve is obviously worse (in the sense of matching the population curve) than that for Fig. 1a, though d_{∞} is again preserved to the population value as desired. The unconstrained NLS fit is poor, with an estimated maximum diameter that is much larger than the population value of d_{∞} due to the high degree of scatter in increment values around the upper DBH limit, effectively lifting the curve at the population d_{∞} value into the mean of the increment values at the larger diameters. This simple example clearly illustrates that the noise corruption added under the standard deviation σ_1 can provide reasonable estimates close to the population parameter values for a large sample size. On the other hand, increments showing more scatter under σ_2 may not lead to parameter estimates that are as close to the population values. Keep in mind that these results are based on a larger sample of diameters and growth than what are used in the simulations below. Corresponding graphs for the other three distributional shapes listed in Table 1 are shown in Figure S.5 in the Additional file 1, where similar interpretations apply.

Convergence in growth

The methods described in the previous section were applied to the estimation of the CR growth function under each of the four distributional shapes with each of four levels of sample size, n , and two levels of standard deviation, each repeated $\mathfrak{N} = 500$ times. Figure 2 presents the results for the Rotated-Sigmoid shape with a subset of individual growth curves shown. As would be expected, regardless of perturbation level, the individual Monte Carlo estimated curves tend to converge (the spread becomes tighter) to the population growth curve as sample size increases. This phenomenon is much more pronounced for the σ_1 curves, such that the $n_{\sigma_1} = 25$ ensemble set (where n_{σ_1} denotes the sample size n for the σ_1 set) is roughly equivalent in spread to the $n_{\sigma_2} = 100$ curves, while the $n_{\sigma_1} = 50$ and $n_{\sigma_2} = 200$ results are also approximately comparable. The mean curves in each panel denote the curve that corresponds to the set of mean parameter values for each of the \mathfrak{N} ensemble members. These mean curves track the population curve

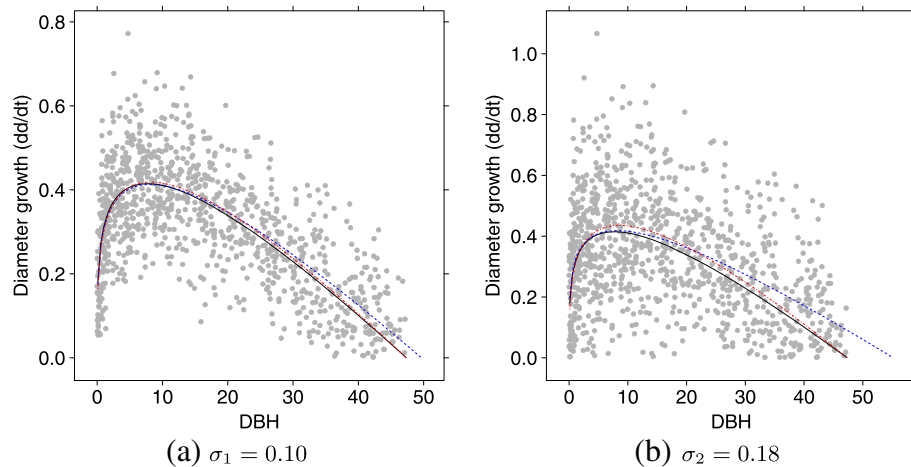


Fig. 1 A sample of $n = 1,000$ diameters from the Rotated-Sigmoid population for **a** $\sigma_1 = 0.10$ and **b** $\sigma_2 = 0.18$, demonstrating the effect of noise on $\frac{dd}{dt}$ increment variability, with the population curve (solid), the unconstrained fit (dashed blue), and the constrained fit (dot-dashed red). Note the difference in scale for the growth increments between the two figures. Similar results for the remaining three shapes in Table 1 are in Additional file 1: Figure S.5

well for all levels of n_{σ_1} , and for larger sample sizes in n_{σ_2} , but the mean curve for $n_{\sigma_2} = 25$ is quite poor, yielding a very low estimate for d_{∞} . The results for the U-shaped distribution growth curves (Figure S.8) are similar to those for the Rotated-Sigmoid distribution. The U-shaped ensemble spread appears tighter in most cases than those for Rotated-Sigmoid but this is due in part to the scale of the two curves: while the shape is similar, the

U-shaped population growth curve has a maximum diameter growth at twice that of the Rotated-Sigmoid curve. The mean curves for σ_2 converge more slowly for the U-shaped ensembles.

The growth curves for the other two distributional shapes, Concave and Positive-Skew, show a continuous decline in growth from the smallest to largest diameter classes yielding mildly reverse *J*-shaped curves overall

