

The Richit–Richards family of distributions and its use in forestry

Mingliang Wang, N. I. Ramesh, and Keith Rennolls

Abstract: Johnson's S_B and the logit–logistic are four-parameter distribution models that may be obtained from the standard normal and logistic distributions by a four-parameter transformation. For relatively small data sets, such as diameter at breast height measurements obtained from typical sample plots, distribution models with four or less parameters have been found to be empirically adequate. However, in situations in which the distributions are complex, for example in mixed stands or when the stand has been thinned or when working with aggregated data, then distribution models with more shape parameters may prove to be necessary. By replacing the symmetric standard logistic distribution of the logit–logistic with a one-parameter “standard Richards” distribution and transforming by a five-parameter Richards function, we obtain a new six-parameter distribution model, the “Richit–Richards”. The Richit–Richards includes the “logit–Richards”, the “Richit–logistic”, and the logit–logistic as submodels. Maximum likelihood estimation is used to fit the model, and some problems in the maximum likelihood estimation of bounding parameters are discussed. An empirical case study of the Richit–Richards and its submodels is conducted on pooled diameter at breast height data from 107 sample plots of Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.). It is found that the new models provide significantly better fits than the four-parameter logit–logistic for large data sets.

Résumé : Les équations Johnson S_B et logit–logistique sont des modèles de distribution à quatre paramètres qui peuvent être développés à partir des distributions standard normale et logistique par une transformation à quatre paramètres. Pour les jeux de données relativement de petite taille, tels que les mesures de diamètre à hauteur de poitrine obtenues à partir des placettes échantillons typiques, les modèles de distribution à quatre paramètres ou moins se sont avérés empiriquement adéquats. Cependant, dans les situations où les distributions sont complexes, dans les peuplements mélangés par exemple, ou lorsque le peuplement a été éclairci ou encore lorsqu'on veut travailler avec des données agrégées, des modèles de distribution avec plus de paramètres peuvent s'avérer nécessaires. En remplaçant la distribution logistique symétrique standard du modèle logit–logistique par une distribution de « Richards standard » à un paramètre et en faisant la transformation avec la fonction de Richard à cinq paramètres, nous obtenons un nouveau modèle de distribution à six paramètres appelé « Richit–Richards ». Ce nouveau modèle inclut les sous-modèles « logit–Richards », « Richit–logistique » et logit–logistique. L'estimation par maximum de vraisemblance est utilisée pour ajuster le modèle et certains des problèmes associés à l'estimation par maximum de vraisemblance des paramètres de démarcation sont discutés. Une étude de cas du modèle Richit–Richards et de ses sous-modèles est réalisée avec des données agrégées de diamètre à hauteur de poitrine provenant de 107 placettes échantillons de sapin de Chine (*Cunninghamia lanceolata* (Lamb.) Hook.). Le nouveau modèle permet d'obtenir, avec ses sous-modèles, un ajustement significativement meilleur que le modèle logit–logistique à quatre paramètres pour les jeux de données de grande taille.

[Traduit par la Rédaction]

Introduction

Diameter distribution models are a central tool in characterizing the size-class distribution of a forest stand and are

used to predict volume production and the range of products that might be expected from a stand.

A wide range of probability density functions, and fitting methods, have been considered and compared for use in forestry to model tree diameter distributions. These distribution models include lognormal (Bliss and Reinker 1964), gamma (Nelson 1964), Weibull (Bailey and Dell 1973; Rennolls et al. 1985), beta (Zohrer 1972; Li et al. 2002), S_B (Hafley and Schreuder 1977; Kamziah et al. 1999; Zhang et al. 2003; Rennolls and Wang 2005), and the logit–logistic, generalized Weibull, and Burr XII (Wang and Rennolls 2005). Wang and Rennolls (2005) found that the logit–logistic gave the best performance in their case study.

The utility of a distribution model depends on its flexibility of distribution shape representation, tractability, and parsimony. The “principle of parsimony” (otherwise known as Occam's razor or the Pareto principle) states that if there are two alternative models that describe a data set equally well,

Received 28 November 2006. Accepted 25 January 2007.
Published on the NRC Research Press Web site at cjfr.nrc.ca on 2 November 2007.

M. Wang¹ School of Computing and Mathematical Sciences, University of Greenwich, London, SE10 9LS, UK; Chinese Academy of Forestry, Beijing, PR China; Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30601, USA.

N.I. Ramesh and K. Rennolls² School of Computing and Mathematical Sciences, University of Greenwich, Old Royal Naval College, London, SE10 9LS, UK.

¹Corresponding author (e-mail: mwang@forestry.uga.edu).

²Corresponding author (e-mail: k.rennolls@gre.ac.uk).

then the simplest model, usually that with fewer parameters, is to be preferred (Baker 2004; Sarkar and Pfeifer 2005).

In studies of the distribution of diameter at breast height (DBH), it has often been found that a simple three-parameter Weibull distribution is an adequate description of the data from sample plot studies. The literature of distribution modelling of typical sample plot data suggests that models with more than four parameters would be unnecessarily complex. However, for atypical sample plot DBH distribution data, for example from mixtures or when the stand is multistory or when the stand has been subjected to novel thinning regimes, it is possible that a model with more than four parameters will be better than the four-parameter models currently available. If the distribution data set is very large, as might be the case if data are collected from a whole stand, or from a region, possibly by pooling data sets, then it is also possible that more highly parameterized distribution models will be required to adequately describe the data.

We develop a new six-parameter family of distribution models, the "Richit-Richards" (RR), a generalization of the four-parameter logit-logistic (LL) of Wang and Rennolls (2005). The RR has two five-parameter submodels, the "logit-Richards" (LR) and the "Richit-logistic" (RL).

We present a case study that uses large pooled DBH distribution data sets to demonstrate that the RR (and its five-parameter submodels) provide a better description than the four-parameter LL distribution.

