# A NEURAL NETWORK MODEL FOR WOOD CHIP THICKNESS DISTRIBUTIONS

*Emily B. Schultz*
Associate Professor

*Thomas G. Matney*
Professor

Department of Forestry
Forest and Wildlife Research Center
Mississippi State University
Box 9681
Mississippi State, MS 39762

and

*Jerry L. Koger*
President
Wellrock Company
585 Combs Lane
Somerset, KY 42501

(Received November 1997)

ABSTRACT

Wood chip thickness is an important factor in pulp quality and yield. An artificial neural network model was developed and incorporated into a growth and yield simulator to predict wood chip thickness distributions from stand and tree characteristics. Models based on direct parameter estimation and parameter recovery were also developed for comparison to the neural network. Data were derived from 11,771 individual loblolly pine chip thickness measurements. Four stand ages, five dbh (diameter at breast height) classes, and three stem positions were used to predict the cumulative proportion of chip weight per chip thickness class. Results showed that the neural network model was superior to the two deterministic models on the basis of bias, root mean square error, and index of fit. Sensitivity analyses for the neural network model demonstrated that thicker chips were produced by younger stands and lower stem positions. The neural network was combined with a growth and yield simulator to demonstrate its use as a tool for procurement foresters and mill managers in predicting yields from stands of given characteristics.

*Keywords:* Neural network, wood chips, wood chip thickness, wood chip thickness distributions.

## INTRODUCTION

Wood chip thickness is a major factor in the performance of pulp digesters and in subsequent pulp quality and yield (Borlew and Miller 1970; Becker 1992; Tikka et al. 1993). Different pulping methods require different chip thicknesses (Dubois et al. 1991; Wood and Gosda 1992). Chip thickness below or above an optimum range produce either overcooked or undercooked pulp, thereby reducing process and fiber efficiencies (Worster et al. 1977; Christie 1987). The control of chip thickness distributions readily translates into increased yield per unit cost and has been studied from four different approaches or production phases: 1) prechipping wood conditions (Flowers et al. 1992; Wallace et al. 1992; Koger et al. 1993); 2) mechanical considerations during chipping (Twaddle and Watson 1990; Dubois et al. 1991; Uelmen 1993); 3) screening of chips prior to pulping (Christie 1987;

*Wood and Fiber Science*, 31(1), 1999, pp. 2–14

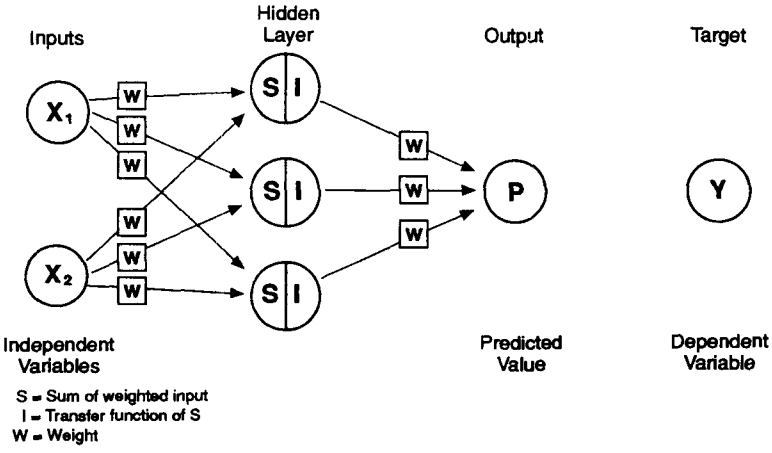


FIG. 1. A simple multilayer neural network.

Tikka et al. 1993); and 4) pulping conditions (Worster et al. 1977; Becker 1992).

Prediction of chip thickness distributions from prechipping wood conditions has several advantages. It can change the magnitude or type of control needed in later production phases, and it allows the purchase price of roundwood to be based on projected yields. Prechipping studies by Flowers et al. (1992) and Koger et al. (1993) showed that stand age, dbh (diameter at breast height) class, and chip position in the stem influenced the distribution of chip weights by thickness class. Given adequate prediction equations, these variables could be readily manipulated to increase pulp yields.

The objective of our study was the enhancement of modeling techniques for predicting chip thickness distributions from prechipping stand and tree characteristics. We developed and compared three models. The best model was integrated with a growth and yield simulator for the purpose of predicting total chip weight by thickness class for stands of a given age, site index, and merchandizing standards. The utility of the integrated model for the procurement forester and mill manager is twofold: 1) the prediction of chip yields from stands under consideration for purchase or scheduled for harvest; and 2) the prediction of yields from the manipulation of management, harvesting, and chipping strategies. The objective of manipulating chipping strategies could be either optimum chip thickness or optimum chip thickness mixes.

## MODELING APPROACHES

We selected two parametric methods and one nonparametric method for predicting chip thickness distributions. The parametric methods, based on the Weibull distribution function, were direct parameter prediction and parameter recovery. The nonparametric approach was a neural network. Each approach and its advantages are discussed below. Bias, precision, and index of fit statistics were used to compare the three models and to select the best one for integration with the growth and yield simulator.

### *Neural networks*

The use of neural networks, formally called *artificial neural networks,* is a technique from the field of artificial intelligence that attempts to simulate human cognitive behavior. There are numerous neural network types, but perhaps the most popular is the back-propagation network. A multilayered, back-propagation network (Fig. 1) is composed of an input layer, an output layer, and one or more middle layers called hidden layers. Each layer contains a

number of nodes that hold and transmit calculated values. The input layer contains one node for each independent variable in the model, and the output layer contains one node for each dependent variable in the model. Trial and error testing during network construction determines the number of hidden layers and number of nodes per hidden layer. Values travel in one direction along connecting links, from nodes in the input layer to nodes in the output layer. Each link is associated with an iteratively calculated weight.