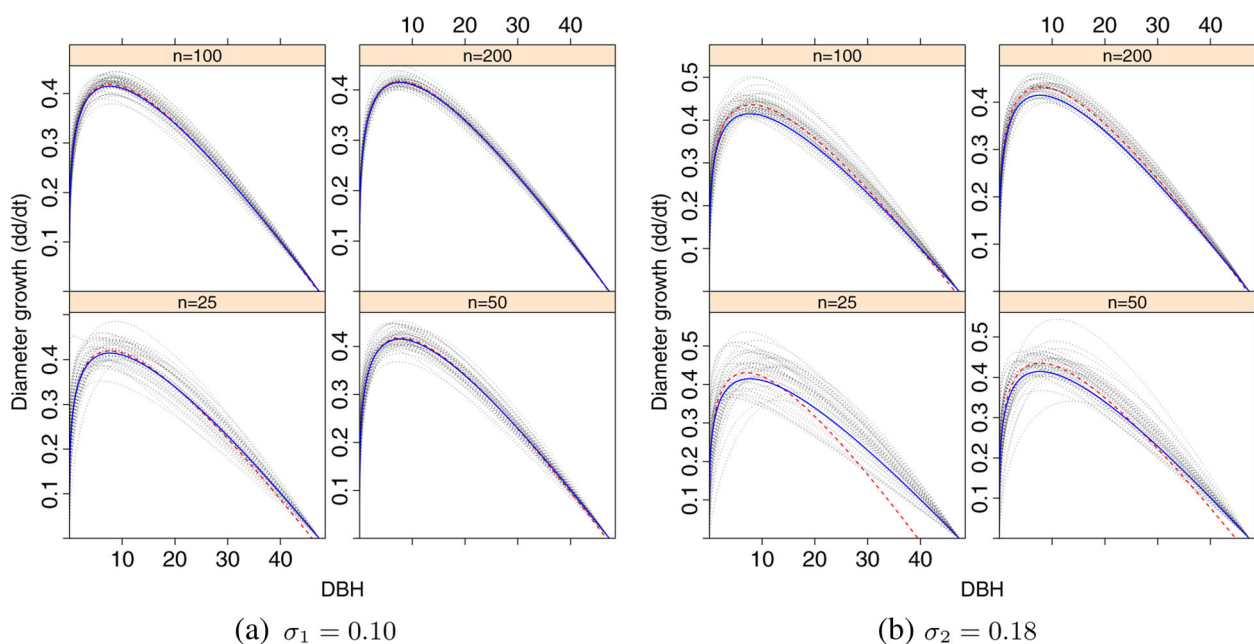


Fig. 2 Convergence of growth in sample size, n , represented by a subsample of $\mathfrak{N} = 50$ Monte Carlo replicates (dotted) for the Rotated-Sigmoid population (solid blue) and mean (of the full $\mathfrak{N} = 500$ replicates) curve (dashed red) for **a** $\sigma_1 = 0.10$ and **b** $\sigma_2 = 0.18$. Similar results for the remaining three shapes in Table 1 are in Additional file 1: Figure S.6–S.8

(Figures S.6 and S.7), unlike the concave Rotated-Sigmoid and U-shaped curves. As diameter nears zero in both sets, the predicted growth can become quite large and is truncated for display at a diameter of 0.1 cm in both cases. Also, in both cases, regardless of noise level, there is at least one estimated curve in the ensemble that flips to a concave shape at the $n = 25$ level. For the Concave- σ_2 curves, only $n_{\sigma_2} = 200$ lacks such a flip. The ensemble growth curves are all well-estimated for $n_{\sigma_1} > 25$ over both distributional shapes. When more noise is added to the estimation scenario, n_{σ_2} of at least 100 appears necessary for consistent estimates under both distributional shapes.

MLE Mortality Estimator Convergence

The estimation of the CR growth parameters through nonlinear least squares is, in general, well supported. The results in the previous section indicate that lower noise in the growth observations lead to better behaved curves; and in each case, the model curves tend to converge as sample size increases. None of this is surprising. In the two-phase estimation scheme proposed, the results from the last section may now be considered as fixed constants and incorporated into the ML estimator for mortality, (15). This section provides details on the convergence of this estimator where the simulation results from the last section are used as a basis for conditional estimation. Recall that diameter growth only enters the model estimation phase through the estimation of the growth equations; and the same diameter values drawn for use in those simulations are used here for the mortality results.

Efron and Hinkley (1978) advocate the use of the observed rather than the expected Fisher information for establishing confidence intervals on individual estimates. It is easily shown (Additional file 1: Section S.6.1) that an estimator for the variance of the MLE using the observed information is given by

$$\widehat{\text{Var}}(\hat{M}) = \frac{M^2}{n} \quad (16)$$

Likelihood theory then provides that an asymptotic confidence interval for the estimate is given by

$$\hat{M} \pm 1.96 \sqrt{\widehat{\text{Var}}(\hat{M})} \quad (17)$$

The relationship in (17) was used to establish approximate confidence intervals on the MLE for mortality in each of the Monte Carlo simulations. The percentage of times these confidence intervals capture the population mortality parameter, M , is given in Fig. 3 (with full results in Additional file 1: Table S.1). The capture rates for $n_{\sigma_1} \geq 100$ are in the range of ~ 94 – 96% for all shapes, with only slightly lower rates for the U-shaped distribution attributable to chance. The smaller sample sizes,

$n_{\sigma_1} < 100$, show mixed results as would be expected. With the exception of the Rotated-Sigmoid shape, the results for $n_{\sigma_1} < 100$ vary in a non-systematic manner for each different shape, with some capture rates approaching $\sim 92\%$ (Positive-Skew) and others as high as $\sim 96\%$ (Concave).

The results for n_{σ_2} are more varied and less conclusive as would be expected due to the less reliable results of the fitted growth equations: if the underlying growth equations are not estimated accurately, then conditional basis for application of the ML estimator is reduced. There is considerable ‘randomness’ associated with all shapes, even the well-behaved Rotated-Sigmoid shape suffers from a decline in capture rate to $\sim 91\%$ at the largest sample size. Clearly, the conditional foundation on which (15) is applied for likelihood estimation is diminished to the point where the results are of questionable value when larger variations in diameter growth observations lead to poorer fit in the CR function. Recall that the intervals given by (17) are asymptotic, and therefore, the tendency for decreased capture rates at the largest sample size is troubling: the lower rates may be a more honest indicator of the MLEs under poorly estimated growth equations.

A second approach is to regard the Monte Carlo estimates as the random samples that they are and establish sample-based confidence intervals based on the distribution of sample means using the average mortality estimate, \hat{M} , and the standard error of the estimate. This approach provides one summary confidence interval for each shape- n_{σ} level, the results of which are presented in Fig. 4. Similar to the likelihood-based intervals, the sample-based intervals tend to decrease as sample size increases. The average mortality estimates for σ_1 converge by approximately $n = 100$; the small upturn for the Concave and Positive-Skew shapes are negligible and attributable to chance. The results for σ_2 appear to be less accurate, with mean estimates converging only the U-shaped distribution. These results corroborate those of Fig. 3, indicating that the estimates for the growth equations produce a less reliable MLE on average as the variability in the observed growth increases.

Convergence in Distribution

The Monte Carlo ensemble sets of growth curve parameter estimates and their associated MLEs for mortality from the two-phase estimation results in the previous two sections were used to generate the corresponding numbers densities (5) (or equivalently (7)); recruitment was estimated using (13), substituting the MLE for mortality into this estimator. This provides a complete set of numbers densities, which can be compared to the population density at each experimental level. Recall that the same diameters were used for both levels of σ in both phases of estimation for each Monte Carlo sample at each given shape-sample size level, so that only the variability