The RR distribution

Before defining the RR distribution, we first define the S_B and LL distributions that lead naturally to it.

The S_B distribution

Johnson (1949) defined the S_B distribution as that bounded distribution that when scaled to the range (0, 1), transformed by the logit function and scaled by a linear transformation, produces a standard normal distribution. Hafley and Schreuder (1977) and Schreuder and Hafley (1977) first demonstrated its utility in forest distribution modelling. Rennolls and Wang (2005) viewed the transformation process in the opposite direction and consequently suggested what they regarded as a more natural parameterization of S_B .

The LL distribution

Tadikamalla and Johnson (1982), following the approach of Johnson (1949), used a logit transformation of a two-parameter logistic distribution (a symmetric distribution similar to the normal but with fatter tails (Berkson 1951)) to give what Wang and Rennolls (2005) have called the LL distribution. Wang and Rennolls (2005), following the approach of Rennolls and Wang (2005), based their definition and parameterization of the LL distribution on a transformation of a standard logistic (SL) distribution (with CDF $1/(1 + e^{-z})$), first by a linear scaling, then by a (symmetric) SL function, and finally by a linear scaling to the range (ξ , $\xi + \lambda$). The overall transformation from $z \sim \text{SL}$ to $x \sim \text{LL}$ is

$$[1] \quad x = \xi + \frac{\lambda}{1 + e^{-(\mu + \sigma z)}}$$

where μ and σ are the usual parameters of a linear scaling transformation of z and ξ and λ are, respectively, the lower bound and range parameters of x . The LL cumulative distribution function (CDF) and probability density function (pdf) are given in Wang and Rennolls (2005) and may be easily obtained from those of the RR given below in this paper.

The RR distribution and its submodels

The symmetry properties of the logistic and logit components of the LL distribution limit its representational flexibility. Hence, in looking for a flexible generalization of the LL, two generalizations are natural: first, to replace the symmetric SL distribution by an asymmetric one-parameter "standard Richards" (SR) distribution (defined below), and second, to replace the (symmetric) SL transformation function by a one-parameter (asymmetric) SR function. This is equivalent to applying a power transformation after the SL transformation in the construction of the LL distribution. We call the inverse of the SR function the "Richit", in analogy to the logit. Hence, the resulting six-parameter distribution model is obtained by applying a linear scaling to an SR distribution, which is then transformed by an SR function, which is finally scaled to an arbitrary range. We call the resulting distribution the Richit-Richards (RR) distribution.

The SR distribution

The SR distribution has CDF

$$[2] \quad F_{Z_{SR}}(z) = \frac{1}{(1 + e^{-z})^k}$$

where k is a shape parameter. This is a simple form of a Richards function (Richards 1959; Nelder 1962). The Richards distribution (the SR with z replaced by $\mu + \sigma z$) is also known as the Burr II distribution (Burr 1942) and has been used in forest diameter distribution modelling by Ishikawa (1991, 1996).

We note, to avoid confusion, that there is another form of (standardized) Richards function (i.e., $(1 - \exp(-z))^k$) that is usually known as the "Chapman-Richards" function in forestry (Pienaar and Turnbull 1973). The two forms of the Richards function are related (see Garcia 2005), but we do not use the Chapman-Richards function in defining the RR distribution.

The standard Richards function

We call the functional form of the SR CDF in eq. 2 the SR function and its inverse function the (standard) Richit function. They are, respectively:

$$[3] \quad y = \frac{1}{(1 + e^{-x})^{1/c}} : x = \ln\left(\frac{y^c}{1 - y^c}\right)$$

where c is a shape parameter. Note that the SR function is to be used to transform distributions and is distinct from the SR CDF of eq. 2. Parameters k of eq. 2 and $(1/c)$ of eq. 3 are distinct parameters.

The Richit–Richards CDF and pdf

If z is distributed as SR with single parameter k , as specified by eq. 2, then x , given by the five-parameter (Richards) transformation

$$[4] \quad x = \xi + \frac{\lambda}{(1 + e^{-(\mu + \sigma z)})^{1/c}}$$

has the RR(θ) distribution with parameter vector $\theta = (k, \mu, \sigma, c, \xi, \lambda)$. The parameters are defined in eqs. 1, 2, and 3. Transformation 4 was used by Mead (1965) to generalize Johnson’s S_B distribution to a generalized logit–normal distribution (that is, the Richit–normal: note that S_B is logit–normal). Note also (following Garcia 2005) the close association between growth curves models and the distributions they produce when used as transformations of basic distribution models such as the standard of the normal, logistic, Richards, and uniform distributions.

Since eq. 4 is a monotonic transformation function applied to the SR distribution of z to produce the RR distribution of x , we have

$$[5] \quad F_{x_{RR}}(x) = F_{z_{SR}}(z) = \frac{1}{(1 + e^{-z})^k} = \frac{1}{\left\{ 1 + e^{\frac{\mu}{\sigma}} \left[\left(\frac{x - \xi}{\lambda} \right)^{-c} - 1 \right]^{\frac{1}{\sigma}} \right\}^k}$$

where z as a function of x is found by inverting eq. 4. This explicit and tractable form of the RR CDF allows the RR distribution to be directly simulated by applying the inverse of the CDF in eq. 5 to $u \sim U(0, 1)$ (or $z = -\ln(u^{-1/k}) - 1$) from eq. 2, $u \sim U(0, 1)$, and $x \sim RR(\theta)$ using eq. 4).

