

United States
Department of
Agriculture

Forest Service



**Southern
Research Station**

Research Paper
SRS-25

Predicting Diameter Distributions of Longleaf Pine Plantations: A Comparison Between Artificial Neural Networks and Other Accepted Methodologies

**Daniel J. Leduc, Thomas G. Matney,
Keith L. Belli, and V. Clark Baldwin, Jr.**



The Authors

Daniel J. Leduc, Computer Specialist, USDA Forest Service, Southern Research Station, Forest Management Research, Pineville, LA; **Thomas G. Matney** and **Keith L. Belli**, Professors, Mississippi State University, School of Forest Resources, Mississippi State, MS; and **V. Clark Baldwin, Jr.**, Research Forester, USDA Forest Service, Southern Research Station, Forest Inventory and Analysis, Asheville, NC, respectively.

Cover: Longleaf pine plantation.

Photo by Bill Lea

September 2001

Southern Research Station
P.O. Box 2680
Asheville, NC 28802

Predicting Diameter Distributions of Longleaf Pine Plantations: A Comparison Between Artificial Neural Networks and Other Accepted Methodologies

Daniel J. Leduc, Thomas G. Matney,
Keith L. Belli, and V. Clark Baldwin, Jr.

Abstract

Artificial neural networks (NN) are becoming a popular estimation tool. Because they require no assumptions about the form of a fitting function, they can free the modeler from reliance on parametric approximating functions that may or may not satisfactorily fit the observed data. To date there have been few applications in forestry science, but as better NN software and fitting algorithms become available, they may be used to solve a wide variety of problems—particularly problems where the underlying relationship between predicted and predictors is unknown. We benchmark tested an alternative to the traditional Weibull probability distribution function, diameter-at-breast-height moment, and direct parameter prediction models for approximating stand-diameter distributions. Using a feedforward backpropagation network, we demonstrated that NN are a somewhat better option. Unlike Weibull approximations, NN solutions cannot easily be mathematically constrained to match known reality constraints, but this difficulty is easy to overcome in practice.

Keywords: Connectionist models, parallel distributed processing systems, parameter recovery, Weibull distribution.

Introduction

The three-parameter Weibull probability distribution function (Weibull 1951) can take on a wide variety of shapes, and has been found to be an applicable model for approximating tree diameter-at-breast-height (d.b.h.) distributions (Bailey and Dell 1973). Because of its plasticity, many stand-level, diameter-distribution growth-and-yield models in use today rely on the Weibull probability distribution function (Bailey and Aleixo da Silva 1988, Matney and Sullivan 1982b, Zarnoch and others 1991). However, the Weibull distribution does not span the entire function space, and its performance as a d.b.h. distribution estimator varies widely among data sets. In some d.b.h. distribution modeling cases, the Weibull function tends to produce poor estimates in the tails of the distribution. In other situations, it may lock down the tails and overcompensate by missing badly in the middle part of the distribution. Because any small miss in the middle and upper d.b.h. ranges of a distribution can have a large impact on derived volume estimates, growth-and-yield models

constructed from poor Weibull estimates will produce biased volume estimates. Thus, general applicability of a procedure does not mean that the procedure is always best, as growth-and-yield modelers are always searching for procedures to improve d.b.h. distribution estimates. Artificial neural networks (NN) may provide better estimates of d.b.h. distributions that do not rely on assuming an imperfect underlying probability model.

In general, artificial NN are appropriate in modeling situations: (1) where the application is data intensive and dependent on multiple, interacting parameters; (2) where the problem area is rich in historical data or examples; (3) where the available data are incomplete, contain errors, and describe specific examples, and (4) when the function to determine solutions is unknown or expensive to discover (Bailey and Thompson 1990). All these conditions are to some degree met by the typical growth-and-yield database. Theorems by Cybenko (1989), Sun and Cheney (1992), and Light (1992) show that a single output, single hidden-layer, feedforward network employing continuous sigmoid and other more general activation functions with a sufficient number of hidden units can approximate any continuous function to any desired accuracy. This makes them ideal for d.b.h. distributions of an unknown form. Some researchers, e.g. Josin 1987, point to Kolmogorov's (1957) theorem on the realization of real-valued functions as strong, albeit not conclusive, evidence that NN models can learn to approximate any continuous real-valued multivariate function. They minimize error in the least-mean-square sense based only on example mapping. Hassoun (1995) gives an excellent discussion on the real-value function approximation capabilities of feedforward networks.

Given the advertised great promise of neural methodology, we decided that a preliminary investigative benchmark comparison of this technique with the traditional methods of diameter-distribution modeling was warranted. The model selected for this comparison is a simple, fully connected, feedforward, backpropagation delta-learning-rule network. The two traditional Weibull probability function-based benchmark test models selected were a diameter-moment

parameter recovery system and a direct-parameter prediction model. Each of these techniques was then used to predict the diameter distributions of three distinctly different, unthinned, planted longleaf pine (*Pinus palustris* Mill.) databases from stands originating under three different conditions. The preliminary results will help us to determine the amount of additional work that will be necessary to refine the unthinned stand model and construct network models for predicting thinned stand-diameter distributions.

We report the analysis and results of the multitude of evaluation criteria selected for the preliminary benchmark testing of the NN mode.

Data

The three diameter-distribution NN training/test data sets selected for study were from longleaf pine plantations on (1) abandoned agricultural land (old fields), (2) land that had received some form of site preparation, and (3) recently cutover sites. The three data sets were plot data assembled from several experiments designed to represent the growth response of planted longleaf pine across the Southern United States. The sites were considered problem free, in terms of having low levels of competing vegetation as well as good initial and subsequent survival of planted trees. Table 1 presents the basic descriptive statistics for each of the three data sets.

The input variables selected for modeling d.b.h. distributions were stand age, base-age-25 site index, average height of dominant and codominant trees, and trees per acre. The output units selected for benchmark testing were trees per acre in each 1-inch (in.) d.b.h. class up to 20 in. In effect, by selecting these outputs we had the NN approximate the probability density function of d.b.h. at points spaced 1 in. apart. There are many other output formats that could be used, but this one fits well within the limitations of a backpropagation NN.

In regression model-building work, there is little to be gained by separating a data set into parts for fitting and testing (Hirsch 1991). The only good way to test a regression model is to obtain independent test samples from the same population. However, because NN contain a large number of nodes, and thus many parameters, there is a good chance of overfitting the training data set and capturing not just general patterns but sampling variation as well (Leahy 1994). Therefore, it is important to have both fitting/training and testing data sets as insurance against overfitting. In this study, we randomly partitioned each of the three data sets

into training and test data sets, designating a training data set approximately twice the size of the test data set.

Model Development

Neural Network

The type of NN we selected for evaluation is known as a hetero-associative, fully connected, feedforward network. This means that the network is composed of an input layer, one or more hidden layers, and a multiple unit output layer; and that all of the nodes in every layer are connected to all of the nodes in each successive layer. It is also called a backpropagation network because, during training, errors are backpropagated from the output layer to the inner connections using the delta rule. While there is a large number of possible network models to choose from, we selected the backpropagation network because it is recognized as being best suited to function approximations (Hassoun 1995). Attempting to evaluate all possible networks for a particular problem would be an overwhelming task.

Figure 1 shows the general, single hidden-layer NN backpropagation model structure adopted for each data set. Between data sets, the number of nodes was varied to obtain the lowest mean-squared error on the outputs. Otherwise, no changes in the structure were necessary to obtain a satisfactory fit for each data set. The bias term connected to both the hidden layer and the output layer is a constant value (typically equal to one) that is analogous to the intercept term in a linear regression equation.

An article by Freeman (1992) and a book by Hassoun (1995) include descriptive summaries and sufficient mathematics to explain the actual backpropagation training process. An abbreviated explanation of this process is shown in figure 2.

In the course of deciding to use a single hidden-layer network, we tested numerous network structures to determine the optimum number of hidden layers and the numbers of nodes in each. In a typical NN, there can be one or more hidden layers. Recent research into backpropagation networks shows that almost any function can be synthesized using a sufficiently complex network with a single hidden layer (NeuralWare 1991d). We made several attempts to use more than a single hidden layer but observed no noticeable improvements in precision or accuracy. Thus the single hidden-layer NN was constructed and trained for each data set.

