# Evaluation of simple quantile estimation functions for modeling forest diameter distributions in even-aged stands of interior Douglas-fir

JAMES D. NEWBERRY[1]
*Potlatch Corporation, P.O. Box 1016, Lewiston, ID 83501, U.S.A.*
AND
JAMES A. MOORE AND LIANJUN ZHANG[2]
*Department of Forest Resources, University of Idaho, Moscow, ID 83843, U.S.A.*

Received August 12, 1992
Accepted March 29, 1993

NEWBERRY, J.D., MOORE, J.A., and ZHANG, L. 1993. Evaluation of simple quantile estimation functions for modeling forest diameter distributions in even-aged stands of interior Douglas-fir. Can. J. For. Res. **23**: 2376–2382.

The method of percentiles usually involves simultaneously solving equations for probability distribution parameters as functions of sample-based estimates of the appropriate quantiles. Eight simple distribution-free methods for estimating quantiles from sample-based order statistics were evaluated empirically using even-aged Douglas-fir (*Pseudotsuga menziesii* var. *glauca* (Beissn.) Franco) diameter distributions from the Inland Northwest. Two methods, calculated by weighting adjacent order statistics, consistently gave the best results for both the Weibull and Johnson's $S_B$ distributions. Certain distributional shapes were also evaluated to determine if they influenced the quantile estimation method. Although some influence was detected, the best methods were usually best across all categories.

NEWBERRY, J.D., MOORE, J.A., et ZHANG, L. 1993. Evaluation of simple quantile estimation functions for modeling forest diameter distributions in even-aged stands of interior Douglas-fir. Can. J. For. Res. **23** : 2376–2382.

La méthode des percentiles est habituellement impliquée simultanément dans la solution d'équations pour des paramètres de fonction de probabilités comme fonction des estimations échantillonnales des quantiles appropriés. Huit méthodes simples de distribution pour l'estimation des quantiles en utilisant des statistiques d'ordre échantillonnales ont été évaluées empiriquement à partir des distributions de diamètres de sapins de Douglas (*Pseudotsuga menziesii* var. *glauca* (Beissn.) Franco) équienne du Nord-Ouest intérieur. Deux méthodes, calculées par les statistiques d'ordre adjacents pondérés ont donné les meilleurs résultats, de façon consistante, pour les deux distributions de Weibull et de Johnson $S_B$. Certaines formes de distribution ont été aussi évaluées pour déterminer si elles influençaient la méthode d'estimation du quantile. Même si une certaine influence a été détectée, les meilleures méthodes étaient généralement les meilleures, toute catégorie.

[Traduit par la rédaction]

## Introduction

Probability distributions are often used to describe the frequency distribution of tree breast-height diameters in forest stands. These distributions can take on a variety of shapes and typically can be described as nonsymmetric (either positively or negatively skewed). Although many different distribution functions have been used, the Weibull and Johnson's $S_B$ distributions have probably received the most attention in the forestry literature.

The method of percentiles or quantiles is one of several methods used to estimate the parameters of these probability distributions. Percentile parameter estimation involves solving equations for the distribution parameters as functions of one or more estimated quantiles, which can be derived as functions of the sample-based order statistics. Several distribution-free methods have been proposed for estimating the quantiles. The purpose of this study was to compare and evaluate several of these sample quantile estimators in the process of modeling forest diameter distributions with either the Weibull or Johnson's $S_B$ distributions for even-aged Douglas-fir (*Pseudotsuga menziesii* var. *glauca* (Beissn.) Franco) stands in the Inland Northwest.

### *Quantile estimation functions*

Let $X$ be a continuous random variable and let $x_1, x_2, \ldots, x_n$ be a sample of size $n$ from $X$. Let $\hat{F}(x)$ be the fraction of $x_1, x_2, \ldots, x_n$ less than or equal to some value $x$. Further, let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the sample-based order statistics such that $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$. Based on these assumptions, the sample-based functions used to estimate the $p$th ($0 < p < 1$) quantile are given in Table 1. The notation and naming conventions used in Table 1 follow those in the UNIVARIATE procedure (SAS Institute Inc. 1985) and Parrish (1990).

Most of the estimators, $Y_1$, $Y_2$, $Y_3$, $Y_5$, $Y_6$, and $Y_7$ are based on either $x_{(j)}$ and (or) $x_{(j+1)}$ where $np = j + g$, $j$ and $g$ are the integer and decimal portions of $np$, respectively. The estimate of $Y_1$ for the $p$th quantile as discussed by Parzen (1979) is the weighted average, based on $g$, of the order statistics on either side of the point $np$. This method was used by Bailey et al. (1981) in the process of estimating parameters of the Weibull distribution for modeling diameter distributions of thinned slash pine (*Pinus elliottii* Engelm.) plantations. Methods $Y_2$, $Y_3$, $Y_6$, and $Y_7$ utilize a single order statistic as the quantile estimate. The $Y_2$ method of estimating the $p$th quantile is the order statistic closest to $np$. Method $Y_3$, also discussed by Parzen (1979) and used by Dubey (1967), can be described as the smallest value of $x$ such that $\hat{F}(x) \ge p$. Method $Y_7$ can generally be described as the smallest value of $x$ such that $\hat{F}(x) > p$ and method $Y_6$ can generally be described as the largest value of $x$ such that $\hat{F}(x) \le p$. Minor exceptions to these descriptions of $Y_6$ and $Y_7$ do occur at the extremes of the distribution. Zarnoch and Dell (1985) and Shiver (1988) estimated parameters of the Weibull distribution for stand diameter distributions using $Y_6$. Csörgő and Révész (1978) and David (1981) defined the $p$th quantile after $Y_7$. Method $Y_5$'s

[1]Author to whom all correspondence should be addressed.
[2]Present address: Department of Plant and Soil Science, Alabama A & M University, Normal, AL 35762, U.S.A.