The RR pdf is obtained by differentiating the RR CDF in eq. 5 to give

$$[6] \quad f_{x_{RR}}(x) = \frac{\left(\frac{kc}{\sigma\lambda}\right) e^{\frac{\mu}{\sigma}} \left[\left(\frac{x - \xi}{\lambda} \right)^{-c} - 1 \right]^{\frac{1}{\sigma} - 1} \left(\frac{x - \xi}{\lambda} \right)^{-(c+1)}}{\left\{ 1 + e^{\frac{\mu}{\sigma}} \left[\left(\frac{x - \xi}{\lambda} \right)^{-c} - 1 \right]^{\frac{1}{\sigma}} \right\}^{k+1}}$$

$$[8] \quad L = n \left(\ln k + \ln c - \ln \sigma + c \ln \lambda + \frac{\mu}{\sigma} \right) - \left(\frac{c}{\sigma} + 1 \right) \sum_{i=1}^n \ln(x_i - \xi) + \left(\frac{1}{\sigma} - 1 \right) \sum_{i=1}^n \ln[\lambda^c - (x_i - \xi)^c] - (k + 1) \sum_{i=1}^n \ln \left\{ 1 + e^{\frac{\mu}{\sigma}} \left[\left(\frac{x_i - \xi}{\lambda} \right)^{-c} - 1 \right]^{\frac{1}{\sigma}} \right\}$$

For fitting purposes, a simple reparameterization was used that replaced the range parameter λ by the upper bound parameter $v = \xi + \lambda$. This made the boundary constraints simpler to specify, as $\hat{\xi} \leq x_{(1)}$ (the minimum sample value) and $\hat{v} \geq x_{(n)}$ (the maximum sample value). The invariance property of MLE means that $\hat{\lambda}_{ML} = \hat{v}_{ML} - \hat{\xi}_{ML}$. We have left the parameterization in the above treatment in terms of λ , since this is the conventional parameterization following Johnson (1949). MLE has been achieved by minimizing $-L$ with respect to the parameter vector θ using the function *optim*

The RL and LR RR submodels

The five-parameter RL distribution is obtained from setting $k = 1$ in the RR distribution, corresponding to the use of the SL instead of SR as the source distribution to be transformed by eq. 4. The RL is an analogue of the Richit–normal introduced by Mead (1965) as a generalization of S_B . The five-parameter LR distribution is obtained from setting $c = 1$ in the RR distribution. LR results from the transformation of the SR distribution by eq. 1. The CDFs and pdfs of the RL, LR, and LL are obtained from eqs. 5 and 6 by setting $k = 1$, $c = 1$, and $k = c = 1$, respectively.

Inferential methods

Parameter estimation method

There has been a wide range of criteria used for fitting empirical distribution models and for evaluating the goodness-of-fit, including error sum of squares, likelihood, performance in volume prediction usage, etc. (Zhou and McTague 1996). In some cases, performances over a range of criteria-of-fit have been considered (Wang 2005, pp. 69–95).

We use maximum likelihood estimation (MLE) and adopt the corresponding measure of goodness-of-fit, $-2 \log$ -likelihood, to compare via likelihood ratio tests the relative performances of the models in the RR \rightarrow (RL/LR) \rightarrow LL hierarchy of models. We note that MLE of bounded distribution models is known to be problematic (Davison 2003, p. 146).

If we assume that the sample data $\underline{x} \equiv \{x_i\}_{i=1, \dots, n}$ are independently and identically distributed of size n from the RR(θ) distribution, then the likelihood of the observed data is

$$[7] \quad l(\underline{x}|\theta) = \prod_{i=1}^n f_{RR}(x_i|\theta)$$

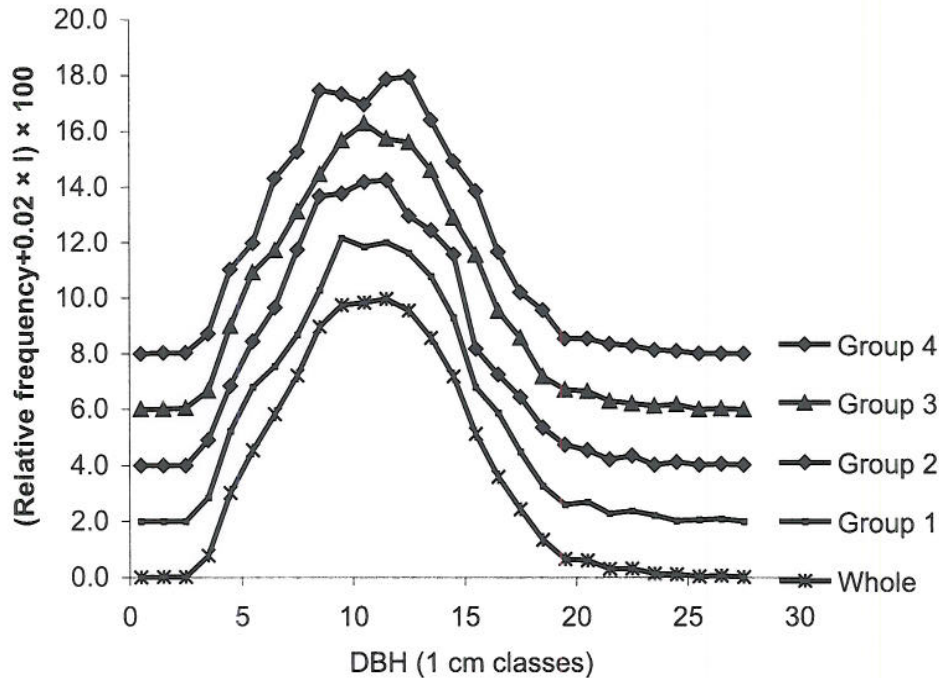
where $\theta = (k, \mu, \sigma, c, \xi, \lambda)$. In the case study, x is DBH. The log-likelihood function L is

of the statistical package R (R Development Core Team 2006).