Table 1—Descriptive statistics for the training and testing data sets used in the comparison of neural network methodology to traditional Weibull models

Variable	Old field		Site prepared		Cutover	
	Training	Testing	Training	Testing	Training	Testing
Observations (no.)	292	122	200	108	636	333
Age (years)						
Average	14.8	14.8	15.6	15.8	18.2	18.7
Range	8 – 30	9 – 30	8 – 25	10 – 25	9 – 48	9 – 48
Trees per acre						
Average	715.4	711.7	609.9	584.3	590.7	593.3
Range	74 – 1,185	284 – 1,160	123 – 1,036	133 – 1,089	133 – 1,210	133 – 1,136
Trees per acre > 1 in. d.b.h.						
Average	704.7	699.7	559.2	542.8	538.6	552.2
Range	74 – 1,185	284 – 1,160	123 – 1,000	133 – 988	133 – 1,210	133 – 1,136
Total height of trees with d.b.h. > 1 in. (ft)						
Average	32.2	32.5	20.2	20.1	34.4	35.5
Range	15 – 62	16 – 62	7 – 38	8 – 37	8 – 80	9 – 82
Total height of dominants and codominants (ft)						
Average	36.7	36.8	23.7	23.5	38.6	39.7
Range	14 – 67	15 – 69	9 – 41	10 – 41	8 – 86	8 – 88
Basal area (sq. ft.)						
Average	96.4	94.9	34.2	32.8	74.5	80.1
Range	8 – 230	28 – 194	5 – 113	6 – 104	5 – 221	5 – 229
Quadratic mean diameter (in.)						
Average	4.94	4.98	3.18	3.20	4.83	5.02
Range	2.3 – 8.4	2.7 – 7.9	1.3 – 5.7	1.6 – 6.4	1.6 – 12.7	1.5 – 12.9
Site index (b.a. 25) (ft)						
Average	56.8	57.5	38.7	37.9	56.5	56.7
Range	34 – 80	36 – 76	20 – 77	20 – 77	20 – 103	26 – 104

The number of nodes in a hidden layer is another parameter that can be changed in a NN (fig. 1). The number used for each data set was determined by experimentally finding the optimum (based on the minimum mean-squared error criterion) fit of the training data set using different numbers of nodes. These optimums were three nodes for old-field sites, four nodes for cutover sites, and six nodes for prepared sites.

Matching the attributes of a NN to the requirement that a diameter distribution must have exactly as many trees as there are in the stand required some additional calculations. First, rather than work with the actual number of trees, the output of this network was designed to be the percentage of trees in each 1-in. diameter class from 1 to 20 in. Thus, in effect, output for any d.b.h. class is the probability of a given tree belonging to that class. For a network to be

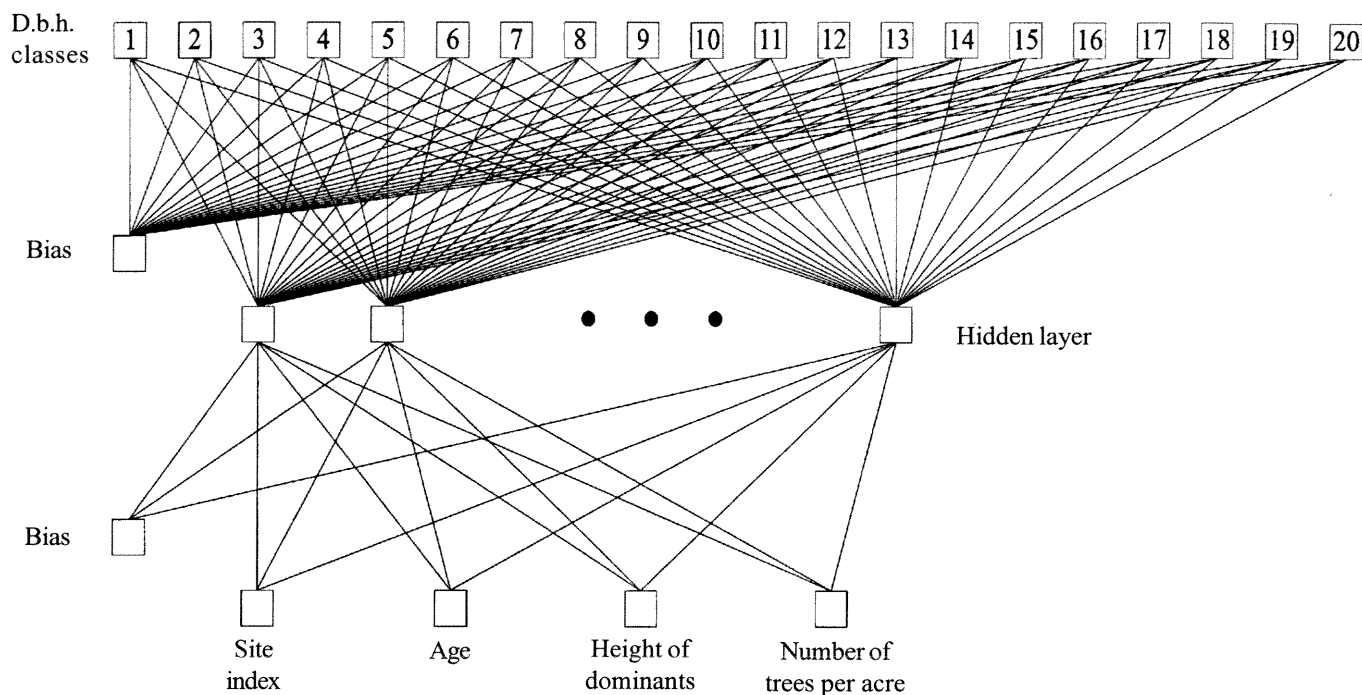


Figure 1—Diagram of the basic structure used for neural networks tested in this paper. They are all similar in structure except for a variable number of nodes in the hidden layer. Each square block is a node. Nodes can be input like age, output like the proportion of trees in any diameter class, or hidden, which are just internal summation and transformation points. Each line is a connection where variable weights are applied.

logically consistent with a probability function, all diameter-class probabilities must be nonnegative and sum to one. These constraints cannot be imposed on a NN of the type used here, but they can be obtained with minor postprocessing. This is done simply by (1) setting any output less than zero to zero, and (2) dividing each output by the sum of the outputs.

While there are many computer packages that could be used to train a NN, most of the work for this paper was done with NeuralWorks Professional II/Plus (NeuralWare 1991c). This is significant because there are many assumptions and choices that vary among different software packages. All of the assumptions will not be explained here, but specific variations from the defaults are detailed.

NeuralWorks allows the network developer to choose from among several transfer or squashing functions. We used the hyperbolic tangent function with a range of -1 to 1 along a smooth, sigmoidal curve to train all of the networks in this study. NeuralWare (1991b) recommends the function as very effective in backpropagation networks.

NeuralWorks automatically scales all inputs and outputs into a subset of the range of the transfer function chosen. For the

hyperbolic tangent, this range is from -0.8 to 0.8 . Guan and Gertner (1991) chose to use a logistic function in their survival-probability network model partly because it has a range of 0 to 1 . It is desirable to have the network outputs already in the correct range, but NeuralWare's automatic scaling eliminates the need for preliminary data scaling and makes the choice dependent only upon the slope properties of the transfer function. Extensive testing showed that the hyperbolic tangent worked best with the NN we used.

Standard backpropagation uses the delta error correction rule, which adjusts the node weights after each example is presented. We used the normalized, cumulative delta rule for all the networks trained. The cumulative delta rule accumulates the error for each output and does not adjust weights until all of the observations in an epoch are presented. It also avoids learning rate changes by dividing the sum of the errors by the square root of the number of observations in an epoch. An epoch may constitute the entire data set or any subset thereof. We used an epoch size of 100 observations.

Multiplying the backpropagated error by a learning coefficient and adding the result to the current weight makes weight adjustments in standard backpropagation. Two