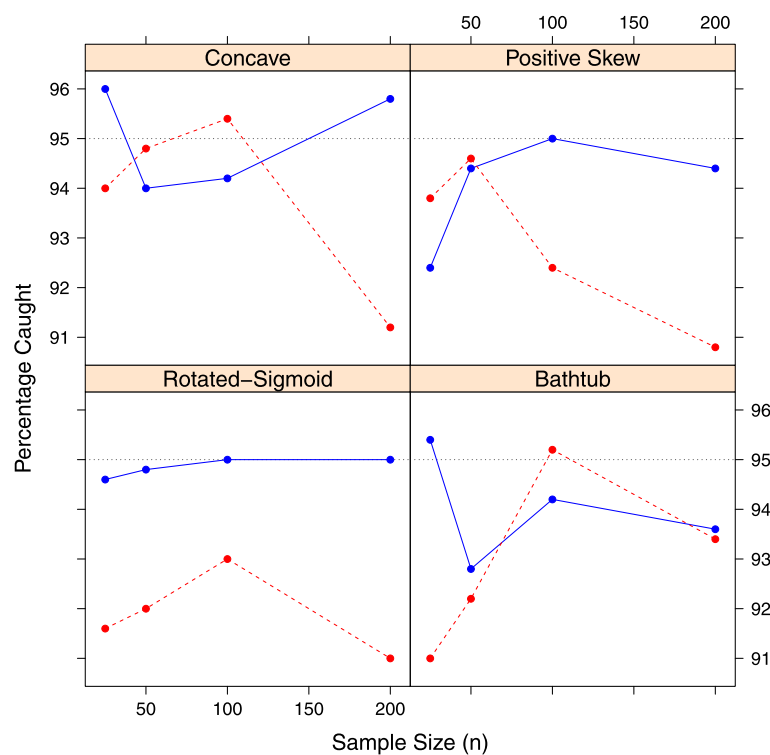


Fig. 3 Percent catch for $\mathfrak{N} = 500$ individual Monte Carlo observed Fisher information confidence intervals on estimated mortality at the nominal $\alpha = 0.05$ level by sample size, n , for σ_1 (solid blue) and σ_2 (dashed red) within each shape. Complete results are found in Additional file 1: Table S.1

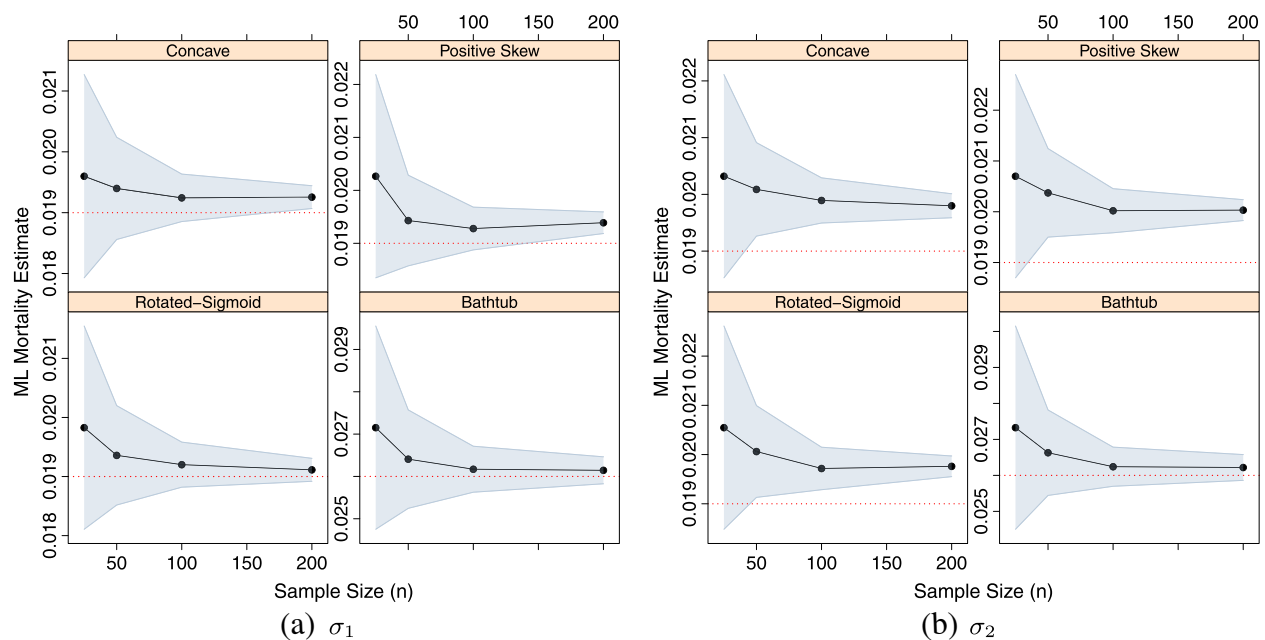
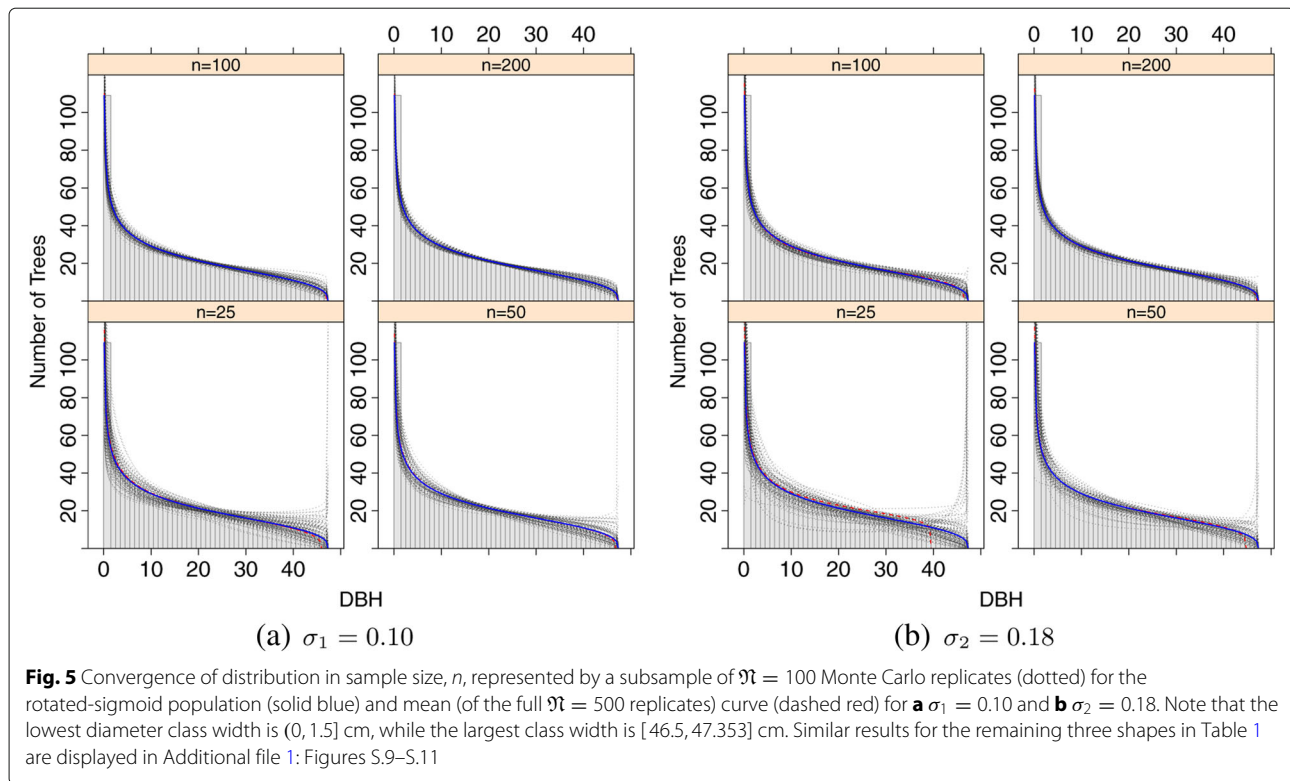