Estimation of the boundary parameters ξ and v

Previous studies (e.g., Hafley and Schreuder 1977; Li et al. 2002; Zhang et al. 2003) have determined the bounding parameter by heuristic means, for example, either by setting the lower boundary parameter to zero or by setting $\hat{\xi} = x_{(1)}$ (or a little less) before estimating the remaining model pa-

Fig. 1. Empirical relative frequency polygons for the four aggregated data sets and the whole data set. The plotted points are at the mid-points of 1 cm intervals and the relative frequency for subset I ($RF(I)$) has been scaled to $(RF(I) + 0.02 \times I) \times 100$. For the whole data set, $I = 0$.



rameters. In practical terms, such procedures would seem to be justified. Visual examination of the empirical DBH distributions in Fig. 1 suggests that the lower bound parameter ξ might reasonably be set to a value of about 3 cm. Mead (1965) used a two-stage estimation procedure, estimating the boundary parameters first and subsequently the remaining parameters. There seems to be little other published guidance on formal methods of estimating the boundary parameters of a distribution.

We have estimated the bounding parameters ξ and ν using MLE. The advantages of doing so are that the same MLE procedure is used for all the model parameters, and hence in principle, an asymptotic variance-covariance matrix of the whole parameter vector can be obtained and hence asymptotic standard errors and correlations of all parameter estimates.

However, there is a technical difficulty that arises in using MLE for the boundary parameters ξ and ν , and this is probably the reason why previous studies have estimated these parameters heuristically. If there is a boundary constraint on a parameter, such as $\xi \leq x_{(1)}$ (the minimum observed diameter), then if the MLE of ξ approaches its bounding value $x_{(1)}$, we are in what is called the “nonstandard conditions” of MLE (Moran 1971; Self and Liang 1987). The likelihood equation $\frac{\partial(-L)}{\partial\xi} = 0$ is not satisfied in general under nonstandard conditions. For example, for an independently and identically distributed sample from $U(\xi, \nu)$, the MLEs are the biased estimates: $\hat{\xi} = x_{(1)}$ and $\hat{\nu} = x_{(n)}$. Under nonstandard conditions for more general distribution models, such as the RR family, the likelihood as $\xi \rightarrow x_{(1)}$ can have vari-

ous nonstandard properties depending on the nature of the shape of the lower tail of the model adopted. The shape of $-L$ is not explicitly convex-down if $\hat{\xi} \rightarrow x_{(1)}$, since it terminates at the boundary in parameter space. We discuss in the case study section some of the problems encountered in using MLE of the bounding parameters ξ and ν .

Model comparisons

The comparisons of the fit of submodels in the RR family are achieved using likelihood ratio tests.

Correlation matrix of parameter estimates

After ML parameter estimation, the asymptotic information matrix, the variance-covariance matrix, and the correlation matrix (of the estimated parameter vector) can be computed. The (i, j) th element of Fisher’s information matrix (of the parameter estimates) is $I(\hat{\theta}) = E\left(-\frac{\partial L(\theta)}{\partial\theta_i\partial\theta_j}\right)_{\theta=\hat{\theta}}$. The asymptotic variance-covariance matrix of parameter estimates, $V(\hat{\theta})$, is the inverse of the information matrix (Cox and Hinkley 1974). The correlation matrix (of the estimated parameter vector) is obtained from the variance-covariance matrix in the usual way.

In this study, to estimate V , the following approximation was used:

$$[9] \quad I(\hat{\theta}) \cong \text{Observed}\left(-\frac{\partial L(\theta)}{\partial\theta_i\partial\theta_j}\right)_{\theta=\hat{\theta}}$$

The partial derivatives may be evaluated symbolically using the S-Plus function *deriv*, and we have used such an approach in S-Plus implementations. Such symbolic evaluation

Table 1. Summary statistics for the four aggregated data sets and the whole data set.

Data set	Count	Mean	SD	Skew	Kurtosis	Minimum	Maximum
1	3 172	11.23	3.75	0.38	0.18	3.5	27
2	3 106	11.11	3.67	0.42	0.23	3.5	27.5
3	3 210	11.20	3.71	0.34	-0.01	2.2	26.9
4	3 228	11.23	3.74	0.38	0.11	1.9	27.5
Whole	12 716	11.19	3.72	0.38	0.12	1.9	27.5

is only possible because of the explicit form of L for the RR distribution, as given in eq. 8. However, we have also used the usual numerical method of evaluating the Hessian in eq. 9 when using R, and this approach may be used when the analytic approach is not possible. The results obtained are essentially the same, subject to small numerical differences resulting from the differing numerical algorithms and stopping criteria used.

Case study with Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.) sample plot DBH distribution data

Sample plot data

The diameter data of 107 plots for Chinese fir plantations were provided by the Chinese Academy of Forestry. These plots were located at Kaihua forestry farm, Zhejiang Province, southeastern China. The plot size ranges from 400 to 600 m², age varies from 10 to 29 years, and density ranges from 1000 to 4500 stems/ha. The number of trees per plot ranges from 63 to 239 with a mean of about 119. Further details about the data may be found in Wang and Rennolls (2005).

Results for sample plot data

The RR family of models has been fitted to all of the sample plot data sets individually. As expected, the six-parameter RR and five-parameter LR and RL models do not demonstrate significantly better fits than the four-parameter LL, since the range and complexity of the distributional forms exhibited by the sample plot data sets are rather limited and the sample sizes are not sufficiently high to warrant the use of the models with more parameters than the four of the LL distribution.

Aggregated case study data

The 107 sample plots were therefore aggregated into four sets in a random manner so that each set contained about the same number of DBH measurements. These pooled data sets were of sizes 3172, 3106, 3210, and 3228, large compared with typical sample plot sample sizes but possibly around the size expected for an intensive study or a regional inventory. The DBH data from all the plots resulted in a sample size of 12 716.

The basic statistics for the four aggregated data sets and all data are shown in Table 1. The means and standard deviations of all data sets are similar, as expected from the method of construction. The skew of the data sets varies between 0.34 and 0.42, corresponding to a fairly well-defined lower limit to the distributions and fairly long upper distribution tails. The kurtosis statistics vary from -0.01 to 0.23.