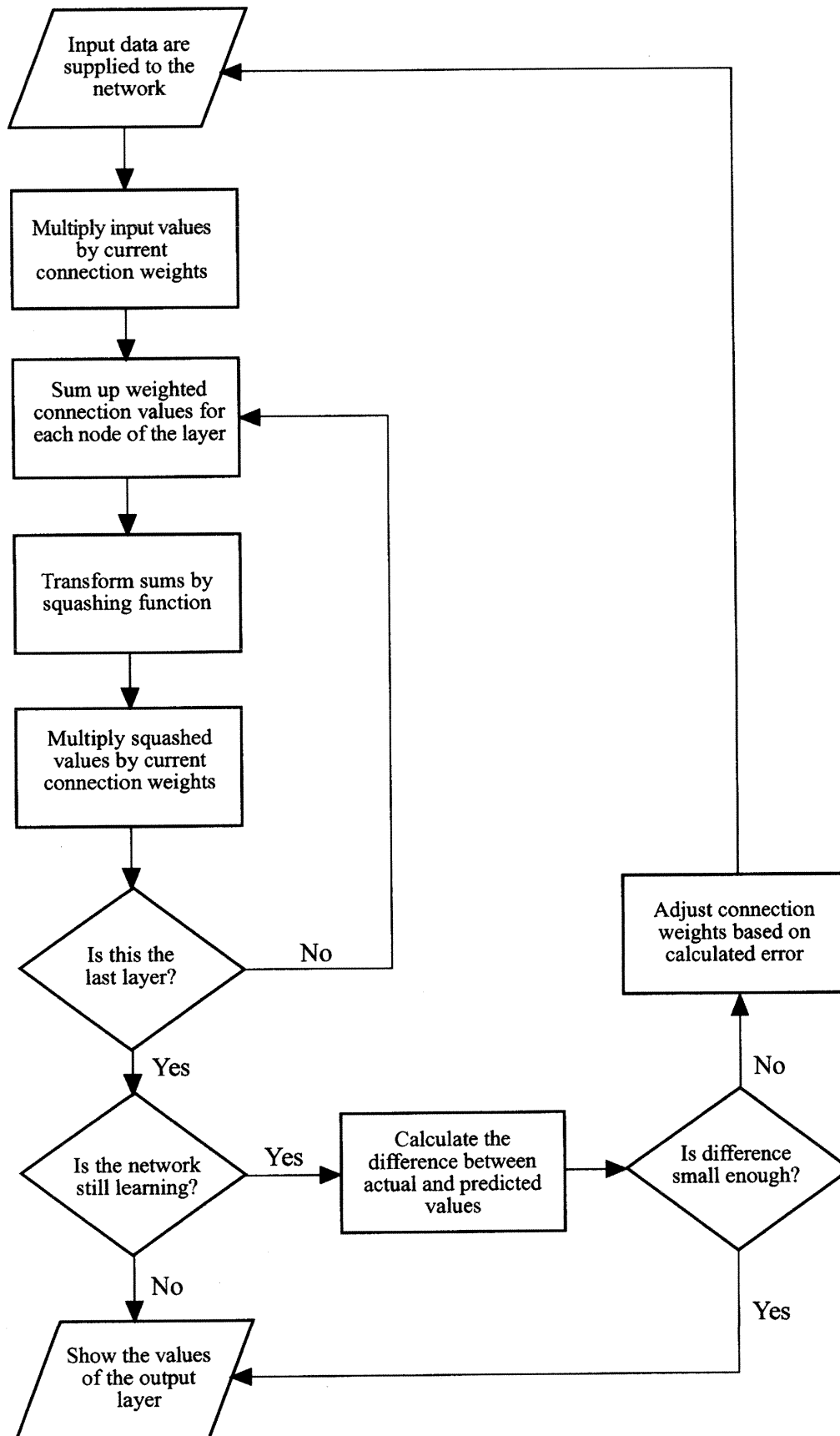


Figure 2—A flowchart of the training and use of a backpropagation neural network. A squashing function is a function such as the hyperbolic tangent, which transforms its input into a number between -1 and 1 .

common modifications to this technique were used in the training. The first is the addition of momentum that adds some portion of the previous weight adjustment to the current weight adjustment. The second, developed by Tariq Samad (1988), is called fast learning. This technique adds a multiple of the error from the next lower layer to the activation value before the weight update. Equations (1) through (3) show the change in weight for standard backpropagation, backpropagation with momentum, and backpropagation for fast learning with momentum, respectively.

$$\Delta w_{ji[s]} = \text{lcoef} * e_{j[s]} * x_{i[s-1]} \quad (1)$$

$$\Delta w_{ji[s]} = \text{lcoef} * e_{j[s]} * x_{i[s-1]} + \text{momentum} * \Delta w_{ji[s]} \quad (2)$$

$$\Delta w_{ji[s]} = \text{lcoef} * e_{j[s]} * (x_{i[s-1]} + \text{fastlrn} * e_{i[s-1]}) + \text{momentum} * \Delta w_{ji[s]} \quad (3)$$

where

$\Delta w_{ji[s]}$ = the change in weight to be applied to the connection between the i th node in layer $s-1$ to the j th node in layer s ;

$e_{j[s]}$ = the error at node j of layer s ;

$x_{i[s-1]}$ = the current output of the i th node in layer $s-1$; and
 lcoef , momentum , and fastlrn = the coefficients for learning, momentum, and the fast-learning adjustment, respectively (NeuralWare 1991a).

A final characteristic that needs to be considered in training these networks is how many iterations are necessary. We did not set a fixed number of iterations, nor did we set a fixed level of error that we considered small enough. Rather, we observed trends in the reduction of the root-mean-squared error (RMS) during the training process. During training, RMS usually drops quickly in the beginning then slows and may even climb again. For each structure that we tested, we selected parameters providing the lowest value for RMS. Observing the trend as well as the lowest value, helped us to avoid some of the spurious, apparently optimum results that can occur in any nonlinear iterative fitting process. When results appeared spurious, we reinitialized the network to see if similar or different optimum values would be generated.

Weibull with Parameters Recovered from Stand Distribution Characteristics

Parameter recovery systems for probability distribution functions have two parts. First, equations are developed to predict the expected values of d.b.h. moments and/or order

statistics such as arithmetic mean d.b.h., quadratic mean d.b.h., minimum d.b.h., and percentiles. Second, the resulting system of equations is used to solve for the unknown parameters of the distribution (Matney and Sullivan 1982a, Zarnoch and others 1991).

Matney and Farrar (1992) and Farrar and Matney (1994) successfully used the three-parameter Weibull distribution parameter recovery system selected for the benchmark test to simulate the unthinned stand yields of cutover, site-prepared loblolly pine (*P. taeda* L.) plantations and natural longleaf pine, respectively. The steps used to construct the parameter recovery system are:

1. Develop equations for predicting stand minimum d.b.h., arithmetic mean d.b.h., and quadratic mean d.b.h. from age, surviving trees per acre, and average height of dominants and codominants. The arithmetic mean d.b.h. is the first-order moment of d.b.h., and the quadratic mean d.b.h. is the square root of the second-order moment of d.b.h..
2. Set the location parameter (a) of the Weibull probability distribution equal to one-half the estimated minimum stand d.b.h.

$$a = 0.5 \text{ dbh}_{\min} \quad (4)$$

3. Equate the first-order and second-order d.b.h. moments to their expected values. That is

$$E(\text{dbh}) = \overline{\text{dbh}}_a = a + b \Gamma(1 + 1/c) \quad (5)$$

$$E(\text{dbh}^2) = \overline{\text{dbh}}_q^2 = a^2 + 2ab \Gamma(1 + 1/c) + b^2 \Gamma(1 + 2/c) \quad (6)$$

where

$E(\text{dbh}) = \overline{\text{dbh}}_a$ is the expected value of d.b.h., and the arithmetic mean d.b.h. (first-order moment of d.b.h.);
 $E(\text{dbh}^2) = \overline{\text{dbh}}_q^2$ is the expected value of d.b.h.², and the square of the quadratic mean d.b.h. (second-order moment of d.b.h.);

b = the scale parameter of the Weibull;

c = the shape parameter of the Weibull; and

Γ = the value of the gamma function with argument a .

4. Solve equation (5) for the b parameter and substitute the result for the b parameter in equation (6). This yields a nonlinear equation involving only the unknown parameter c and the known parameter a that is solvable for c . From equation (5),

$$b = \frac{\overline{dbh_a} - a}{\Gamma(1 + 1/c)} \quad (7)$$

On substitution of equation (7) for parameter b in equation (6), the equation involving only parameters a and c that is in a convenient, nonlinear equations solution form is:

$$f(c) = -a^2 + 2(a)\overline{dbh_a} + \left(\frac{\overline{dbh_a} - a}{\Gamma(1 + 1/c)} \right)^2 \Gamma(1 + 2/c) - \overline{dbh_q}^2 \quad (8)$$

5. Solve equation (8) for the value of c satisfying the condition $f(c) = 0$ using an algorithm, such as the bisection or secant methods, for finding the roots (zeros) of nonlinear equations (Burden and others 1981).

Weibull with Regression Predicted Parameters—Direct Prediction

To obtain regression equations to directly predict the parameters of the Weibull distribution, we fitted the cumulative distribution function of the Weibull to the sample plot empirical cumulative d.b.h. distributions using the SAS Gaussian NONLIN procedure (SAS Institute Inc. 1990). We then calculated regression equations to predict the estimated parameters from age, surviving trees per acre, and average height of dominant and codominant trees. Dell and others (1981) as well as other researchers, used the direct prediction method to build unthinned stand-yield models.

Methods of Model Testing and Comparison

A NN, as used in this modeling effort, is highly nonlinear; and no assumptions about the residual error structure were made. Because of this, the NN models are compared to the two Weibull models using mathematical measures of difference that resemble standard goodness-of-fit statistics, but which cannot be used in actual hypothesis testing because the probability distributions remain unknown. Simply, the numbers calculated are observed to be higher or lower than those of competing models. In these calculations, the degrees of freedom used is assumed to equal the number of observations, which is equal to the number of distributions compared multiplied by a fixed value of 20, the assumed number of classes in each distribution. Each of the test observations was estimated with the appropriate site-specific model using each of the three methods. Then these individual results were combined for an overall view of model quality.