Printed in Canada / Imprimé au Canada

TABLE 1. Quantile estimators of the $p$th quantile from a sample of size $n$

$$Y_1 = (1 - g)x_{(j)} + gx_{(j+1)}$$
$$Y_2 = x_{(j)} \text{ if } g < 0.5$$
$$= x_{(j+1)} \text{ if } g \geq 0.5$$
$$Y_3 = x_{(j)} \text{ if } g = 0$$
$$= x_{(j+1)} \text{ if } g > 0$$
$$Y_4 = (1 - h)x_{(k)} + hx_{(k+1)}$$
$$Y_5 = \frac{(x_{(j)} + x_{(j+1)})}{2} \text{ if } g = 0$$
$$= x_{(j+1)} \text{ if } g > 0$$
$$Y_6 = x_{(j)}$$
$$Y_7 = x_{(j+1)}$$
$$Y_8 = (0.5 + i - np)x_{(i)} + (0.5 - i + np)x_{(i+1)}$$

NOTE: $x_{(j)}$ is the $j$th order statistic; $np = j + g$ and $j = [np]$; $(n + 1)p = k + h$ and $k = [(n + 1)p]$; and $i = [np + 0.5]$, where $[a]$ is the greatest integer $\leq a$.

TABLE 2. Summary characteristics of the 92 managed Douglas-fir installations from the Inland Northwest

| DBH distribution | Minimum | Median | Maximum |
|---|---|---|---|
| Mean (in.) | 5.41 | 9.74 | 15.39 |
| $n$ | 22 | 45.5 | 145 |
| Coefficient of variation (%) | 11.9 | 26.5 | 52.1 |
| Skewness | −0.790 | 0.285 | 1.590 |
| Kurtosis | −1.210 | −0.380 | 6.510 |

NOTE: Skewness $= \frac{\Sigma(x_j - \bar{x})^3}{ns^3}$, for $j = 1, 2, ..., n$

Kurtosis $= \frac{\Sigma(x_j - \bar{x})^4}{ns^4} - 3$, for $j = 1, 2, ..., n$

where $s$ is the sample standard deviation, $\bar{x}$ is the sample mean, and $n$ is the sample size.

estimate of the $p$th quantile is similar to $Y_7$ except that an average of $x_{(j)}$ and $x_{(j+1)}$ is used when $g = 0$. This is the definition used for the $p$th quantile in Conover (1980).

Methods $Y_4$ and $Y_8$ estimate the $p$th quantile as the weighted average of adjacent order statistics as does $Y_1$. Method $Y_4$, mentioned by Harrell and Davies (1982) as a "traditional" quantile estimator, weights the order statistics around the point $(n + 1)p$, while $Y_8$ weights the order statistics around $np + 0.5$. Parzen (1979) indicated that $Y_8$ should provide good quantile estimates under conditions of small sample sizes and symmetric distributions.

Using the equations from Table 1, seven of the eight quantile functions can be ranked according to the relative value of their estimates. It can be shown, for $g = 0$, that $Y_1 = Y_2 = Y_3 = Y_6 < Y_5 = Y_8 < Y_7$; for $0 < g < 0.5$, that $Y_2 = Y_6 < Y_1 < Y_8 < Y_3 = Y_5 = Y_7$; for $0.5 \leq g < 1.0$, that $Y_6 < Y_1 < Y_2 = Y_3 = Y_5 = Y_7 \leq Y_8$. Method $Y_4$ cannot be so easily characterized. In addition to $g$, the ranking of $Y_4$ depends on $p$. In general, for a given value of $g$, the value of $Y_4$ increases (or at least does not decrease) as $p$ increases.

Parrish (1990) compared 10 methods for estimating quantiles for normal distributions, eight of which are evaluated in this study. The best estimators were $Y_8$ and an estimator (Parrish's $Y_9$) calculated as the weighted average of all sample order statistics. This estimator has much greater computational requirements than does $Y_8$, and Parrish concluded this may make $Y_8$ the most practical. We chose to evaluate only those estimators that are relatively simple and easy to calculate.

## Materials and methods

### *Data base description*

The data used to evaluate the quantile estimation methods were obtained from 92 installations established by the Intermountain Forest Tree Nutrition Cooperative. The geographic regions covered by these installations are central and northeastern Washington, northeastern Oregon, northern and central Idaho, and western Montana. These installations were established in even-aged and single-species Douglas-fir stands to test effects of nitrogen fertilization regimes and represent a broad range of second growth conditions. Most stands had been thinned 5–12 years prior to installation establishment. Each installation consists of six plots, two of which received no fertilization treatments. Plot sizes ranged from 0.1 to 0.2 acres (1 acre = 0.405 ha). The trees from the two control plots were combined for the purposes of this analysis. All trees were measured for diameter at 4.5 feet (1 ft = 0.305 m) above ground level (DBH) to the nearest 0.01 inch. Table 2 summarizes some important characteristics of these installations. A more complete description of these data is given by Moore and others (1983).