Even though our aggregated data sets are the result of pooling many sample plot data sets, the central limit theorem does not apply to the resulting distributions: Normality (with skew and kurtosis both zero) is not to be expected.

The empirical relative frequency polygons are shown in Fig. 1. The relative frequency (RF) axis for data set I (RF(I), say) has been scaled to $(\text{RF}(I) + 0.02 \times I) \times 100$ for clarity of presentation. Data sets 1, 3, and 4 seem to display some bimodality, but this has been averaged out in the whole data set.

Aggregated data results

Model fits

The estimated parameters (and standard errors) for the RR family of models are presented in Table 2. The estimation procedure converged in all cases, although a selection of parameter starting values had to be tried to avoid convergence to local minima of $-L$. Only for the RL model, for subset 2, were standard errors unobtainable.

Figure 2 shows the fitted models on the whole data set using 2 cm DBH classes that are conventional for management purposes in China. It is apparent that the RR model provides a qualitatively better fit to the empirical DBH distribution than its submodels. Figure 3 shows the empirical DBH distribution for subsample 4 and 99 simulations from the fitted RR model. This demonstrates that the fitted RR model has a significant lack of fit to this bimodal empirical distribution, rather as expected, even though the RR family of models is able to take a bimodal form for appropriate parameter settings. It is possible that the beta distribution, with its polynomial pdf form, might provide a better fit to multimodal distributions than the RR family. Observation of a multimodal tree size distribution can indicate that the population is a mixture, in which case, a mixture distribution model, for example, a Weibull mixture (Zhang et al. 2001), would be most appropriate.

Boundary parameter estimates

To illustrate the issues relating to nonstandard conditions of MLE, consider the RR model fitted to data set 2. Figure 4 shows plots of $-L$ as functions of the individual parameters (the other parameters being set to their MLE values) for this case. The MLE of $(k, \mu, \sigma, c, \xi, \nu = \xi + \lambda)$ was (0.3477, -1.4298, 0.3499, 1.6735, 3.4129, 31.4579) in this data set, with the minimal value of $-L$ attained being 8361.4. It may be seen from Fig. 4 that the parameters (k, μ, σ, c, ν) have values at which $-L$ is both minimal and convex-down. However, for the lower boundary parameter ξ , $-L$ approaches its minimal value as the lower boundary parameter approaches the boundary of its feasible region, in this case the minimal observed data value of 3.5 cm.

Table 2. Summary of estimated parameters for the RR family of distribution models with standard errors.

		ξ	ν	μ	σ	c	k
Subsample 1	RR	3.46	32.39	-2.09	0.39	2.34	0.27
	Standard error	0.04	4.21	1.02	0.13	1.12	0.05
	RL	3.50	27.55	0.74	0.33	0.33	
	Standard error	0.02	0.34	0.27	0.02	0.08	
	LR	3.48	38.00	-0.81	0.24		0.34
	Standard error	0.02	9.26	0.33	0.05		0.06
Subsample 2	RR	3.41	31.46	-1.43	0.35	1.67	0.35
	Standard error	0.07	4.18	1.53	0.17	1.56	0.12
	RL	3.30	28.44	1.06	0.29	0.25	
	Standard error	na	0.52	na	0.01	na	
	LR	3.45	33.08	-0.67	0.28		0.42
	Standard error	0.04	3.95	0.16	0.03		0.05
Subsample 3	RR	2.15	27.62	-2.32	0.52	3.09	0.38
	Standard error	0.05	0.91	0.67	0.10	0.93	0.05
	RL	2.15	27.78	0.30	0.32	0.52	
	Standard error	0.06	0.64	0.28	0.02	0.12	
	LR	2.15	28.26	-0.46	0.34		0.67
	Standard error	0.05	1.05	0.05	0.02		0.06
Subsample 4	RR	1.83	28.41	-2.19	0.50	2.80	0.43
	Standard error	0.06	1.03	0.78	0.11	1.04	0.08
	RL	1.83	28.82	0.11	0.31	0.59	
	Standard error	0.09	0.82	0.28	0.03	0.13	
	LR	1.83	29.03	-0.51	0.34		0.73
	Standard error	0.07	0.99	0.05	0.02		0.07
Whole	RR	1.88	28.18	-2.30	0.52	2.94	0.42
	Standard error	0.02	0.48	0.35	0.05	0.47	0.03
	RL	1.88	28.59	0.11	0.31	0.59	
	Standard error	0.02	0.41	0.14	0.01	0.07	
	LR	1.88	29.18	-0.52	0.33		0.71
	Standard error	0.02	0.63	0.03	0.01		0.03
	LL	1.74	27.81	-0.61	0.39		
	Standard error	0.10	0.16	0.01	0.005		

Note: na, not available.

Fig. 2. Fitted distribution models of the RR family for the whole data set. The ordinate indicates the number of sample trees in a 2 cm DBH class.