The first set of comparators calculated is based on the goodness-of-fit statistics that are usually presented for fitted models. While none of these calculations adequately describes model quality, as a set they provide a comprehensive description. Included in this set are mean-squared error, fit index (Schlaegel 1981), generalized R^2 compared to the null model (Anderson-Sprecher 1994), a count of the number of times the model diameter class was predicted correctly, the number of times the probability of a tree falling in the model class was closest to the actual probability, the number of times the predicted probability of a tree falling in the upper or lower quartile was closest to the actual probability, the number of times the calculated x^2 was lowest, the number of times the calculated Kolmogorov-Smirnov statistic was lowest, the number of times the predicted arithmetic and quadratic means were closest to those of the actual distribution, and the number of times that the standard deviation of the predicted distribution was closest to that of the actual distribution. For the x^2 and Kolmogorov-Smirnov calculations, each check was done for the entire distribution and also for the upper and lower quartiles.

It is easy to pick one's favorite test statistic and then comparatively rank one model above or below another based on it. Unfortunately, this does not tell the whole story of how the model performs on various segments of the data. To better grasp this, we calculated three indicators of performance along the entire diameter distribution. In each of the following indicators, error is calculated as the predicted value minus the actual value (a positive error indicates overprediction). The first is average error for each diameter class. It is an indicator of bias. The second is the average absolute value of errors. It provides an indicator of the typical misfits of the model when positive and negative values are not allowed to cancel. Lastly, the maximum absolute value of an error allows one to see how poorly the model can perform in extreme cases.

Beyond these tests of prediction quality, we subjected each methodology to a sensitivity analysis. This was done by comparing the predicted probabilities for each diameter class to the same prediction when the input values for age, height of dominant trees > 1 in. in diameter, and number of trees per acre > 1 in. in diameter were each individually varied by a 10-percent increase and a 10-percent decrease. These differences were observed by diameter class. The effect this procedure had on basal area, arithmetic mean d.b.h., and quadratic mean d.b.h. was also examined for each site type and for all sites combined.

Results and Discussion

The results of the trained NN are contained in FORTRAN subroutines not shown in this paper because of their large size involving one parameter for every connection. There are 95 parameters for old-field sites, 120 parameters for cutover sites, and 170 parameters for prepared sites.

The Weibull distribution with parameters recovered from stand-distribution characteristics relies on prediction equations of various stand parameters. These prediction equations are presented in figure 3 for each of the three data sets.

The Weibull distribution with regression predicted parameters produces actual prediction equations for the

----- Old Field -----	
$\ln dbh_{min} = 3.7388 - 0.5645(\ln tpa) - \frac{1.7884}{\sqrt{h_d}} + 0.0461 age - \frac{1.3301}{age}$	$R^2 = 0.11 \quad s_{y.x} = 0.68$
$\ln \overline{dbh}_a = 3.3181 - 0.1540(\ln tpa) - \frac{4.0713}{\sqrt{h_d}} + 0.0097 age - \frac{2.5193}{age}$	$R^2 = 0.76 \quad s_{y.x} = 0.12$
$\ln \overline{dbh}_q = 3.3757 - 0.1502(\ln tpa) - \frac{4.1490}{\sqrt{h_d}} + 0.0081 age - \frac{2.7199}{age}$	$R^2 = 0.80 \quad s_{y.x} = 0.11$
----- Site Prepared -----	
$\ln dbh_{min} = 2.9072 - 0.2762(\ln tpa) - \frac{4.3556}{\sqrt{h_d}} - 0.0271 age - \frac{4.1527}{age}$	$R^2 = 0.19 \quad s_{y.x} = 0.40$
$\ln \overline{dbh}_a = 3.3685 - 0.1067(\ln tpa) - \frac{8.6348}{\sqrt{h_d}} - 0.0005 age + \frac{3.0011}{age}$	$R^2 = 0.92 \quad s_{y.x} = 0.09$
$\ln \overline{dbh}_q = 3.3636 - 0.1043(\ln tpa) + \frac{8.4831}{\sqrt{h_d}} + 0.0007 age + \frac{2.9436}{age}$	$R^2 = 0.92 \quad s_{y.x} = 0.09$
----- Cutover -----	
$\ln dbh_{min} = 3.5046 - 0.4352(\ln tpa) - \frac{7.6429}{\sqrt{h_d}} + 0.0171 age + \frac{3.8441}{age}$	$R^2 = 0.42 \quad s_{y.x} = 0.54$
$\ln \overline{dbh}_a = 3.3259 - 0.1558(\ln tpa) - \frac{8.8211}{\sqrt{h_d}} + 0.0165 age + \frac{6.1361}{age}$	$R^2 = 0.90 \quad s_{y.x} = 0.14$
$\ln \overline{dbh}_q = 3.1260 - 0.1259(\ln tpa) + \frac{8.3879}{\sqrt{h_d}} + 0.0175 age + \frac{5.6397}{age}$	$R^2 = 0.91 \quad s_{y.x} = 0.13$

Figure 3—Equation sets for predicting the minimum, arithmetic, and quadratic mean d.b.h. in inches from age, surviving trees per acre (tpa), and average height in feet of dominant and codominant trees (h_d) for the three longleaf pine plantation data sets.

----- Old Field -----
$a = 0.46, R^2 = 0.00, s_{y.x} = 0.45$
$b = 8.7143 + 0.2343 \text{ age} - 0.0289 \text{ age}(\ln \text{ tpa}) - \frac{26.8092}{\sqrt{h_d}}, R^2 = 0.56, s_{y.x} = 1.07$
$\ln c = 1.7780 - \frac{1.7370}{b}, R^2 = 0.24, s_{y.x} = 0.37$
----- Site Prepared -----
$a = 0.24, R^2 = 0.00, s_{y.x} = 0.53$
$b = 22.511 - \frac{127.6806}{\text{age}} - 0.293 \text{ age}(\ln \text{ tpa}) - \frac{70.228}{\sqrt{h_d}} + \frac{482.4427}{\text{age} \sqrt{h_d}}, R^2 = 0.70, s_{y.x} = 0.72$
$\ln c = 1.3805 - \frac{0.4578}{b}, R^2 = 0.15, s_{y.x} = 0.34$
----- Cutover -----
$a = 0.33, R^2 = 0.00, s_{y.x} = 0.86$
$b = 8.939 + 0.4232 \text{ age} - 0.0439 \text{ age}(\ln \text{ tpa}) - \frac{55.9923}{\sqrt{h_d}} + \frac{248.8361}{\text{age} \sqrt{h_d}}, R^2 = 0.73, s_{y.x} = 1.22$
$\ln c = 1.3879 - \frac{0.9166}{b}, R^2 = 0.17, s_{y.x} = 0.38$

Figure 4—Equation sets for predicting the a, b, and c parameters of the Weibull distribution from age, surviving trees per acre (tpa), and average height in feet of dominant and codominant trees (h_d) for the three longleaf pine plantation data sets.

Weibull parameters. Figure 4 presents the prediction equation sets for each of the three data sets.

The first comparison of note is a graphical examination of the performance of various diameter-distribution recovery methods. Figures 5 and 6 show the three predicted distributions as lines against a gray backdrop of actual example distribution. Figure 5 shows a unimodel distribution and simply illustrates that all of the examined methods produce a smooth curve that does a fair job of modeling the intended distribution. Figure 6 shows a multimodel distribution that no method fits very well, but it illustrates the flexibility of the NN to fit multimodel distributions. In each, the NN provided the best fit based on the χ^2 test statistic.

Mathematical measures of difference are shown in table 2. All of the models provided very good fits to the actual distributions, but the NN was best by almost all measures.

Figure 7 is a plot of the average error (or bias) by diameter class for each of the tested methodologies. All of the models are least accurate in the smaller diameter classes and most accurate in the larger. It can also be seen that the regression method has the greatest deviation and the NN the smallest.

Figure 8 shows the average absolute deviations that result from each of the three methods. When the positive and negative deviations are not allowed to cancel each other out, the average errors are much greater, but still the largest deviations occur in the smaller diameter classes where they have the least influence on volume and value. By this measure, all of the methods are about equal.