### *Methods*

The three-parameter Weibull distribution (Weibull 1951) has a probability density function (PDF) as follows:

$$f(x) = \left(\frac{c}{b}\right)\left(\frac{x - a}{b}\right)^{c-1} \exp\left[-\left(\frac{x - a}{b}\right)^c\right] \quad \text{for } a \leq x < \infty$$

$$= 0 \text{ elsewhere}$$

where $x$ is the Weibull random variable, $a$ is the location parameter ($a \geq 0$ and $x \geq a$), $b$ is the scale parameter ($b > 0$), and $c$ is the shape parameter ($c > 0$). We used the method proposed by Zanakis (1979) to estimate the three parameters using percentiles. This method utilizes $x_{(1)}$, $x_{(2)}$, and $x_{(n)}$ to estimate the location parameter ($a$):

$$\hat{a} = \frac{x_{(1)}x_{(n)} - x_{(2)}^2}{x_{(1)} + x_{(n)} - 2x_{(2)}}$$

if $x_{(2)} - x_{(1)} < x_{(n)} - x_{(2)}$; otherwise $\hat{a} = x_{(1)}$. The scale parameter ($b$) is estimated from the relationship $\hat{a} + \hat{b} = X_{63}$, where $X_{63}$ is the 63rd percentile of the Weibull distribution. The shape parameter ($c$) is then estimated as

$$\hat{c} = \frac{\ln\left[\frac{\ln(1 - 0.97)}{\ln(1 - 0.17)}\right]}{\ln\left[\frac{X_{97} - \hat{a}}{X_{17} - \hat{a}}\right]}$$

where $X_{17}$ and $X_{97}$ are the 17th and 97th percentiles, respectively, of the Weibull distribution. Dubey (1967) showed the 17th and 97th percentiles provided the asymptotically most efficient estimate of $c$. Shiver (1988) and Zarnoch and Dell (1985) used this parameter estimation method in their

TABLE 3. Average differences among eight quantile estimation methods for the 17th, 50th, 63rd, 95th, and 97th percentiles for 92 managed Douglas-fir installations

| | Average difference for percentile (in.) | | | | |
|---|---|---|---|---|---|
| Difference | 17 | 50 | 63 | 95 | 97 |
| $Y_1 - Y_2$ | 0.019 46 | −0.036 63 | −0.013 37 | 0.013 59 | −0.017 39 |
| $Y_1 - Y_3$ | −0.109 89 | −0.036 63 | −0.063 15 | −0.274 46 | −0.270 00 |
| $Y_1 - Y_4$ | −0.037 93 | −0.071 63 | −0.089 02 | −0.581 52 | −0.897 17 |
| $Y_1 - Y_5$ | −0.109 89 | −0.071 63 | −0.063 15 | −0.299 35 | −0.270 00 |
| $Y_1 - Y_6$ | 0.111 09 | 0.035 43 | 0.082 72 | 0.212 17 | 0.340 76 |
| $Y_1 - Y_7$ | −0.109 89 | −0.107 28 | −0.063 15 | −0.324 35 | −0.270 00 |
| $Y_1 - Y_8$ | −0.109 78 | −0.071 63 | −0.069 35 | −0.282 17 | −0.425 11 |
| $Y_2 - Y_3$ | −0.129 35 | 0.000 00 | −0.049 78 | −0.288 04 | −0.252 61 |
| $Y_2 - Y_4$ | −0.057 39 | −0.035 00 | −0.075 65 | −0.595 11 | −0.879 78 |
| $Y_2 - Y_5$ | −0.129 35 | −0.035 00 | −0.049 78 | −0.312 93 | −0.252 61 |
| $Y_2 - Y_6$ | 0.091 63 | 0.072 07 | 0.096 09 | 0.198 59 | 0.358 15 |
| $Y_2 - Y_7$ | −0.129 35 | −0.070 65 | −0.049 78 | −0.337 93 | −0.252 61 |
| $Y_2 - Y_8$ | −0.129 24 | −0.035 00 | −0.055 98 | −0.295 76 | −0.407 72 |
| $Y_3 - Y_4$ | 0.071 96 | −0.035 00 | −0.025 87 | −0.307 07 | −0.627 17 |
| $Y_3 - Y_5$ | 0.000 00 | −0.035 00 | 0.000 00 | −0.024 89 | 0.000 00 |
| $Y_3 - Y_6$ | 0.220 98 | 0.072 07 | 0.145 87 | 0.486 63 | 0.610 76 |
| $Y_3 - Y_7$ | 0.000 00 | −0.070 65 | 0.000 00 | −0.049 89 | 0.000 00 |
| $Y_3 - Y_8$ | 0.000 11 | −0.035 00 | −0.006 20 | −0.007 72 | −0.155 11 |
| $Y_4 - Y_5$ | −0.071 96 | 0.000 00 | 0.025 87 | 0.282 17 | 0.627 17 |
| $Y_4 - Y_6$ | 0.149 02 | 0.107 07 | 0.171 74 | 0.793 70 | 1.237 93 |
| $Y_4 - Y_7$ | −0.071 96 | −0.035 65 | 0.025 87 | 0.257 17 | 0.627 17 |
| $Y_4 - Y_8$ | −0.071 85 | 0.000 00 | 0.019 67 | 0.299 35 | 0.472 07 |
| $Y_5 - Y_6$ | 0.220 98 | 0.107 07 | 0.145 87 | 0.511 52 | 0.610 76 |
| $Y_5 - Y_7$ | 0.000 00 | −0.035 65 | 0.000 00 | −0.025 00 | 0.000 00 |
| $Y_5 - Y_8$ | 0.000 11 | 0.000 00 | −0.006 20 | 0.017 17 | −0.155 11 |
| $Y_6 - Y_7$ | −0.220 98 | −0.142 72 | −0.145 87 | −0.536 52 | −0.610 76 |
| $Y_6 - Y_8$ | −0.220 87 | −0.107 07 | −0.152 07 | −0.494 35 | −0.765 87 |
| $Y_7 - Y_8$ | 0.000 11 | 0.035 65 | −0.006 20 | 0.042 17 | −0.155 11 |

evaluation of different estimation methods for the Weibull distribution.