Fig. 4 Convergence of average ML mortality estimates, \hat{M} , (solid dots) in sample size, n , from $\mathfrak{N} = 500$ Monte Carlo replicates per level. The population value, M (dotted) and sample-based 95% confidence intervals (shaded) are also shown. **a** σ_1 . **b** σ_2



in the growth observations is responsible for observed differences between distributions at each level.

The results for the Rotated-Sigmoid distribution are presented in Fig. 5 using a random subset of approximately $\mathfrak{N} = 100$ ensemble members out of the full complement of 500 at each shape- n_σ level. For $n_{\sigma_1} < 100$ the individual distributions appear to fit reasonably well, except that a few aberrant U-shaped distributions tend to appear. Recall that the growth functions for Rotated-Sigmoid and U-shaped distributions both share a similar shape (compare Figs. 2 and S.8), with the main difference being that of larger growth increments and higher average mortality rate (Fig. 4) in the latter; such a random draw coupled with the mortality estimate will evidently cause a flip in shape in these few realizations. Distributions for $n_{\sigma_1} \geq 100$ are all generally well-behaved, with the mean densities (over all $\mathfrak{N} = 500$ members) converging quickly by $n_{\sigma_1} \approx 50$. Similar trends hold for the σ_2 results in Fig. 5b, except that the number of aberrant distributions has increased for $n_{\sigma_2} < 100$, with a few U-shaped forms continuing to appear in the larger sample sizes. The difference in overall fit is also shown by the mean curves, which are poorly estimated for the levels with $n_{\sigma_2} < 100$.

The other shapes (Additional file 1: Figures S.9–S.11) share many features with that of the Rotated-Sigmoid results. All distributions tend to converge with increasing sample size and do so more rapidly at the lower noise level, σ_1 . The Concave ensembles show some sigmoid-shaped

distributions arising at $n_{\sigma_1} < 100$, while a few U-shaped forms also appear in the $n_{\sigma_2} < 100$ groups. Notable also are Concave distributions with some degree of positive skewness. The mean distributions are all remarkably close to the population distribution given the anomalous forms. The Positive-Skew forms are also well-behaved in general with a few tending towards reverse J -shaped or Rotated-Sigmoid, especially in the smallest sample size. The mean distributions are all well-formed for all n_σ levels. Of particular note, d_∞ is very well estimated in all scenarios for both distribution shapes. Finally, in the U-shaped distributions a small quantity of aberrant shapes in the smaller sample sizes for $n_\sigma < 100$ appear, which tend to L-shaped or mildly rotated-sigmoid in form. The mean distribution is well estimated for $n_{\sigma_1} > 50$, but it is poorly estimated for $n_{\sigma_2} < 100$, apparently due in large part to average parameter values that underestimate d_∞ in (4).

In an effort to provide a form of numerical index to facilitate the above qualitative comparison of distribution convergence, KS tests for agreement in distribution were calculated at each combination level. Thus, the test compared each Monte Carlo realization of the distribution against its population counterpart as given in Table 1, resulting in $\mathfrak{N} = 500$ KS tests performed at each shape- n_σ level. The null hypothesis for each such test is that there is no difference between the simulated and population distributions; the rejection level used was $\alpha = 0.05$.

This simulated distribution versus population test is similar to traditional treatment versus control, and the tests are therefore dependent in nature within a given shape. The p -values for each level were adjusted using a method that controls the false discovery rate (FDR) developed by Benjamini and Hochberg (1995); this method was subsequently shown to be applicable to dependent tests by Benjamini and Yekutieli (2001). In essence, the FDR is simply the expected proportion of all rejections that are false as described by these authors; and it is applied to the family of distributions at each shape- n_σ level in the simulations. Fully cognizant of the recent concerns and discussions over the use and misuse of p -values ([e.g., see also associated comments online] (Wasserstein and Lazar 2016)), it must be emphasized that our use of these methods is not for formal testing *per se*, but simply to provide a more tangible way to quantify how the distributions converge than by the Monte Carlo distribution plots alone (e.g., Fig. 5). The results are therefore presented simply as the percentage of rejections at each experimental level, and it is important to note that the α -level used is immaterial in this context as any reasonable cutoff could have been chosen: it is largely the trend of ‘rejections’ with sample size regardless of level that is of interest.

The results of the convergence index calculations just described are presented in Fig. 6 (for details see the

Additional file 1: Table S.2). The results clearly illustrate that the smaller sample sizes are inadequate for estimation of the numbers density regardless of the amount of variability in the growth observations. It is interesting to note that evidently the Concave and U-shaped distributions were the least affected by the difference in noise level in terms of percentage of model rejections at the larger sample sizes. This is all the more interesting because the two growth curves are of different shape class. This suggests the conclusion that these particular distributional shapes are more robust in terms of slight variations in growth curve than perhaps are the other two forms studied (however, see the Discussion). The results for the Rotated-Sigmoid shapes show the largest separation in fit index of any shape-class in the $n = 100$ sample size. As mentioned above, note in Fig. 5 that there are a few σ_2 distributions that still turn U-shaped at this sample size (the full set of $\mathfrak{N} = 500$ —not shown—corroborate this proportionally), while those in the σ_1 set are all quite well behaved. Indeed, the Rotated-Sigmoid $n_{\sigma_2} = 100$ distributions record the most rejections, while the $n_{\sigma_1} = 100$, the least, of the four shape classes (Additional file 1: Table S.2). The results for the Positive-Skew distributions are unremarkable in terms of the index results, shedding no new interpretation on the graphical results.