Both figures 7 and 8 show average performance of the presented methodologies. Figure 9 shows the worst-case performance of them. Once again, the lowest accuracy is in the smallest diameter classes and the methods are all about the same. The maximum deviations are strongly negative and

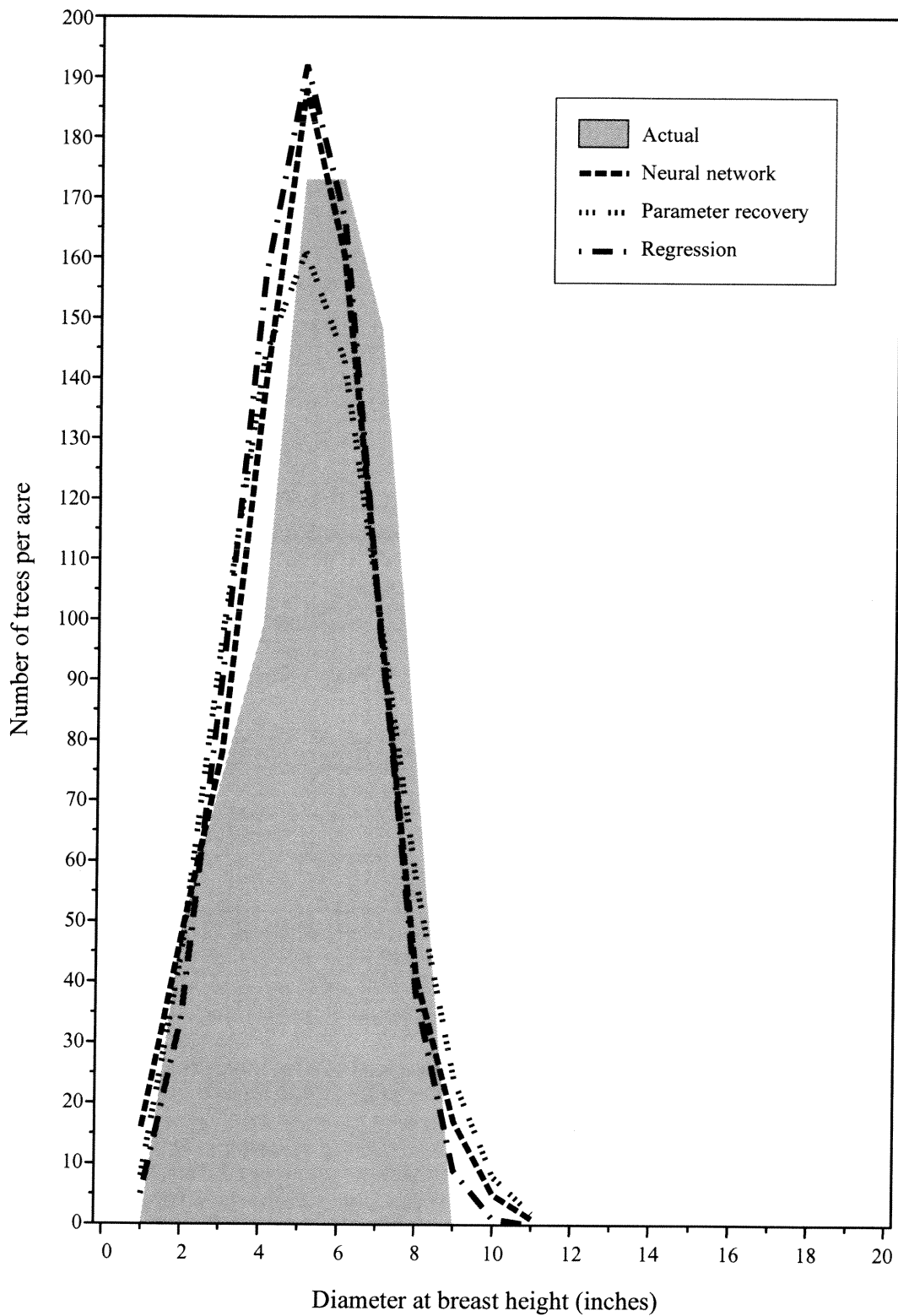


Figure 5—Three methods of predicting a diameter distribution compared to an actual unimodal distribution from a cutover site at age 21, with a height of dominant trees equal to 48 feet and 790 trees per acre.

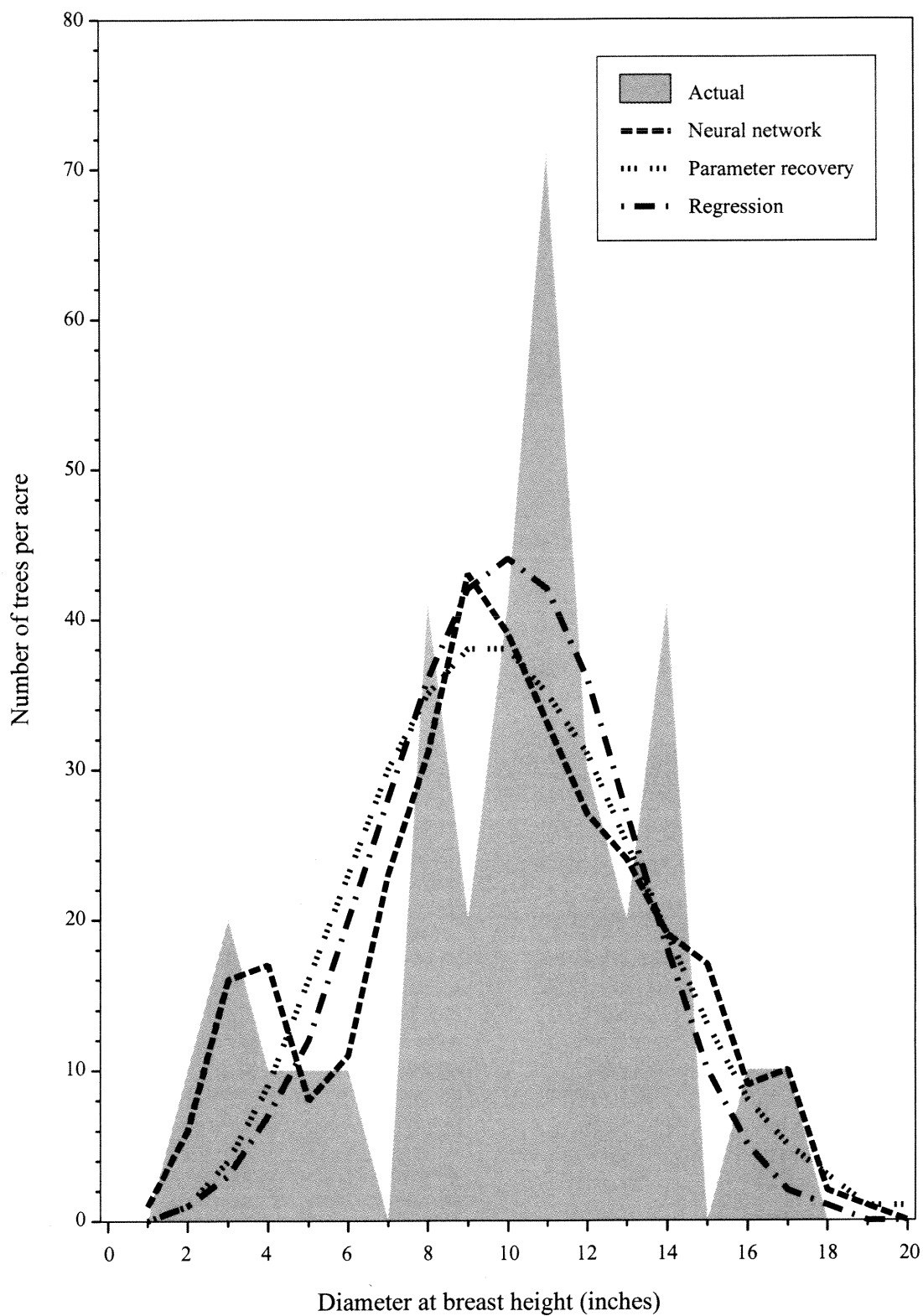


Figure 6—Three methods of predicting a diameter distribution compared to a complex multimodel distribution from a cutover site at age 43, with a height of dominant trees equal to 81 feet and 334 trees per acre.

Table 2—Some mathematical measures of difference for artificial neural networks and two methods of fitting Weibull distributions

Comparator	Neural network	Parameter recovery Weibull	Regression Weibull
Mean-squared error	0.002378	0.002592	0.002861
Fit index	.78	.76	.74
Generalized R ²	.82	.81	.79
Model class same as actual	12	2	2
Percentage of trees in model class closest to actual	216	190	157
Percentage of trees in lower quartile closest to actual	253	170	140
Percentage of trees in upper quartile closest to actual	202	195	139
Lowest x^2	259	136	168
Lowest x^2 in lower quartile	295	125	143
Lowest x^2 in upper quartile	287	202	74
Lowest Kolmogorov-Smirnov statistic	244	161	158
Lowest Kolmogorov-Smirnov statistic in lower quartile	257	165	141
Lowest Kolmogorov-Smirnov statistic in upper quartile	203	160	200
Arithmetic mean d.b.h. closest to actual	208	201	154
Quadratic mean d.b.h. closest to actual	187	199	177
Standard deviation of d.b.h. closest to actual	141	171	251

disturbingly large. The strong negative bias shows that one large error encountered is the presence of trees in a diameter class when none is predicted. One cause of this is that, in general, smaller diameter trees grow or die out over time, but there are several cases where a small diameter tree will stagnate and neither grow nor die. The ability to predict this phenomenon would greatly enhance our ability to model the diameter distributions of longleaf pine stands.