The PDF of the Johnson's $S_B$ distribution (Johnson 1949) can be expressed as

$$f(x) = \left(\frac{\delta}{\sqrt{2\pi}}\right)\left(\frac{\lambda}{(x-\varepsilon)(\varepsilon+\lambda-x)}\right) \times \exp\left\{-\frac{1}{2}\left[\gamma + \delta \ln\left(\frac{x-\varepsilon}{\varepsilon+\lambda-x}\right)\right]^2\right\} \quad \text{for } \varepsilon < x < \varepsilon + \lambda$$

$$= 0 \text{ elsewhere}$$

where $x$ is the $S_B$ random variable, $\lambda$ and $\delta > 0$, $-\infty < \varepsilon < \infty$, and $-\infty < \gamma < \infty$. The choice of percentiles for estimating the parameters of the $S_B$ distribution is not clear. Typically, the two shape parameters ($\gamma$ and $\delta$) have been estimated following independent estimates of the minimum ($\varepsilon$) and range ($\lambda$) parameters (see for example, Knoebel and Burkhart (1991) and Newberry and Burk (1985)). Both of these studies used the 50th and 95th percentiles to estimate the two shape parameters using the following system of equations:

$$Z_{50} = \hat{\gamma} + \hat{\delta} X_{50}$$

$$Z_{95} = \hat{\gamma} + \hat{\delta} X_{95}$$

where $X_{50}$ and $X_{95}$ are the 50th and 95th percentiles, respectively, from the $S_B$ distribution and $Z_{50}$ and $Z_{95}$ are the 50th and 95th percentiles, respectively, from the standard normal distribution. Knoebel and Burkhart (1991) used $x_{(1)} - 0.5$ and $x_{(n)} - x_{(1)} + 1.5$ to independently estimate the minimum and range, respectively. We, therefore, chose the 50th and 95th percentiles to estimate the two shape parameters given Knoebel and Burkhart's estimates of the minimum and range.

The first step in our analysis was to estimate the 17th, 50th, 63rd, 95th, and 97th percentiles for each of the 92 installations using each of the eight sample quantile functions. For each percentile and installation, pairwise differences and pairwise absolute differences were calculated among all eight quantile estimates. These differences were averaged over all 92 installations for comparison and evaluation.

Next, using the appropriate percentiles and estimates for the minimum and range as discussed above, we estimated the parameters of both the Weibull and $S_B$ distributions for each installation by the method of percentiles. The predicted cumulative distribution function (CDF using the parameters estimated by the method of percentiles) was then compared with the empirical CDF at each sample diameter for each installation. Our definition of the empirical CDF was

$$F^*(x_{(i)}) = \frac{i - \frac{1}{3}}{n + \frac{1}{3}}, \quad \text{for } 1 \le i \le n$$

where $F^*$ is the empirical CDF and $x_{(i)}$ is the $i$th order statistic from a sample of size $n$. This definition was recommended

TABLE 4. Average absolute differences among eight quantile estimation methods for the 17th, 50th, 63rd, 95th, and 97th percentiles for 92 managed Douglas-fir installations in the Inland Northwest