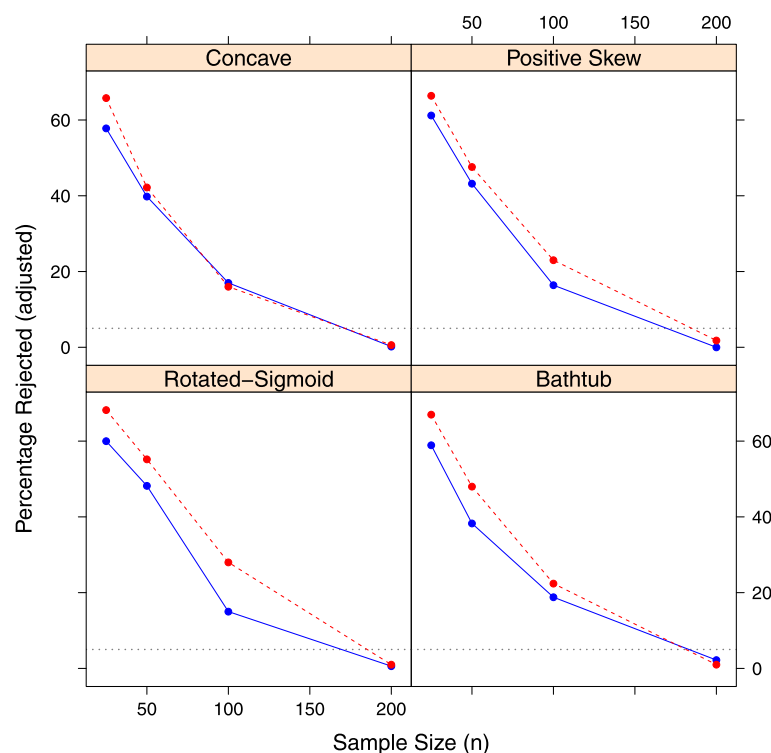


Fig. 6 Percentage rejections of KS tests using adjusted p -values at the $\alpha = 0.05$ level illustrating convergence of distribution in sample size, n , with $\mathfrak{N} = 500$ Monte Carlo replicates for σ_1 (solid blue) and σ_2 (dashed red); the 5% level is also shown (dotted). Complete results are found in Table S.2

Discussion

The Chapman-Richards distribution described here is a fairly flexible family of numbers density curves that are intrinsically tied to the stand demographic drivers of (i) diameter growth that follows the Chapman-Richards growth function, (ii) constant per-capita tree mortality over all size classes and, (iii) constant recruitment. The distribution is presented in its equilibrium form, because this facilitated the link to the GB₁ distribution and the subsequent derivation of the ML mortality estimator. It also facilitated the design and implementation of simulations for the assessment of the proposed two-phase estimation method and should make future comparison with alternative estimation schemes, should they be proposed, simpler. As noted earlier, the CRD has the ability to take a fairly large set of final shapes; a subset of only four were presented here in order to identify the behavior of the proposed model and estimation scheme over a range of shapes. These included two potentially useful forms in forestry in the Positive-Skew and Rotated-Sigmoid shapes, along with the two more extreme shapes represented by the Concave and U-shaped distributions, which would perhaps find less use in forestry diameter distribution applications.

The target, or population distributions in Table 1, were chosen to have approximately the same values for R and M , yielding the same population size, N , for three levels of shape; the U-shaped distribution was found to be unable to support this number of stems at a reasonable level of basal area and maximum diameter. This latter point is important because not every combination of parameters will yield realistic growth curves, final distributional forms, and consequent stocking parameters. However, controlling these variables to be as close as possible as part of the experimental design also reveals very succinctly how the different shapes lead, unsurprisingly, to different basal areas, and demonstrates how changes in the growth parameters alone can affect the curve shapes. This particular design was a determined choice because of the complexity of the model and the number of free parameters and it is straightforward to show that one can similarly fix the growth parameters and manipulate M (since recruitment is a simple scale factor) resulting in a similar range of shapes. Manipulating all parameters can thus yield a very complete set of shapes and upper diameter limits, a more complete enumeration of which, is well beyond the experimental limits of the study.

Given the number of free parameters and the range of shapes just described that could be elicited by either manipulation of the CR growth parameters while holding mortality constant, or changing mortality for a given growth parameter set, it is not surprising that the two-step conditional approach to estimation was a reasonable solution to the estimation problem. It also may be of no

surprise that a maximum likelihood solution on the full parameter set was attempted and failed because of this interaction effect between growth and mortality in the model, leading to multiple parameter set solutions for a given data set. However, this full ML approach was based only on the observed diameter distribution; inclusion of the increments was not attempted. Inclusion of the increments into a joint likelihood could produce issues with scaling, leading to multi-modal likelihoods (e.g., (Gove 1995)).

Indeed, each of the parameters has the ability to affect the shape of the distribution with the exception of recruitment, which is a simple scale factor. The growth function serves to shape the survival function, which, in turn is also dependent on growth. Therefore, each of the growth parameters can change the distribution shape. However, the growth function is fitted to the growth increments and is independent of the stand distribution shape (except in the sense that growth can be density-dependent); and it has been noted that the growth sample need not come from the stand where diameters were sampled, only represent its underlying increments. Thus, it is the stand diameter data that fine tunes the final shape of the distribution. This is determined only by mortality. As noted above, it is easy to show that for given set of growth parameters, changing mortality can change the stand distribution through a variety of shapes similar to those presented here. The two-step procedure for estimation, therefore, appears to make sense. While it would be interesting to have a joint procedure as noted previously, this appears to be unnecessary, and would actually restrict the increment data to be from the same trees as the diameter data (or a subset thereof).

As noted previously with regard to the KS statistic results, the Concave and U-shaped distributions appear to be more robust at all but the smallest sample sizes. However, this interpretation is conditional on the σ values chosen; thus, it may simply be that the chosen values did not impart as much variation for the high level as expected, hence the better results. A certain amount of subjectivity went into the selection of the two σ levels for each growth equation. This was due in part to the monotone reverse J-shape associated with the Concave and Positive-Skew distributions. In the very small diameters the curve increases quickly so that perturbations added to such values can result in abnormal growth estimates (Additional file 1: Figures S.6 and S.7). Therefore, these two sets of σ_2 especially, were set to levels that did not produce extreme synthetic observations. A natural constraint on the standard deviation of the perturbations appeared in that too large a noise level can lead to NLS results with poor fit statistics with models that depart from the population curve not only in degree within growth curve shape, but also in overall shape as has been demonstrated

at smaller sample sizes, especially for the σ_2 level. The results point out that if the CR model is fitted to growth data with moderately large degree of variation, aberrant distributional shapes are possible in combined parameter estimation at moderately small sample sizes. This phenomenon may also occur if a growth equation that is inappropriate for the stand diameters under consideration is used in the model. Even though mortality can have a large effect on overall distribution shape as noted above, one can not expect ML to correct for an inappropriate growth model, and if it does in some instances, then the entire model estimation results are suspect, including both growth and mortality.