Our results show that NN can perform at least as well as traditional methods and often better. They may be reducible to nonlinear models (Sarle 1994), however unwieldy, but their strength is that the model form does not have to be specified in advance. This is a great advantage because, in spite of many efforts in process modeling, we still do not understand the processes of growth that would allow us to create models not tied to empirical data. Even the Weibull function has no biological meaning. It is simply a mathematically handy function with the ability to assume a variety of appropriate shapes (Weibull 1951).

There are three principal weaknesses in NN as used in this paper. The first is that although a model does not have to be specified, the number of hidden nodes and layers and the transfer function to be used still must be determined. A correct choice of options can make a great difference in a modeling effort's success. Also, there are many rules of

thumb for selecting variables, but it really comes down to a matter of trial and error.

The second limitation is that one cannot count on some of the desired constraints that can easily be imposed by mathematical model selection. Constrained least squares d.b.h. moment-recovery algorithms such as those presented in Matney and others (1990) and in Matney and Belli (1995) provide for the calculation of logically constrained residual and future d.b.h. distributions. On the other hand, natural constraints built into the data sets may allow for development of NN models that are almost logically constrained without imposing formal constraints. For many data sets, the strength of the data should generate NN solutions that are naturally constrained. In general, an unconstrained solution to a problem will have lower mean-square error than a constrained solution to the same problem. In linear space, a constrained model with one or more parameter constraints imposed always has a higher mean-square error than the same model with no parameter constraints. Hence, unnecessary constraints should not be imposed on a solution to a problem. However, NN are nonlinear and the theory for linear systems is not directly applicable. It is possible to have a constrained nonlinear estimator with lower mean-square error than the same model without parameter constraints.

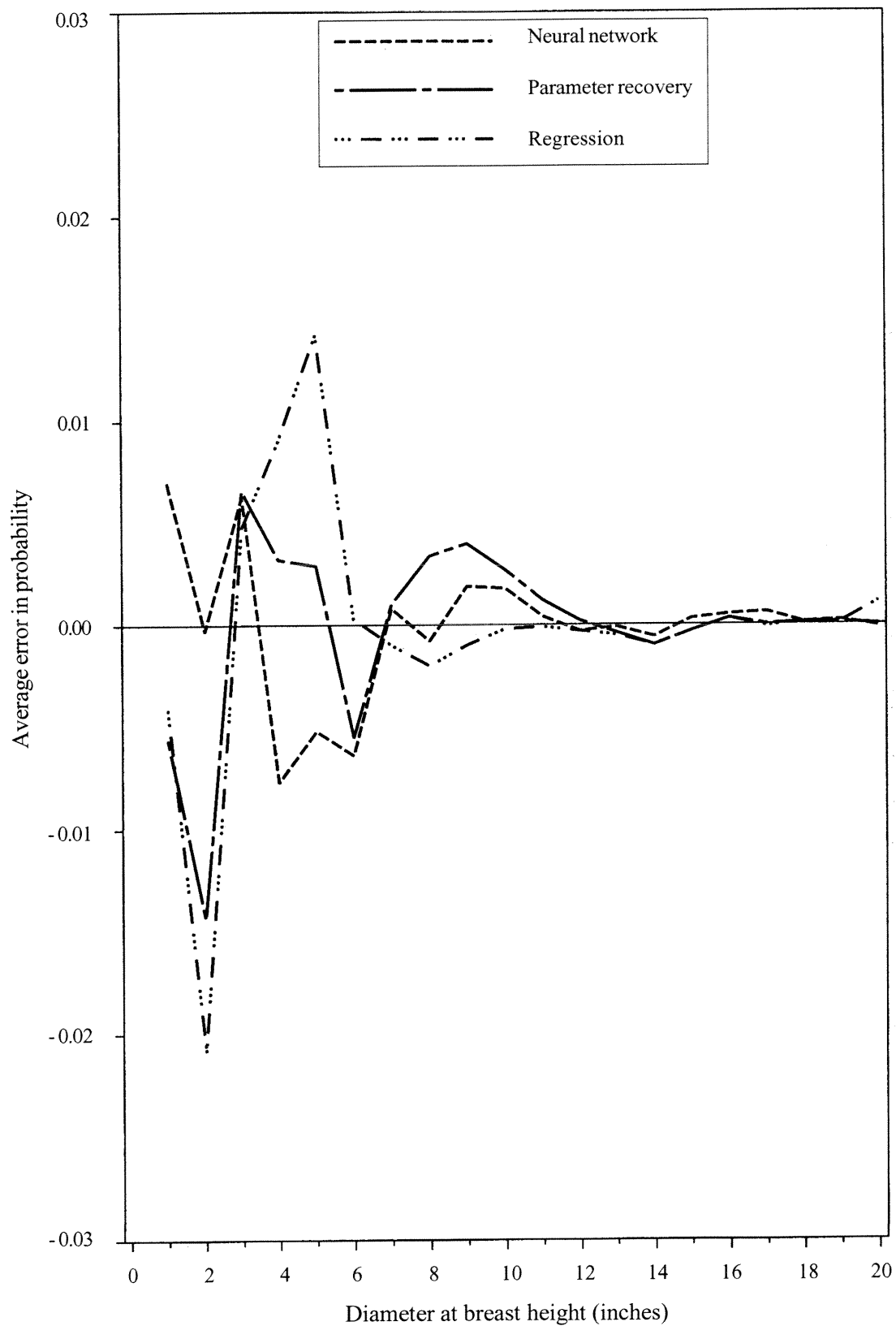


Figure 7—Average error in probability estimates (predicted to actual) in each diameter class for each of three methods of estimating diameter distributions.

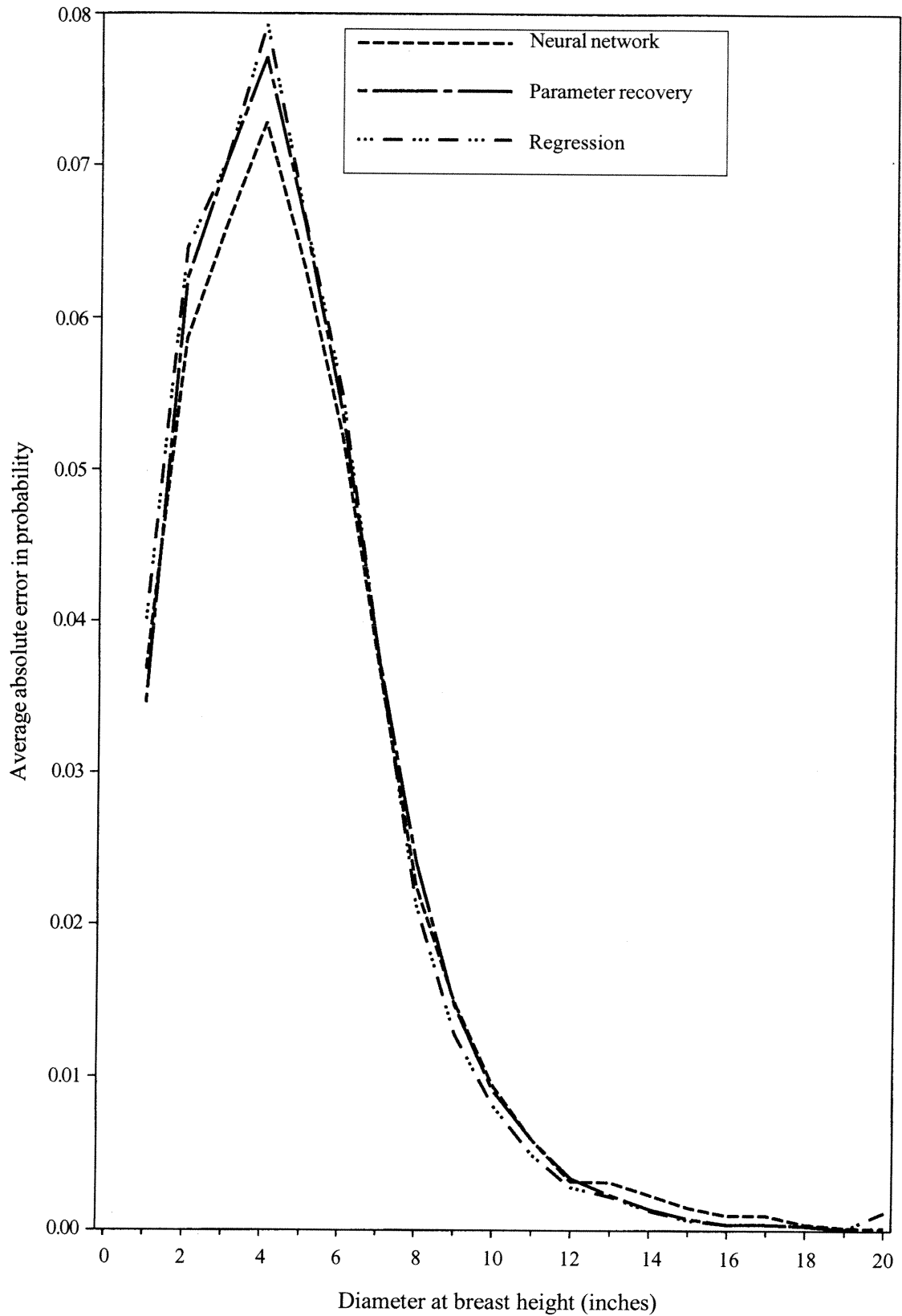


Figure 8—Average absolute value of the error in probability estimates in each diameter class for each of three methods of estimating diameter distributions.