| | Average absolute difference for percentile (in.) | | | | |
|---|---|---|---|---|---|
| Difference | 17 | 50 | 63 | 95 | 97 |
| $Y_1 - Y_2$ | 0.060 98 | 0.036 63 | 0.036 41 | 0.109 89 | 0.126 74 |
| $Y_1 - Y_3$ | 0.109 89 | 0.036 63 | 0.063 15 | 0.274 46 | 0.270 00 |
| $Y_1 - Y_4$ | 0.037 93 | 0.071 63 | 0.089 02 | 0.581 52 | 0.897 17 |
| $Y_1 - Y_5$ | 0.109 89 | 0.071 63 | 0.063 15 | 0.299 35 | 0.270 00 |
| $Y_1 - Y_6$ | 0.111 09 | 0.035 43 | 0.082 72 | 0.212 17 | 0.340 76 |
| $Y_1 - Y_7$ | 0.109 89 | 0.107 28 | 0.063 15 | 0.324 35 | 0.270 00 |
| $Y_1 - Y_8$ | 0.109 78 | 0.071 63 | 0.069 35 | 0.282 17 | 0.425 11 |
| $Y_2 - Y_3$ | 0.129 35 | 0.000 00 | 0.049 78 | 0.288 04 | 0.252 61 |
| $Y_2 - Y_4$ | 0.075 65 | 0.035 00 | 0.075 65 | 0.595 11 | 0.879 78 |
| $Y_2 - Y_5$ | 0.129 35 | 0.035 00 | 0.049 78 | 0.312 93 | 0.252 61 |
| $Y_2 - Y_6$ | 0.091 63 | 0.072 07 | 0.096 09 | 0.198 59 | 0.358 15 |
| $Y_2 - Y_7$ | 0.129 35 | 0.070 65 | 0.049 78 | 0.337 93 | 0.252 61 |
| $Y_2 - Y_8$ | 0.129 24 | 0.035 00 | 0.055 98 | 0.295 76 | 0.407 72 |
| $Y_3 - Y_4$ | 0.080 22 | 0.035 00 | 0.040 87 | 0.307 07 | 0.628 04 |
| $Y_3 - Y_5$ | 0.000 00 | 0.035 00 | 0.000 00 | 0.024 89 | 0.000 00 |
| $Y_3 - Y_6$ | 0.220 98 | 0.072 07 | 0.145 87 | 0.486 63 | 0.610 76 |
| $Y_3 - Y_7$ | 0.000 00 | 0.070 65 | 0.000 00 | 0.049 89 | 0.000 00 |
| $Y_3 - Y_8$ | 0.049 46 | 0.035 00 | 0.032 28 | 0.172 28 | 0.298 15 |
| $Y_4 - Y_5$ | 0.080 22 | 0.000 00 | 0.040 87 | 0.282 17 | 0.628 04 |
| $Y_4 - Y_6$ | 0.149 02 | 0.107 07 | 0.171 74 | 0.793 70 | 1.237 93 |
| $Y_4 - Y_7$ | 0.080 22 | 0.035 65 | 0.040 87 | 0.262 17 | 0.628 04 |
| $Y_4 - Y_8$ | 0.071 85 | 0.000 00 | 0.019 67 | 0.299 35 | 0.472 07 |
| $Y_5 - Y_6$ | 0.220 98 | 0.107 07 | 0.145 87 | 0.511 52 | 0.610 76 |
| $Y_5 - Y_7$ | 0.000 00 | 0.035 65 | 0.000 00 | 0.025 00 | 0.000 00 |
| $Y_5 - Y_8$ | 0.049 46 | 0.000 00 | 0.032 28 | 0.147 39 | 0.298 15 |
| $Y_6 - Y_7$ | 0.220 98 | 0.142 72 | 0.145 87 | 0.536 52 | 0.610 76 |
| $Y_6 - Y_8$ | 0.220 87 | 0.107 07 | 0.152 07 | 0.494 35 | 0.765 87 |
| $Y_7 - Y_8$ | 0.049 46 | 0.035 65 | 0.032 28 | 0.172 39 | 0.298 15 |

by Hoaglin et al. (1983) since it closely approximates the median of the distribution of $X_{(i)}$ for any continuous distribution where $X_{(i)}$ is the random variable associated with $x_{(i)}$. The 1/3 in the numerator and denominator is used to ensure that $x_{(n)}$ is not considered the largest population value. The difference between the estimated cumulative probabilities of the empirical and predicted distribution functions was calculated at each sample diameter observation in an installation. The average difference, average absolute difference, and maximum difference were then determined for each installation. These differences were averaged over all installations to determine which quantile estimation method had the least error.

We divided the installations into three categories and repeated the comparisons of the empirical and predicted distribution functions in an attempt to assess whether certain distributional shapes affect the choice of the quantile estimation function. To this end, installations were categorized by the amount and direction of distribution skewness. We used the skewness classes of less than −0.20, −0.20 to 0.20, and greater than 0.20. Although the choice of these classes is somewhat arbitrary, it does separate the more highly negatively skewed distributions from the more highly positively skewed distributions and provides at least 10 installations in each category.

## Results

The 17th, 50th, 63rd, 95th, and 97th percentiles were estimated using each of the eight quantile functions. Tables 3 and 4 show the pairwise differences and pairwise absolute differences, respectively, among the eight quantile estimates averaged over all 92 installations for each percentile. As we would expect from our earlier analytical evaluations of the functions in Table 1, most of the average differences (relative to the absolute differences) are in a particular direction (bias). The differences among methods are, in general, greatest in the tails of the distributions. Method $Y_6$ estimates are most different from the other methods for all five percentiles. For the 17th percentile, the largest difference is 0.22 inches (1 inch = 2.54 cm) between $Y_6$ and $Y_3$, $Y_5$, and $Y_7$. Methods $Y_6$ and $Y_7$ have the largest difference of 0.14 inches for the 50th percentile. For the 63rd, 95th, and 97th percentiles, $Y_4$ and $Y_6$ have the largest differences of 0.17, 0.79, and 1.24 inches, respectively.

For estimating the median, $Y_2$ and $Y_3$ provide the same estimate in all cases. This follows since $Y_2$ and $Y_3$ equal $x_{(j)}$ for $g = 0$ ($n$ even) and $x_{(j+1)}$ for $g = 0.5$ ($n$ odd). Methods $Y_4$, $Y_5$, and $Y_8$ which give identical estimates of the median are the only methods, out of the eight evaluated, which produce median estimates consistent with the traditional definition.

Methods $Y_3$, $Y_5$, and $Y_7$ yield the same estimates for the 17th, 63rd, and 97th percentiles since $g \neq 0$ for these percentiles and number of sample observations. This implies these three methods will lead to the same parameter estimates for the Weibull distribution using the method of percentiles. However, estimates of the 50th and 95th percentiles are different among methods $Y_3$, $Y_5$, and $Y_7$ and result in different parameter estimates for the $S_B$ distributions.