Limiting Forms

McDonald (1984) presents a hierarchy of different well-known distributions that can be derived by simple mathematical arguments from the full GB₁ distribution. Likewise, the CR-GB₁ also has similar distributions as subordinate forms. Two examples that will be discussed are the two-parameter Weibull and the negative exponential distributions. These particular distributions are found using a limiting argument that is outlined in the appendix by McDonald (1984) (details of the derivation are available from the authors on request). The first part of the derivation entails the substitution $b = \beta(p + q)^{\frac{1}{a}}$, which implies that $\beta = \frac{b}{(p+q)^{\frac{1}{a}}}$. Following substitution, the CR-GB₁ is written in terms of β rather than b . This density is simplified by letting $q \rightarrow \infty$ after employing Sterling's approximation and other simplifications. This limit in q is not applied to the terms in β , however. Thus, all of the parameters in the original CR-GB₁ are retained in the limiting forms. The final form is again a subset of the generalized gamma distribution (GGD) with $p = 1$. However, the GGD with $p = 1$ is simply the two-parameter Weibull distribution; *viz.*,

$$f(d; a, \beta) = \frac{a d^{a-1} e^{-\left(\frac{d}{\beta}\right)^a}}{\beta^a} \quad a, \beta > 0 \quad (18)$$

with shape parameter, a , and scale parameter β . The negative exponential distribution follows immediately from (18) by letting $a = 1$, which forces the condition that $m = 0$ in the CR growth equation since $a = 1 - m$, implying linearly decreasing growth. The negative exponential form is given as

$$f(d; \beta) = \frac{1}{\beta} e^{-\frac{d}{\beta}} \quad \beta > 0 \quad (19)$$

with rate parameter $\lambda = \frac{1}{\beta}$.

At first glance, the above results seem like a very simple way to connect the underlying vital rates with the oft-used Weibull and negative exponential distributions

in forestry. However, there is a problem with this interpretation for two reasons. First, under the CR-GB₁, the parameter $b \equiv d_{\infty}$ by definition; this substitution means that there is now an added constraint on the underlying parameters such that $b = \beta(p + q)^{\frac{1}{a}}$ must equal (4). The original reparameterization of b by McDonald (1984) was simply an expedient way to create a new parameterization of the GB₁ density that would yield itself to further simplification. There was no reason to be concerned with this originally under the full GB₁ because b was not already defined in terms of other model parameters as it is here. However, under the CR-GB₁ this leads to a potentially conflicting redefinition of parameter b .

The second reason why this is problematic is found in the fact that a limiting argument, $q \rightarrow \infty$, was used in the derivation procedure to arrive at the final distributional forms of the Weibull and negative exponential distributions in (18) and (19). Because q is comprised of three of the vital rate parameters, γ , m , and M , information on the relationship of these parameters to the original vital rate assumptions has been lost through the application of this asymptotic argument. Thus, the limiting argument for the development of the Weibull and negative exponential forms has decoupled the remnants of these parameters (still found in a and β) from the underlying growth and mortality equations that were inherent components in the original CRD model (5) through the loss of parameter information. The crux of the problem stems from the factorization of (5) in getting to (7), which creates a situation where the full set of parameters are no longer concentrated in the recognizable forms of the CR growth equation, $g(d)$, and the CRD survival function, $S'(d)$. Thus, application of the limiting argument made some of the pertinent terms in the dispersed model (7) vanish, and the full forms of these important functional components—growth and survival—of the full CR-GB₁ density were broken as a consequence.

The conclusion to this is that while (18) and (19) still retain all of the vital rate parameters (they can be multiplied by \tilde{N} to include recruitment and form numbers densities), there is no relation now between the driving vital rates and the final densities. The two-step estimation procedure no longer provides the relationship between estimated growth parameters and the CRD or CR-GB₁ because of the information loss. If the CNLS method were used to estimate the CR growth function, it would not aid in estimating (18) and (19) because the simplification has destroyed the fundamental link between the vital rates and the density. It is too much to ask that a ML scheme involving the only remaining unknown parameter, M , would yield meaningful results. The result of the limiting forms is shown in Fig. 7, where the same population parameter sets from Table 1 are used for each density. Note that there is no relationship between either the Weibull or negative

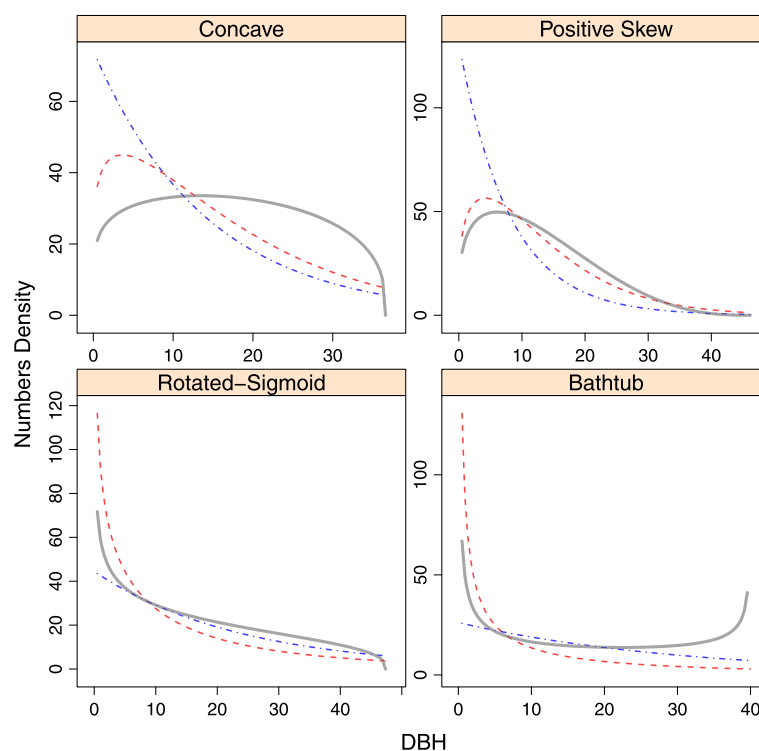


Fig. 7 The numbers densities for the limiting two-parameter Weibull (red, dashed) and negative exponential (blue, dot-dashed) in comparison to the true population distribution (gray, solid). (Refer to Table 1 for the parameter values used)

exponential densities to underlying true population CR-GB₁ density in any case; thus illustrating the complete decoupling of vital rates from limiting forms.