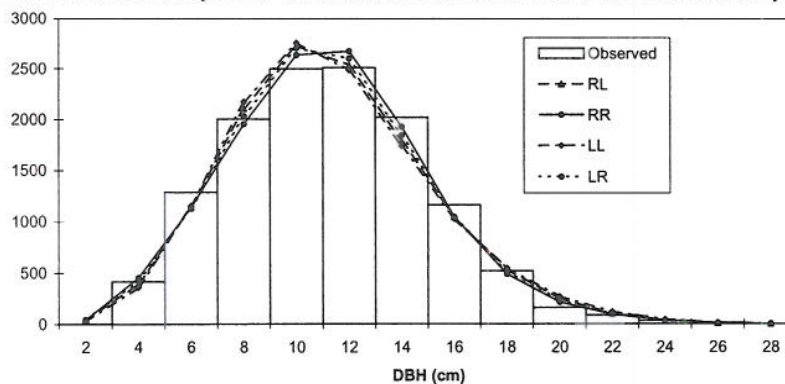
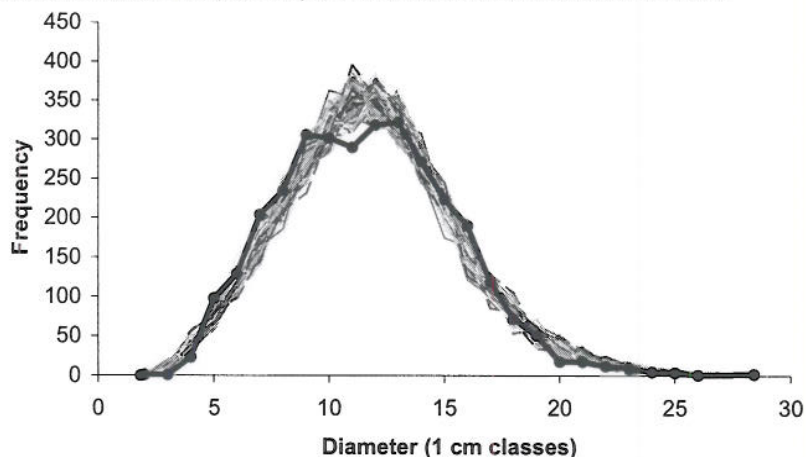


Fig. 3. Empirical DBH distribution for subset 4 (bold line) and 99 simulations from the fitted RR model.



These features of MLE under nonstandard conditions can lead to some software optimization packages indicating a “nonconvergence condition” even though the minimum of $-L$ has been achieved. Some software packages fail to calculate the variance–covariance matrix in this case, since the numerical calculation involves second derivatives of $-L$ about the estimated parameter value. For our five case study data sets, use of R software with numerical variance–covariance estimation did find the ML parameter estimates of the RR model satisfactorily, and variance–covariance estimates were obtained. The variance–covariance estimates obtained should be treated with caution, since they are biased estimates in these nonstandard conditions (Self and Liang 1987).

Correlation matrices

The estimated correlation matrix for the parameters of the RR model for the whole sample is shown in Table 3. The estimated correlations between $\hat{\mu}$ and the other two shape parameter estimates ($\hat{\sigma}$ and \hat{k}) and between $\hat{\sigma}$ and \hat{k} are very high, at -0.96 , -0.99 , and 0.97 , respectively. Two-dimensional sections of the $-L$ function about the MLE have essentially elliptical contours, and so are not presented. The correlation of -0.99 between $\hat{\mu}$ and \hat{k} would normally be taken to indicate that the later parameter is redundant in this case. However, the $-L$ of the fitted RR model in this case is 34 533.5, a significant improvement over the $-L$ of 34 551.7 for the LR submodel (which is equivalent to the RR model with c set to 1). See Table 5.

A “well-parameterized” model is often characterized as one in which the $-L$ manifold at the MLE fit to the data set concerned is expressed orthogonally in terms of the parameters (Cox and Reid 1987; Ratkowsky 1990; Ross 1990; Rennolls 1995b; Davison 2003). In such an “orthogonal” parameterization, the estimated correlations between MLE parameter estimates will be close to zero. To summarize our correlation results, and the lack of orthogonality of the RR parameterization in our five data sets, we take the value of 0.95 as indicating excessive correlation between parameter estimates.

For LL, no excessive correlations were found among the four parameter estimates of LL. However, for LR, high cor-

Table 3. Estimated correlation matrix for the parameter estimates of the RR model for the whole sample.