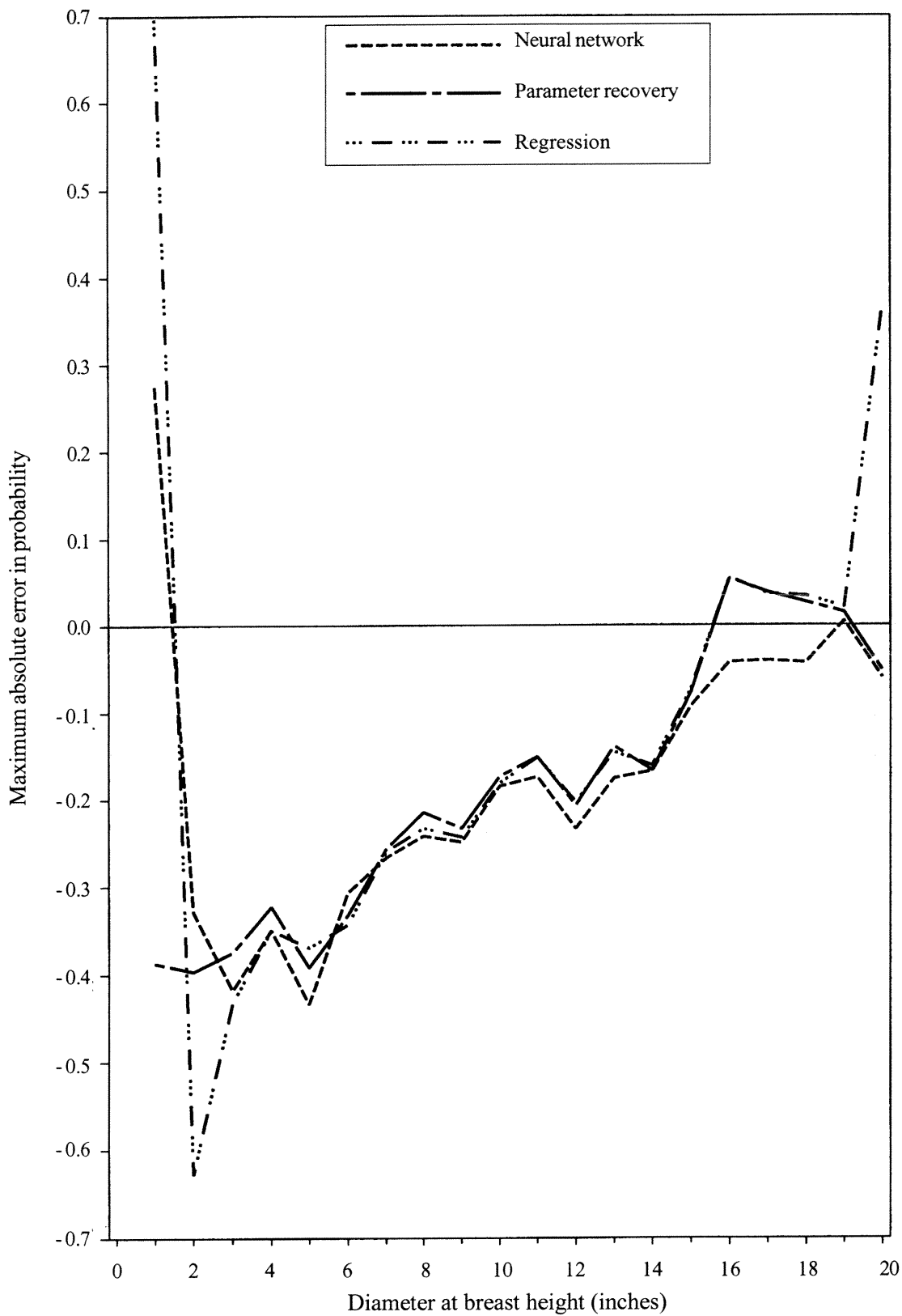


Figure 9—Maximum absolute value of error in probability estimates (predicted to actual) in each diameter class for each of three methods of estimating diameter distributions. Absolute values were used to determine magnitude, but signs were kept to indicate direction.

If constraints are necessary, they can be difficult to apply to the output of a NN model when using commercial software. Users of growth-and-yield models require that the residual stand after-thinning and before-thinning diameter distributions be logical in relation to each other. Likewise, they insist that projected diameter distributions be logically related to the initial diameter distribution. For thinned stands this means, for example, that the estimated diameter distribution should not have more trees in a diameter class after thinning than before thinning, and that the after-thinning stand should have the desired basal area or trees per acre. For growth projections, users expect the future distribution to shift to the right and have a nondecreasing mean diameter. The NN methods already are strictly empirical; and achieving these kinds of constraints is easy if one is willing to use additional programming in computer models when necessary, or to combine neural methodologies with other, more traditional statistical approaches.

The third limitation is that the network that results from a training program appears to be a black box. You can see the inputs and the outputs, but the process of moving from one to the other is not obvious. In this case, it is implemented as a C or FORTRAN program, but an examination of this program just reveals so many interacting equations that clear relationships are hard to determine. Fortunately, this latter

difficulty can be easily addressed through the use of sensitivity analysis (Klimasauskas 1991).

We conducted a sensitivity analysis on the input parameters of age, height of dominant trees, and number of trees per acre to see what effect changes in these parameters would have upon stand basal area, arithmetic mean d.b.h., and quadratic mean d.b.h. These results for all site types combined are shown in table 3. This table shows that all of the methods are very similar in their sensitivity to input, with only the age influence on basal area showing a change that exceeded the change in input. Also it shows that trees-per-acre has a moderate influence, and the height of dominant trees has almost none. These results are rather similar across all three site types (not expanded here) with old-field sites being the least sensitive and prepared sites the most.

Conclusion

Artificial NN are an excellent alternative to the traditional method of predicting unthinned stand-diameter distributions with estimated Weibull probability functions. The superiority of the NN arises because the Weibull is not the correct model for the data in this study. On other data sets, when the Weibull distribution fits better, either the NN or Weibull

Table 3—Percentage deviations in some key outputs caused by 10-percent changes in the input variables for each of the three examined models

Result changed	Age		Height of dominant trees		Trees per acre	
	Increase	Decrease	Increase	Decrease	Increase	Decrease
Artificial neural network						
Basal area	13.8	-13.7	- 0.6	1.9	7.5	- 7.7
Arithmetic mean d.b.h.	7.0	- 7.4	- .6	1.0	8.8	- 9.0
Quadratic mean d.b.h.	6.7	- 7.1	- .4	.8	3.7	- 3.9
Parameter recovery Weibull						
Basal area	13.5	-13.5	.9	.2	7.4	- 7.6
Arithmetic mean d.b.h.	6.8	- 7.2	.1	.4	8.5	- 8.6
Quadratic mean d.b.h.	6.5	- 7.0	.4	0	3.6	- 3.9
Regression Weibull						
Basal area	14.9	6.3	8.4	5.2	7.5	- 7.6
Arithmetic mean d.b.h.	7.1	- 6.3	- .3	2.2	8.7	- 8.8
Quadratic mean d.b.h.	7.0	- 5.3	- .3	2.1	3.7	- 3.9

probability distribution approaches will perform well. The clear advantage of the NN over the parametric function modeling techniques is that in almost all cases a NN solution will minimize the RMS error. If the Weibull distribution or other assumed probability function do not fit the data, the modeler is left with the very difficult task of piecing together and/or finding a new parametric model form for the problem.

The next evolutionary step for applying NN in growth-and-yield modeling is to find network models for estimating diameter distributions after thinning and for estimating future diameter distributions using the current distribution.