It must be emphasized that the fact that these limiting forms, while interesting, are not useful in the sense of a full demographic model, this does not mitigate against the actual CRD (CR-GB₁) itself. The forms contained within the limiting distributions are also largely found in family of shapes within the CRD models (5) and (7), though in a less obvious manner.

Conclusions

The so-called CRD and associated CR-GB₁ is based on the vital rate assumptions that diameter growth increments follow a CR growth curve, and mortality rates are constant over diameter, with constant recruitment. These assumptions have been investigated under the equilibrium solution to (1) for a range of distribution shapes that depend on two general forms of the underlying growth equation: convex and concave responses. Simulation experiments designed to assess the sample size requirements for the two-phase estimation scheme yielded variable requirements for both growth and mortality estimation; however, a conservative estimate of $n = 100$ to 200 observations in each shape- n_σ case seems necessary for convergence, depending on the amount of variation in the diameter increments. The variation

observed in the simulation results for the numbers densities was, of course, also linked to the individual Monte Carlo samples drawn, which obviously deviate from the target population distribution. However, the results of the KS rejection index indicates that more of the individual distributions tend to conform to those of the population as sample size approaches $n = 200$ observations. In addition, for each shape- n_σ combination, the mean of the distributions converges as expected to the population distribution as sample size increases. Thus, somewhere in the neighborhood of $n = 200$ observations would be a reasonable target for estimation of the CRD and its vital rate components.

We have referred to the CRD and more generally those arising from (1) as inherent distributions because of the mathematical link between the numbers density and the underlying vital rates, which shape the density and associated survival function. It might be argued that in one sense, the CR growth equation and the constant mortality, being empirical models, makes the CRD a form of assumed model at the size dynamics level in the model structure. Of course, this is quite common in growth models and is also found in the many studies cited. However, the vital rate models are not required to be empirically based: in general, process models or a more mechanistic approach could also be used. In such cases, this would lend a more complete connection between

the processes driving the vital rate models and the final numbers distribution form.

The CRD arises from simple assumptions on growth and mortality. As noted earlier, these may be generalized into more complex models which will produce far more flexible SSD forms, including multi-modal distributions (VanSickle 1977; Botsford et al. 1994). Though the more complex the vital rate models, the less likely there will be any link to known statistical distributions as was found here. A further complexity in the SSD model (1) can be accommodated in the sense of density dependence. The current model is considered density independent, since the number of trees in the stand (or some function of stocking like basal area) does not appear within the growth equation component of the model. However, if a more complex density dependent model were assumed, with feedback through any of the vital rates including birth or mortality, then model (1) becomes nonlinear, generally requiring more sophisticated solution methods (Murphy 1983) with concomitant loss of connection to known PDF forms such as the GB_1 described here.

Additional file

Additional file 1: Supplementary Material. The Chapman-Richards Distribution and its Relationship to the Generalized Beta. (PDF 4134 kb)

Funding

This work was partially supported by the USDA National Institute of Food and Agriculture, McIntire Stennis Project OKL0 3063 and the Division of Agricultural Sciences and Natural Resources at Oklahoma State University. Additional support was provided by the USDA Forest Service, Research Joint Venture 17-JV-11242306045, Old-Growth Forest Dynamics and Structure, to Mark Ducey.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

JHG designed the study, wrote the manuscript and software, derived the CRD and contributed to the derivations of the CR- GB_1 . TBL made the major contributions to the derivations of the CR- GB_1 , helped with the study design and commented on the manuscript. MJD contributed the derivations for the size-biased GB_1 , helped with the study design and commented on the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹USDA Forest Service, Northern Research Station, 271 Mast Road, Durham 03824, NH, USA. ²Professor Emeritus, Department of Natural Resource Ecology and Management, Oklahoma State University, Ag Hall Room 008C, Stillwater 74078, OK, USA. ³Department of Natural Resources and the Environment, University of New Hampshire, 114 James Hall, Durham 03824, NH, USA.

Received: 15 November 2018 Accepted: 17 April 2019

Published online: 21 May 2019

References

- Anderson RT (1937) The application of Fourier's series in forest mensuration. *J For* 35(3):293–299
- Bailey RL (1980) Individual tree growth derived from diameter distribution models. *For Sci* 26(4):626–632
- Bailey RL, Dell TR (1973) Quantifying diameter distributions with the Weibull function. *For Sci* 19:97–104
- Bailey RL, Burgan TB, Jokela EJ (1989) Fertilized midrotation-aged slash pine plantations—stand structure and yield prediction models. *Southern Journal of Applied Forestry* 13(2):76–80
- Baldwin Jr VC, Feduccia DP (1987) Loblolly pine growth and yield prediction for managed West Gulf plantations. Research Paper SO-236, USDA Forest Service. Southern Forest Experiment Station, New Orleans
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57(1):289–300
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29(4):1165–1188
- Bliss CI, Reinker KA (1964) A lognormal approach to diameter distributions in even-aged stands. *For Sci* 10(3):350–360
- Botsford LW, Smith BD, Quinn JF (1994) Bimodality in size distributions: The red sea urchin *Strongylocentrotus franciscanus* as an example. *Ecol Appl* 4(1):42–50
- Brooks JR, Borders BE, Bailey RL (1992) Predicting diameter distributions for site-prepared loblolly and slash pine plantations. *South J Appl For* 16(3):130–133
- Burk TE, Newberry JD (1984) A simple algorithm for moment-based recovery of Weibull distribution parameters. *For Sci* 30(2):329–332
- Burkhart HE (1971) Slash pine plantation yield estimates based on diameter distributions: An evaluation. *For Sci* 17(4):452–453
- Cajanus W (1914) Ueber die entwicklung gleichaltriger waldbestände. *Acta Forestalia Fennica* 3(1):1–142. <https://helda.helsinki.fi/handle/1975/8405>
- Cao QV (1997) A method to distribute mortality in diameter distribution models. *For Sci* 43(3):435–442
- Cao QV (2004) Predicting parameters of a weibull function for modeling diameter distribution. *For Sci* 50(5):682–685
- Clutter JL, Bennett FA (1965) Diameter distributions in old-field slash pine plantations. Report 13, Georgia Forest Research Council
- Conover WJ (1980) Practical nonparametric statistics, 2nd edn. Wiley, New York
- de Roos AM (1997) A gentle introduction to models of physiologically structured populations. In: Tuljapourkar S, Caswell H (eds). *Structured-population models in marine, terrestrial and freshwater systems, Population and community ecology series*, vol 18, chap 5. Chapman & Hall. pp 119–204
- Ducey MJ, Gove JH (2015) Size-biased distributions in the generalized beta distribution family, with applications to forestry. *Forestry* 88:143–151
- Efron B, Hinkley BV (1978) Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* 65(3):457–487
- Gove JH (1995) Bimodality of the combined removal and signs-of-activities estimator for sampling closed animal populations. *Environ Ecol Stat* 3:65–80
- Gove JH (2004) Structural stocking guides: a new look at an old friend. *Can J For Res* 34:1044–1056
- Gove JH (2015) The French connection: de Liocourt meets Pascal. unpublished manuscript (available upon request)
- Gove JH (2017) A demographic study of the exponential distribution applied to uneven-aged forests. *Forestry* 90:18–31
- Gove JH, Patil GP (1998) Modeling the basal area-size distribution of forest stands: a compatible approach. *For Sci* 44(2):285–297
- Gove JH, Ducey MJ, Leak WB, Zhang L (2008) Rotated sigmoid structures in managed uneven-aged northern hardwood stands: a look at the Burr type III distribution. *Forestry* 81(2):161–176
- Hafley WL, Schreuder HT (1977) Statistical distributions for fitting diameter and height data in even-aged stands. *Can J For Res* 7(3):481–487
- Hyink DM (1980) Diameter distribution approaches to growth and yield modelling. In: Brown KM, Clarke FR (eds). *Forecasting forest stand dynamics. School of Forestry, Lakehead University, Thunder Bay*. pp 138–163