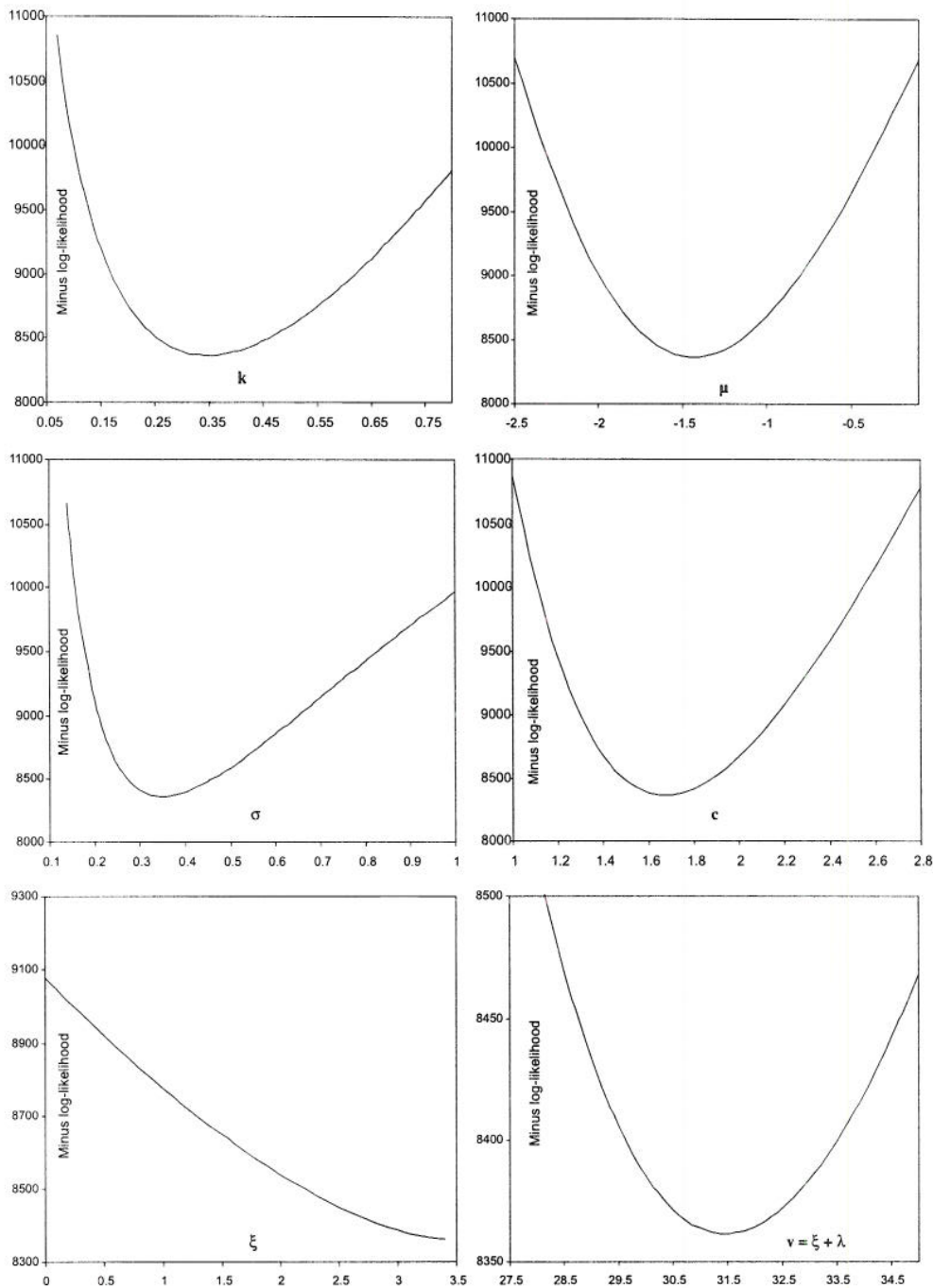
	ξ	ν	μ	σ	c	k
ξ	1.0000	0.0049	-0.0252	0.0179	0.0253	-0.0842
ν	0.0049	1.0000	0.3686	-0.5810	-0.4721	0.2070
μ	-0.0252	0.3686	1.0000	-0.9596	-0.9910	0.8725
σ	0.0179	-0.5810	-0.9596	1.0000	0.9736	-0.7571
c	0.0253	-0.4721	-0.9910	0.9736	1.0000	-0.8802
k	-0.0842	0.2070	0.8725	-0.7571	-0.8802	1.0000

relation was found between pairs $(\hat{\nu}, \hat{\mu})$, $(\hat{\nu}, \hat{\sigma})$, $(\hat{\nu}, \hat{k})$, $(\hat{\mu}, \hat{\sigma})$, and $(\hat{k}, \hat{\sigma})$ on the first subsample and between $(\hat{\nu}, \hat{\mu})$, $(\hat{\nu}, \hat{\sigma})$, and $(\hat{k}, \hat{\sigma})$ on the second subsample. The high correlation between $(\hat{\mu}, \hat{\sigma})$ and (or) $(\hat{k}, \hat{\sigma})$ is not unexpected, since they are all shape parameters. The high correlation between the parameter estimate $\hat{\nu}$ and the estimates of the other (shape) parameters $\hat{\mu}$, $\hat{\sigma}$, and \hat{k} for data sets 1 and 2 was not expected.

For RL, there was high correlation between pairs of $(\hat{c}, \hat{\mu})$ for all data sets and between $(\hat{c}, \hat{\sigma})$ for data sets 1, 2, and 3. For RR, high correlation between pairs of $(\hat{c}, \hat{\mu})$ and pairs of $(\hat{c}, \hat{\sigma})$ for all data sets, between $(\hat{\mu}, \hat{\sigma})$ for data sets 2, 3, and 4, and between $(\hat{k}, \hat{\mu})$ and (\hat{k}, \hat{c}) for the second subsample was found.

These results suggest that some reparameterization of the RR distribution model might be necessary. Rennolls and Wang (2005) suggested such a reparameterization for S_B , which seemed to improve the statistical properties (orthogonality) of the parameter estimates. However, this parameterization has already been used in our specification of the RR family of models. Rennolls (1995a) also used reparameterization of the Richards growth model to overcome convergence problems. Further discussion of such ill-parameterization and reparameterization issues may be found in Grosenbaugh (1965) and Rennolls (1995b). We are unable to suggest at the present time a suitable reparameterization of the RL and RR models with respect to parameter

Fig. 4. RR model fitted to subsample 2: profile (minus) likelihood functions plotted for the ($k, \mu, \sigma, c, \xi, \nu$) parameters.



c , which would remove the high correlations involving parameter estimates of c .

Overall, from the results on correlations among parameter estimates for this case study, it is suggested that introducing an additional (shape) parameter k (to give the SR distribution) is justified as a means of extending the LL model to the LR model. This conclusion is in concordance with the conclusion reached, in the next section, on the superiority of LR over RL in terms of a goodness-of-fit evaluation.

Goodness-of-fit evaluation

Table 4 lists the minimal attained value of $-L$ for each of the models for each of the data sets considered.

For nested models, the deviance ($= 2(L(\text{unconstrained}) - L(\text{constrained}))$) is distributed approximately as the χ^2 distribution with degrees of freedom determined by the number of parameters reduced from the unconstrained model to the constrained (McCullagh and Nelder 1989). Table 5 presents

Table 4. Minimal $-L$ from fitting LL, LR, RL, and RR distributions on the large sample data sets.

	LL	LR	RL	RR
Subsample 1	8 649.79	8 597.03	8 622.97	8 596.50
Subsample 2	8 397.73	8 361.61	8 375.43	8 361.40
Subsample 3	8 729.11	8 718.82	8 722.52	8 713.41
Subsample 4	8 799.09	8 793.46	8 795.21	8 789.99
Whole	34 580.95	34 551.74	34 563.43	34 533.51

Table 5. Likelihood ratio test statistics of compared distribution pairs.