Feedforward networks, as used in this paper, are just one of many kinds of NN; and investigation of other forms may lead to even better results. The NN theory is a rapidly expanding arena of study that will find application in many fields. The field of forestry is an area ripe with potential for NN exploration.

Literature Cited

- Anderson-Sprecher, Richard.** 1994. Model comparisons and R^2 . *The American Statistician*. 48(2): 113–117.
- Bailey, David; Thompson, Donna.** 1990. How to develop neural-network applications. *AI Expert*. 5(6): 38–47.
- Bailey, Robert L.; Aleixo da Silva, J.A.** 1988. Compatible models for survival, basal area growth, and diameter distributions of fertilized slash pine plantations. In: Ek, Alan R.; Shifley, Stephen R.; Burk, Thomas E., eds. *Forest growth modeling and prediction: Proceedings, International Union of Forestry Research Organizations conference; 1987 August 23–27; Minneapolis*. Gen. Tech. Rep. NC-120. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station: 538–546.
- Bailey, Robert L.; Dell, T.R.** 1973. Quantifying diameter distributions with the Weibull function. *Forest Science*. 19(2): 97–104.
- Burden, Richard L.; Faires, J. Douglas; Reynolds, Albert C.** 1981. *Numerical analysis*. 2^d ed. Boston: Prindle, Weber, and Schmidt. 598 p.
- Cybenko, G.** 1989. Approximation by superpositions of a sigmoidal function. *Mathematical Control Systems*. 2: 303–314.
- Dell, T.R.; Feduccia, D.P.; Campbell, T.E. [and others].** 1981. Yields of unthinned slash pine plantations on cutover sites in the west gulf region. Res. Pap. SO-147. New Orleans: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 84 p.
- Farrar, Robert M., Jr.; Matney, Thomas G.** 1994. A dual growth simulator for natural even-aged stands of longleaf pine in the South's east gulf region. *Southern Journal of Applied Forestry*. 18(4): 147–155.
- Freeman, James A.** 1992. Backpropagation in a neural network. In: *Neural network special report*. San Francisco: Miller Freeman Publications: 55–63. [Supplement to *AI Expert*].
- Guan, Bing T.; Gertner, George.** 1991. Modeling red pine tree survival with an artificial neural network. *Forest Science*. 37(5): 1429–1440.
- Hassoun, M.H.** 1995. *Fundamentals of artificial neural networks*. Cambridge, MA: The MIT Press. 511 p.
- Hursch, Robert P.** 1991. Validation samples. *Biometrics*. 47: 1193–1194.
- Josin, Gary.** 1987. Neural network heuristics. *Byte*. 12(11): 183–192.
- Klimasauskas, Casimir C.** 1991. Neural nets tell why. *Dr. Dobbs's Journal*. 16: 16–24.
- Kolmogorov, A.N.** 1957. On the representation of continuous functions of several variables by super-position of continuous functions of one variable and addition. *USSR: Doklady Akademii Nauk*. 114: 679–681.
- Leahy, Kent.** 1994. The overfitting problem in perspective. *AI Expert*. 9(10): 35–36.
- Light, W.A.** 1992. Ridge functions, sigmoidal functions, and neural networks. In: Cheney, E.W.; Chui, C.K.; Schumaker, L.L., eds. *Approximation theory VII*. New York: Academic Press: 163–206.
- Matney, T.G.; Belli, K.L.; Farrar, R.M.** 1990. Parameter-free diameter distribution recovery. In: *Research in forest mensuration, growth and yield: Proceedings of S4.01 mensuration, growth and yield at the World Congress of International Union of Forestry Research Organizations; 1990 August 5–11; Montreal, Canada*. Publ. FWS-2–90. Blacksburg, VA: Virginia Polytechnic Institute and State University, School of Forestry and Wildlife Resources: 95–107.
- Matney, T.G.; Sullivan, A.D.** 1982a. Approximating thinned stand-diameter distribution with statistical probability functions. In: Jones, Earle P., Jr., ed. *Proceedings of the second biennial southern silvicultural research conference; 1982 November 4–5; Atlanta*. Gen. Tech. Rep. SE-24. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station: 315–324.
- Matney, Thomas G.; Belli, Keith L.** 1995. A weighted least squares diameter moment recovery system for cut over site prepared loblolly pine plantations. In: Edwards, M. Boyd, comp. *Proceedings of the eighth biennial southern silvicultural research conference; 1994 November 1–3; Auburn, AL*. Gen. Tech. Rep. SRS-1. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 182–191.
- Matney, Thomas G.; Farrar, Robert M., Jr.** 1992. A thinned/unthinned loblolly pine growth and yield simulator for planted cutover site-prepared land in the mid-gulf south. *Southern Journal of Applied Forestry*. 16(2): 70–75.
- Matney, Thomas G.; Sullivan, Alfred D.** 1982b. Compatible stand and stock tables for thinned loblolly pine stands. *Forest Science*. 28(1): 161–171.
- NeuralWare.** 1991a. *Neural computing: neuralworks professional II/plus and neuralworks explorer*. Pittsburgh: NeuralWare Inc. 356 p.

- NeuralWare.** 1991b. An overview of neural computing. Pittsburgh: NeuralWare Inc. [Diskette]. [Software].
- NeuralWare.** 1991c. Reference guide: neuralworks professional II/plus and neuralworks explorer. Pittsburgh: NeuralWare Inc. 352 p.
- NeuralWare.** 1991d. Using Nworks: neuralworks professional II/plus and neuralworks explorer. Pittsburgh: NeuralWare Inc. 138 p.
- Samad, Tariq.** 1988. Backpropagation is significantly faster if the expected value of the source unit is used for update. In: International neural network society conference abstracts; [Date of meeting unknown]; [Location of meeting unknown]. [Place of publication unknown]: [Publisher unknown]: 216.
- Sarle, Warren S.** 1994. Neural networks and statistical models. In: Proceedings of the nineteenth neural annual SAS users group international conference; 1994 April 10-13; Dallas. Cary, NC: SAS Institute: [Number of pages unknown].
- SAS Institute Inc.** 1990. SAS/STAT user's guide. Version 6. 4th ed. Cary, NC: SAS Institute. 846 p. Vol. 2.
- Schlaegel, Bryce E.** 1981. Testing, reporting, and using biomass estimation models. In: Gresham, Charles A., ed. Proceedings of the 1981 southern forest biomass workshop; 1981 June 11-12; Georgetown, SC. Clemson, SC: The Belle W. Baruch Forest Science Institute: 95-112.
- Sun, X.; Cheney, E.W.** 1992. The fundamentals of sets of ridge functions. *Aequationes Mathematicae*. 44: 226-235.
- Weibull, W.** 1951. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*. 18: 293-297.
- Zarnoch, Stanley J.; Feduccia, Donald P.; Baldwin, Virgil C.; Dell, Tommy R.** 1991. Growth and yield model predictions for thinned and unthinned slash pine plantations on cutover sites in the west gulf region. Res. Pap. SO-264. New Orleans: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 32 p.

Leduc, Daniel J.; Matney, Thomas G.; Belli, Keith L.; Baldwin, V. Clark, Jr. 2001.

Predicting diameter distributions of longleaf pine plantations: a comparison between artificial neural networks and other accepted methodologies. Res. Pap. SRS-25. Asheville, NC: U.S.

Department of Agriculture, Forest Service, Southern Research Station. 18 p.

Artificial neural networks (NN) are becoming a popular estimation tool. Because they require no assumptions about the form of a fitting function, they can free the modeler from reliance on parametric approximating functions that may or may not satisfactorily fit the observed data. To date there have been few applications in forestry science, but as better NN software and fitting algorithms become available, they may be used to solve a wide variety of problems—particularly problems where the underlying relationship between predicted and predictors is unknown. We benchmark tested an alternative to the traditional Weibull probability distribution function, diameter-at-breast-height moment, and direct parameter prediction models for approximating stand-diameter distributions. Using a feedforward backpropagation network, we demonstrated that NN are a somewhat better option. Unlike Weibull approximations, NN solutions cannot easily be mathematically constrained to match known reality constraints, but this difficulty is easy to overcome in practice.

Keywords: Connectionist models, parallel distributed processing systems, parameter recovery, Weibull distribution.



The Forest Service, United States Department of Agriculture (USDA), is dedicated to the principle of multiple use management of the Nation's forest resources for sustained yields of wood, water, forage, wildlife, and recreation. Through forestry research, cooperation with the States and private forest owners, and management of the National Forests and National Grasslands, it strives—as directed by Congress—to provide increasingly greater service to a growing Nation.

The USDA prohibits discrimination in all its programs and activities on the basis of race, color, national origin, sex, religion, age, disability, political beliefs, sexual orientation, or marital or family status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD).

To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, Room 326-W, Whitten Building, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410 or call (202) 720-5964 (voice and TDD). USDA is an equal opportunity provider and employer.