- Hyink DM, Moser JW (1983) A generalized framework for projecting forest yield and stand structure using diameter distributions. *For Sci* 29(1):85–95
- Kerr G (2014) The management of sliver fir forests: de Liocourt (1898) revisited. *Forestry* 87:29–38
- Knowe SA (1992) Basal area and diameter distribution models for loblolly pine plantations with hardwood competition in the Piedmont and upper coastal plain. *South J Appl For* 16(2):93–98
- Leak WB (1964) The J-shaped probability distribution. *For Sci* 11(4):405–409
- Lindsay SR, Wood GR, Woollons RC (1996) Modelling the diameter distribution of forest stands using the Burr distribution. *J Appl Stat* 23(6):609–619
- de Liocourt F (1898) De l'aménagement des sapinières. Tech. rep., Bulletin trimestriel, Société forestière de Franche-Comté et Belfort. (English translation: <http://oak.snr.missouri.edu/silviculture/online/deLiocourt.pdf>)
- Liu C, Zhang L, Davis CJ, Solomon DS, Gove JH (2001) A finite mixture model for characterizing the diameter distributions of mixed-species forest stands. *For Sci* 48(4):653–661
- Lynch TB, Moser JW (1986) A growth model for mixed species stands. *For Sci* 32(3):697–706
- Matney TG, Schultz EB (2008) Deriving tree diameter growth and probability of survival equations from successive diameter distributions. *For Sci* 54(1):31–35
- Matney TG, Sullivan AD (1982) Compatible stand and stock tables for thinned and unthinned loblolly pine stands. *For Sci* 28(1):161–171
- McDonald JB (1984) Some generalized functions for the size distribution of income. *Econometrica* 52(3):647–664
- McGee CE, Della-Bianca L (1967) Diameter distributions in natural yellow-poplar stands. Research Paper SE-25, U. S. Forest Service
- Mehtätalo L (2005) Localizing a predicted diameter distribution using sample information. *For Sci* 51(4):292–303
- Mehtätalo L, Comas C, Pukkala T, Palahí M (2011) Combining a predicted diameter distribution with an estimate based on a small sample of diameters. *Can J For Res* 41:750–762
- Meyer HA (1952) Structure, growth, and drain in balanced uneven-aged forests. *J For* 50:85–92
- Meyer WH (1930) Diameter distribution series in evenaged forest stands. Bulletin 28, Yale University: School of Forestry, New Haven
- Muller-Landau H, Condit RS, Harms KE, Marks CO, et al. (2006) Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecol Lett* 9:589–602
- Murphy LF (1983) A nonlinear growth mechanism in size structured population dynamics. *J Theor Biol* 104:493–506
- Murphy PA, Farrar RM (1988) A framework for stand structure projection of uneven-aged loblolly-shortleaf pine stands. *For Sci* 34(2):321–332
- Pienaar LV, Turnbull KJ (1973) The Chapman-Richards generalization of Von Bertalanffy's growth model for basal area growth and yield in even-aged stands. *For Sci* 19:2–22
- Poudel KP, Cao QV (2013) Evaluation of methods to predict weibull parameters for characterizing diameter distributions. *For Sci* 59(2):243–252
- Rennolls K, Wang M (2005) A new parameterization of Johnson's S_B distribution with application to fitting forest tree diameter data. *Can J For Res* 35:575–579
- Rennolls K, Geary DN, Rollison TJD (1985) Characterizing diameter distributions by the use of the Weibull distribution. *Forestry* 58(1):57–66
- Robinson A (2004) Preserving correlation while modelling diameter distributions. *Can J For Res* 34:221–232
- Shifley SR, Brand GJ (1984) Chapman-Richards growth function constrained for maximum tree size. *For Sci* 30(4):1066–1070
- Siiipilehto J, Sarkkola S, Mehtätalo L (2007) Comparing regression estimation techniques when predicting diameter distributions of scots pine on drained peatlands. *Silva Fenn* 41(2):333–349
- Stuart A, Ord JK (1987) Kendall's Advanced Theory of Statistics, Volume I: Distribution Theory, 5th edn. Oxford University Press, New York
- VanSickle J (1977) Analysis of a distributed-parameter population model based on physiological age. *J Theor Biol* 64:571–586
- Wasserstein RL, Lazar NA (2016) The ASA's statement on p-values: Context, process, and purpose. *Am Stat* 70(2):129–133
- Zavala MA, Angulo O, de la Parra RB, López-Marcos JC (2007) An analytical model of stand dynamics as a function of tree growth, mortality and recruitment: The shade tolerance-stand structure hypothesis. *J Theor Biol* 244:440–450
- Zhang L, Gove JH, Liu C, Leak WB (2001) A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid, uneven-aged stands. *Can J For Res* 31(9):1654–1659
- Zhang L, Packard KC, Liu C (2003) A comparison of estimation methods for fitting weibull and johnson's s_B distributions to mixed spruce–fir stands in northeastern North America. *Can J For Res* 33:1340–1347

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)