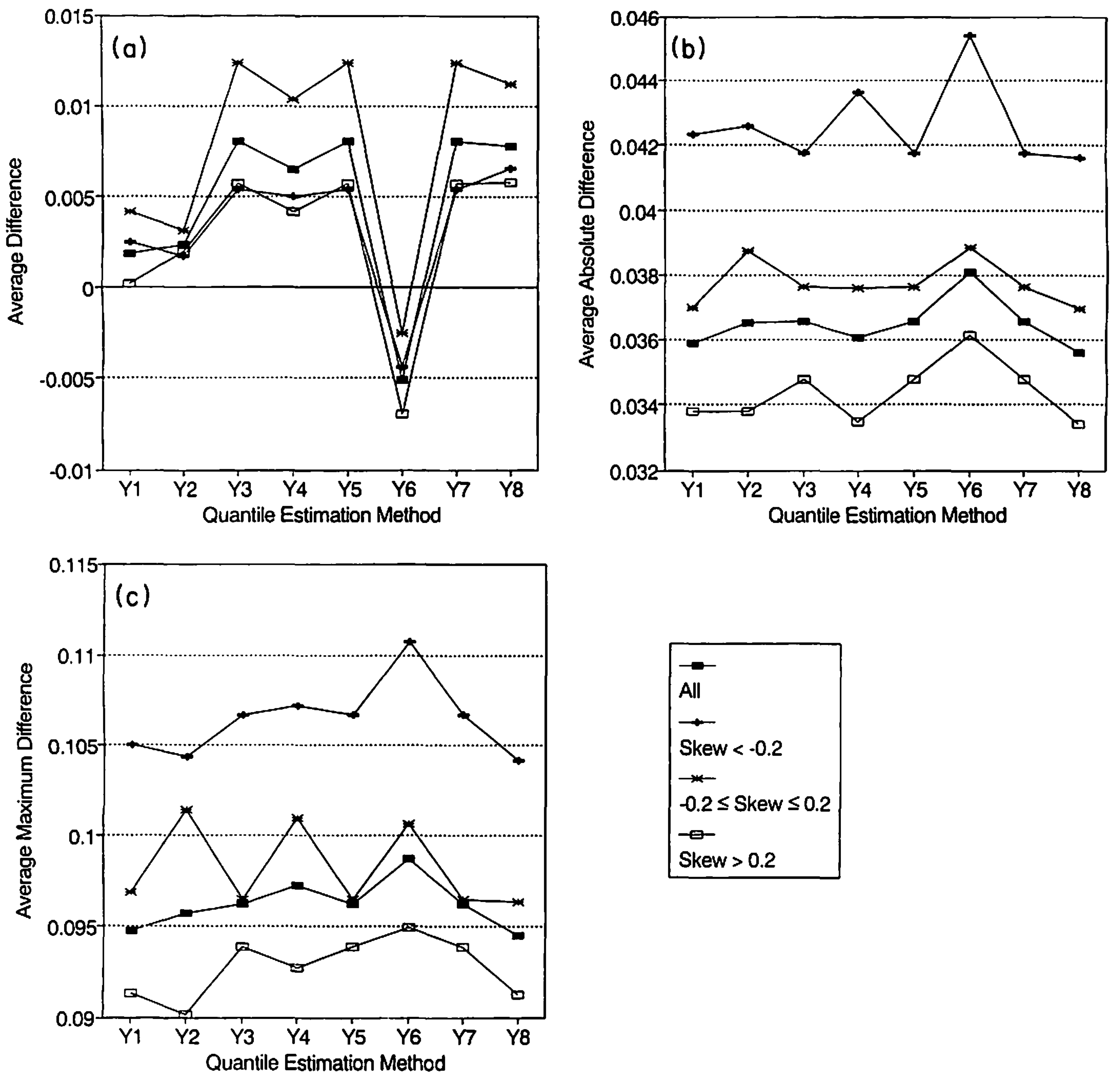


FIG. 1. Average differences (*a*), average absolute differences (*b*), and average maximum differences (*c*) between the empirical and predicted Weibull distribution functions for eight quantile estimation methods. Lines between points are used for ease of interpretation and do not indicate trends between points.

After the percentiles were estimated by each of the eight methods for each installation, the parameters of the Weibull and $S_B$ distributions were estimated using the method of percentiles as outlined previously. The empirical cumulative distribution function $F^*(x)$ was compared with the predicted cumulative distribution function at each sample diameter for all installations. Average differences, average absolute differences, and average maximum differences were determined over all installations. Figures 1 and 2 provide the results of the comparisons for the Weibull and $S_B$ distributions, respectively. These tables include comparisons for all 92 installations: 10 installations with negatively skewed distributions, 33 approximately symmetric distributions, and 49 installations with positively skewed distributions.

### *Weibull distribution*

In general, methods $Y_1$ and $Y_2$ have the smallest average differences across all four categories of comparison. Method $Y_6$ has the smallest average differences when the empirical distributions are approximately symmetric. The quantile methods with the largest average differences are usually the identical estimators; $Y_3$, $Y_5$, and $Y_7$, although they perform somewhat better with the positively skewed distributions. All methods except $Y_6$ show less bias associated with the estima-