	(LL, LR) (df = 1)	(LL, RL) (df = 1)	(LL, RR) (df = 2)	(LR, RR) (df = 1)	(RL, RR) (df = 1)	(RL, LR) (df = 0)
Subsample 1	105.52***	53.64***	106.58***	1.06 ns	52.94***	51.88
Subsample 2	72.24***	44.60***	72.66***	0.42 ns	28.06***	27.64
Subsample 3	20.58***	13.18***	31.40***	10.82***	18.22***	7.40
Subsample 4	11.26***	7.76**	18.20***	6.94**	10.44**	3.50
Whole	58.42***	35.04***	94.88***	36.46***	59.84***	23.38

Note: The asterisks indicate a nominal significance level from the χ^2 distribution for the improvement of fit by the more general model. **Highly significant ($p \leq 0.01$); ***very highly significant ($p \leq 0.001$); ns, not significant.

the estimated deviances between model pairs. The critical values of the χ^2 distribution with 1 df at the 0.05, 0.01, and 0.001 probability levels are 3.84, 6.63, and 10.83 and for 2 df are 5.99, 9.21, and 13.82, respectively. We also include the nonnested pair of (RL, LR) for which the likelihood test cannot be used. Note that the Akaike criterion of model fit (Akaike 1974) is given by $AIC = 2(p - L)$. For models with differing numbers of parameters, the best model is that with the smallest AIC. Hence, use of AIC to compare RL and LR is equivalent to the use of the difference in the $-L$ values for the two models.

We see from Table 5 the impact of large sample sizes on the significance of the improvement of fit provided by the five-parameter and six-parameter models over the four-parameter LL. The p values corresponding to the very high significance level are generally much less than the nominal 0.001 value. The differences in the $-L$ values that reach the nominal significance levels in Table 5 are so large that use of the AIC for model selection reaches the same conclusions as indicated by the nominal significance levels in Table 5.

The summary statistical conclusions, for this case study, from Table 5 are (i) LR is very significantly better than LL for all data sets, (ii) RL is significantly better than LL for all data sets, (iii) RR is very significantly better than LL for all data sets, (iv) RR is significantly better than LR for two of the four aggregated data sets and for the whole data set, and (v) LR is superior to RL for all of the four subsamples and for the whole data set.

Our conclusion from this empirical study is that if the empirical data set has a very large sample size, then the six-parameter RR model may prove to be statistically superior to models with a smaller number of parameters. From our case study data, the indication is that LR is superior to RL. That is, if we are to increase the degree of parameterization from the four-parameter LL by 1 (or equivalently decrease the degree of parameterization from the six-parameter RR by 1), then use of SR as the source distribution is to be preferred over the use of the Richit transformation. However, we can-

not assert from this empirical case study that LR will be superior to RL for other shapes of (diameter) distribution.

Discussion and conclusions

All empirical models are (strictly speaking) "wrong", but some are less wrong than others. The principle of parsimony is a reasonable approach for choosing the most suitable empirical model for characterizing DBH distributions. Significance tests are a reasonable way of assessing if two alternative models describe the data equally well. Using these principles, we find that distribution models of no more than four parameters are sufficient to describe the individual sample plot data of our case study. We have also found in our pooled-data large-sample case study that the five-parameter LR and the six-parameter RR are superior to the four-parameter LL.

For practical forest distribution modelling, the available models comprise a "toolkit" from which the most convenient and useful model can be selected and used in the modelling task at hand. The tractable RR family of models extends the range of distribution model tools. The R code for MLE of the RR family and other Excel-based facilities for using the RR family will be made available on the Forest Model Archive (www.forestmodelarchive.info) in due course.

However, we do feel that there is a limit to the number of parameters that can reasonably be used in an explicit parametric distribution model, and our experience in this study suggests that this is about five or six! Ideally, the parameterization of an explicit parametric model should be such that each of the parameters relates to a separate observable feature of the empirical data (Ross 1990), such as the minimal and maximal bounding parameters in a distribution model. In the large-sample case study of this paper, the high correlations between the estimates of the shape parameters of the RR family follow from the fact that they do not independently represent separate well-defined features of the empiri-

- bution for describing stand structure of tree heights and diameters. *Biometrics*, **33**: 471–478. doi:10.2307/2529361.
- Self, S.G., and Liang, K.-Y. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J. Am. Stat. Assoc.* **82**: 605–610. doi:10.2307/2289471.
- Silverman, B.W. 1986. *Density estimation for statistics and data analysis*. Chapman and Hall/CRC, London and New York.
- Tadikamalla, P.R., and Johnson, N.L. 1982. Systems of frequency curves generated by transformations of logistic variables. *Biometrika*, **69**: 461–465. doi:10.1093/biomet/69.2.461.
- Wang, M. 2005. *Distributional modelling in forestry and remote sensing*. Ph.D. thesis, University of Greenwich, Greenwich, U.K.
- Wang, M., and Rennolls, K. 2005. Tree diameter distribution modelling: introducing the logit–logistic. *Can. J. For. Res.* **35**: 1305–1313. doi:10.1139/x05-057.
- Zhang, L., Gove, J.H., Liu, C., and Leak, W.B. 2001. A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid, uneven-aged stands. *Can. J. For. Res.* **31**: 1654–1659. doi:10.1139/cjfr-31-9-1654.
- Zhang, L., Packard, K.C., and 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. doi:10.1139/x03-054.
- Zhou, B., and McTague, J.P. 1996. Comparison and evaluation of five methods of estimation of the Johnson system parameters. *Can. J. For. Res.* **26**: 928–935.
- Zohrer, F. 1972. The beta distribution for best fit of stem diameter distribution. 3rd Conf. Advisory Group For. Stat. Proc. Publ. 72-3. IUFRO, Institut National Recherche Agronomique, Paris. pp. 91–106.