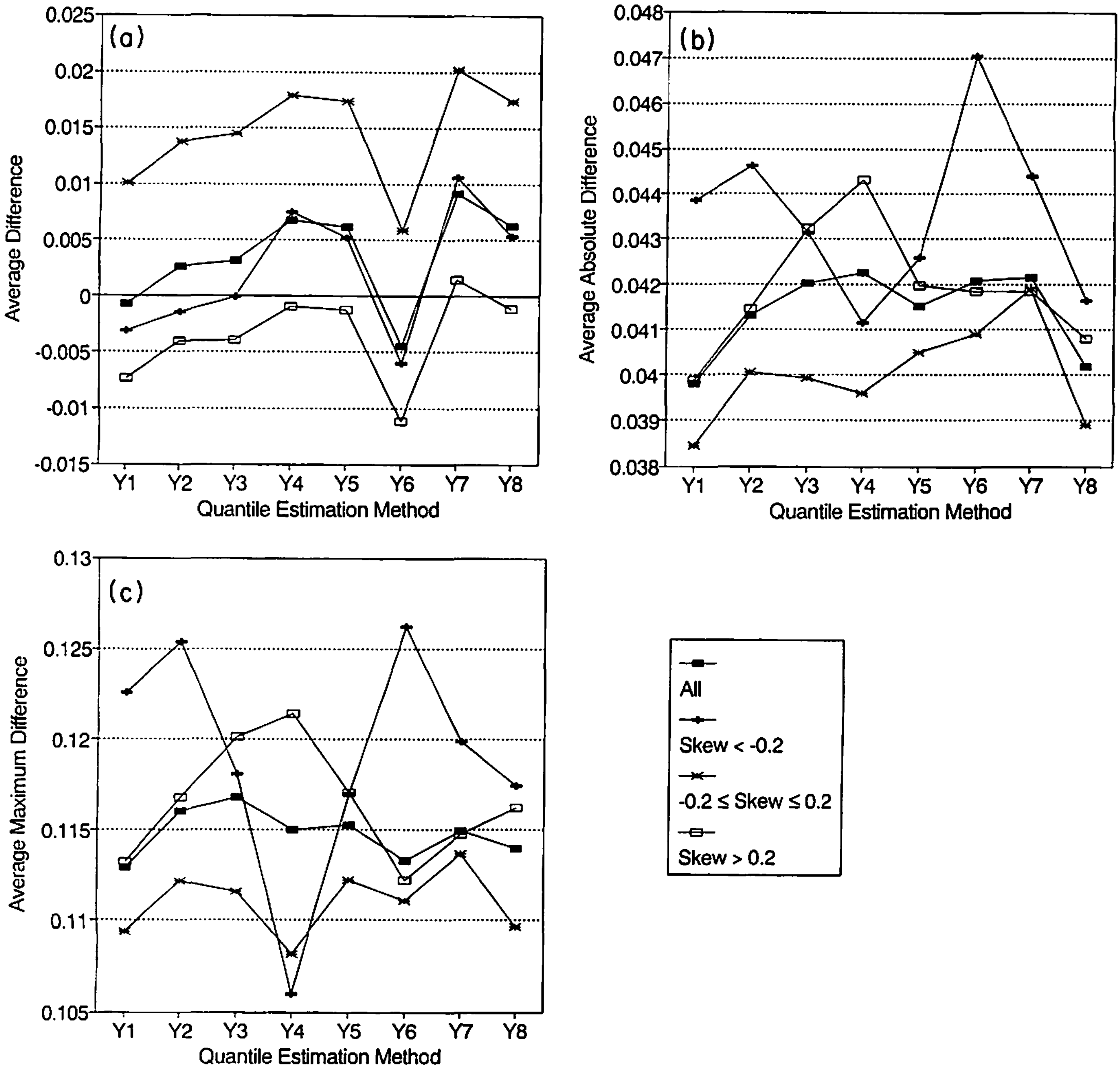


FIG. 2. Average differences (*a*), average absolute differences (*b*), and average maximum differences (*c*) between the empirical and predicted $S_B$ distribution functions for eight quantile estimation methods. Lines between points are used for ease of interpretation and do not indicate trends between points.

tion of skewed distributions. In general, the methods that average adjacent order statistics, $Y_1$, $Y_4$, and $Y_8$ have the least average absolute differences between the empirical and predicted Weibull cumulative probability distributions. Method $Y_8$ has the smallest average absolute difference across all categories. Methods $Y_1$ and $Y_4$ perform well except for the negatively skewed situations. Method $Y_6$ has the largest average absolute differences. All estimators, on average, perform the poorest with the negatively skewed distributions for absolute differences.

For average maximum differences, $Y_1$ and $Y_8$ again provide the smallest values across most categories of comparison. Method $Y_2$ also has small average maximum differences in all categories except when the distributions are approximately symmetric. Methods $Y_4$ and $Y_6$ have the largest differences across most categories. As with the absolute differences, the maximum differences decrease from the negatively skewed to the positively skewed distributions.

Method $Y_1$ consistently had the smallest values across all types of differences and categories evaluated of the methods that always averaged adjacent order statistics. Method $Y_8$ performed as well as $Y_1$ except for larger average differences. Method $Y_4$ was quite inconsistent. Method $Y_2$ consistently had the smallest differences and $Y_6$ had the largest differences for those methods utilizing only one order statistic. In general, shape of the distribution did not greatly affect the ranking of the quantile estimation methods for the Weibull distribution.

*$S_B$ distribution*

Methods $Y_1$, $Y_2$, and $Y_3$ have the smallest average differences between the empirical and predicted $S_B$ cumulative distribution functions for the eight quantile estimation methods. The three methods performed poorly with positively skewed distributions.

The largest average differences were obtained with $Y_4$ and $Y_7$. Recall that percentiles used to estimate the $S_B$ parameters in this study do not lead to identical distributions for methods $Y_3$, $Y_5$, and $Y_7$. All methods had more bias with the symmetric distributions than with either negatively or positively skewed distributions. Methods $Y_1$, $Y_2$, and $Y_8$ have the smallest average absolute differences with $Y_8$ being the most consistent across all categories. Methods $Y_1$ and $Y_2$ have larger differences with the negatively skewed distributions as well as $Y_2$ with the symmetric distributions. The largest average absolute differences were associated with methods $Y_4$ and $Y_7$, although $Y_4$ does quite well with negatively skewed distributions. Methods $Y_1$, $Y_6$, and $Y_8$ were best with respect to average maximum differences. Again there was at least one category where the performance of each of these methods declined. Methods $Y_2$ and $Y_3$ produced the largest average absolute differences. As with the Weibull distribution the estimators performed better with the symmetric and positively skewed distributions.

For the $S_B$ distribution, $Y_1$ outperforms the other methods. Methods $Y_2$ and $Y_8$ also have smaller differences. Similar to its performance with the Weibull, $Y_6$ has larger average differences than most of the other methods. Method $Y_2$ did poorly for the average maximum differences. Methods $Y_4$ and $Y_7$ had the poorest performances for the $S_B$ distribution. Similar to the results for the Weibull, the selection of the best quantile estimation method did not vary greatly with empirical distributional shapes; however, results were more variable for the $S_B$ distribution. Usually, the best overall methods had at least one category with poorer than average differences.

## Discussion and conclusions

The quantile estimation methods that performed best in this evaluation were those that, in some manner, averaged two adjacent order statistics. Methods $Y_1$ and $Y_8$ performed quite well for both the Weibull and $S_B$ distributions. Method $Y_8$ almost always had larger average differences than did $Y_1$. The results for $Y_1$ were slightly more variable across the categories evaluated than $Y_8$. As we discussed earlier, $Y_8$ is consistent with the traditional definition of the median; $Y_1$ is not. Parzen (1979) suggested $Y_8$ should work well for symmetric distributions and Parrish (1990) lends evidence to this assertion. Our evaluation suggests both methods work well for nonsymmetric situations.

With the exception of average maximum differences for the $S_B$ distribution, method $Y_2$ performed best of the methods using one order statistic in their calculation. However, results for $Y_2$ were more highly influenced by distribution shape categories. Method $Y_6$ did not perform as well with the Weibull distribution as with the $S_B$ distribution.

Our results show $Y_1$ and $Y_8$ are the best methods evaluated for estimating quantiles from our sample-based order statistics. It is our hope these results will be tested on other populations and the theoretical properties of these estimators will be evaluated for the types of distributions commonly found in forestry. We see no reason, however, why our results should not apply to situations with similar distributions.

## Acknowledgement

The support of the Intermountain Forest Tree Nutrition Cooperative is acknowledged and greatly appreciated.

Bailey, R.L., Abernathy, N.C., and Jones, E.P., Jr. 1981. Diameter distribution models for repeatedly thinned slash pine plantations. *In* Proceedings of the 1st Biennial Southern Silvicultural Research Conference, 6–7 Nov. 1980, Atlanta, Ga. *Edited by* J.P. Barnett. USDA For. Serv. Tech. Rep. SO-34. pp. 115–126.

Conover, W.J. 1980. Practical nonparametric statistics. 2nd ed. John Wiley & Sons Inc., New York.

Csörgő, M., and Révész, P. 1978. Strong approximations of the quantile process. Ann. Stat. **4**: 882–894.

David, H.A. 1981. Order statistics. 2nd ed. John Wiley & Sons Inc., New York.

Dubey, S.D. 1967. Some percentile estimators for the Weibull parameters. Technometrics, **9**: 119–129.

Harrell, F.E., and Davis, C.E. 1982. A new distribution-free quantile estimator. Biometrika, **69**: 635–640.

Hoaglin, D.C., Mosteller, F., and Tukey, J.W. 1983. Understanding robust and exploratory data analysis. John Wiley & Sons Inc., New York.

Johnson, N.L. 1949. Systems of frequency curves generated by methods of translation. Biometrika, **36**: 149–176.

Knoebel, B.R., and Burkhart, H.E. 1991. A bivariate distribution approach to modeling forest diameter distribution at two points in time. Biometrics, **47**: 241–253.

Moore, J.A., Pregitzer, K.S., Vanderploeg, J.L., and Mital, J.M. 1983. Intermountain Forest Tree Nutrition Cooperative third annual report. College of Forestry, Wildlife and Range Sciences, University of Idaho, Moscow.

Newberry, J.D., and Burk, T.E. 1985. $S_B$ distribution-based models for individual tree merchantable volume – total volume ratios. For. Sci. **31**: 389–398.

Parrish, R.S. 1990. Comparison of quantile estimators in normal sampling. Biometrics, **46**: 247–257.

Parzen, E. 1979. Nonparametric statistical data modeling. J. Am. Stat. Assoc. **74**: 105–121.

SAS Institute Inc. 1985. Statistical analysis system user's guide: basics, version 5 edition. SAS Institute Inc., Cary, N.C.

Shiver, B.D. 1988. Sample sizes and estimation methods for the Weibull distribution for unthinned slash pine plantation diameter distributions. For. Sci. **34**: 809–814.

Weibull, W. 1951. A statistical distribution function of wide applicability. J. Appl. Mech. **18**: 293–297.

Zanakis, S.H. 1979. A simulation study of some simple estimators for the three parameter Weibull distribution. J. Stat. Comput. Simul. **9**: 101–116.

Zarnoch, S.J., and Dell, T.R. 1985. An evaluation of percentile and maximum likelihood estimators of Weibull parameters. For. Sci. **31**: 260–268.