There are two phases of network development, a training phase and a testing phase. Learning (weight adjustment) and model building occur in the training phase. Evaluation of models occurs in the testing phase. During training, the nodes of the input layer receive scaled data values from the independent variables. Each input node value is multiplied by the corresponding weights of its links, and the products are transmitted to connecting nodes in the first hidden layer. The weighted values from each input node are summed, transformed by a smoothing or transfer function, and stored in a hidden layer node. The weighting, summation, and transformation process is repeated from layer to layer until the output layer is reached. Node values are descaled at the output layer.

The difference between the output node values (predicted output) and their paired target values (observed output) is used to adjust the network weights by propagating errors back through the network according to a learning rule. Weights are adjusted in proportion to the product of a learning rate, an error derivative, and the output from the previous layer (NeuralWare 1993; Weiss and Kulikowski 1991). The neural network produces a nonlinear model whose parameters, the network weights, are adjusted after each input or epoch (specified number of inputs).

During the testing phase, values travel through the network in the same manner as in the training phase, except there is no updating of weights. Network inputs either come from random subsets of the training data or from an independent data set. As predicted output values are produced, they are compared to corresponding observed outputs to calculate a collective error. The best of many models is usually chosen on the basis of the lowest root mean square error.

There are several general advantages of neural networks as compared to traditional modeling techniques. Neural networks assume no predetermined functional form, and thus no prior knowledge of the model is needed. They are particularly well adapted to applications where no suitable mathematical model is known or where systems may be composed of complicated interactions.

A noted disadvantage of neural networks is the inability to place logical constraints on the output. Another disadvantage can be the magnitude of data needed during the training phase. Problems that produce small data sets are not well suited to modeling with neural networks.

### *Direct parameter prediction*

The direct parameter prediction approach involves identifying a parametric probability function that closely approximates the observed distribution. The parameters of the selected distribution are estimated by maximum likelihood estimation or by non-linear least squares fittings of the cumulative distribution function to the empirical distribution functions. Parameter estimates are then regressed on the predictor variables to build a distribution prediction model (Bailey and Dell 1973; Dell et al. 1981; Kendal and Stuart 1977).

### *Parameter recovery*

The parameter recovery method for estimating parametric distributions is an indirect technique. First, regression equations are found for predicting selected moments (mean or quadratic mean) and/or order statistics from the predictor variables associated with each distribution. The expected value equation derived for each predicted moment or order statistic is equated to its corresponding regression

equation, and the subsequent linear/nonlinear system of equations is then solved for the desired parameter estimates (Farrar and Matney 1994; Kendal and Stuart 1977; Matney and Farrar 1992; Matney and Sullivan 1982a, b; Zarnoch et al. 1991).

Unlike the neural network, both parametric approaches are bound to a particular functional form and automatically impose logical constraints. These approaches can yield excellent predictions if the assumed distribution closely approximates the actual distribution. On the other hand, when the actual distribution is complex, multi-modal, or discontinuous, these approaches are of limited value.

## DATA

The data consisted of measurements taken on 11,771 individual loblolly pine wood chips. Green Bay Packaging Company's plantations near Morrilton, Arkansas, supplied the trees. Three to ten trees were selected from within each of four stand ages (14, 19, 23, and 29 years) and five diameter classes (5, 7, 9, 11, and 13 inches dbh). Debarked tree length stems were individually identified and chipped in diameter-age groups. Price Industries in Perry, Arkansas, chipped the trees. Chip samples were taken from the butt (1), middle (2), and top (3) thirds of each group of stems. Details on the chipper set-up are found in a previous study by Koger et al. (1993). Chips were classified with a Gradex classifier to determine the percentage of fines, pins, accepts, and overs. Individual chip thickness measurements were made with an electronic caliper, and each chip was weighed to the nearest hundredth of a gram. Chips that were less than 2 millimeters in thickness were not individually measured but were collectively weighted by stand age-diameter-stem position class. The average chip thickness was 3.88 millimeters, and the average chip weight was 0.56 grams (Koger et al. 1993).

The independent variables for the three models were stand age, dbh class, chip position in the stem, and chip thickness. There were a total of 54 bulk samples representing unique combinations of variable levels. The dependent variable was defined as the cumulative proportion of chip weight that is less than or equal to each unique chip thickness. The cumulative proportion, though not a relative frequency probability distribution, does share the same properties of probability distributions, and these properties were important in the development of the parametric approaches. Instead of predicting the relative numbers of chips of a specified thickness, the models predict the cumulative proportion of the total weight of chips having a specified thickness.

## METHODS

### *Neural network*

We used NeuralWare's Neural Works Professional II/Plus (NeuralWare, Inc., Pittsburgh, PA) software to construct the neural network model. The software ran on a SUN 690 MP minicomputer. Many network types are available through the NeuralWorks software, but we selected the commonly used fully connected, hetero-associative, feed forward, back-propagation form. Feed forward networks (like the back-propagation learning system) have been mathematically proven to be capable of approximating continuous functions to any degree of accuracy (Hassoun 1995).

Building a back-propagation network with NeuralWorks requires the selection of various parameters. We based some parameter selections on NeuralWare's recommendation and others on trial and error experimentation and minimum root mean square error. Table 1 summarizes the selected parameters. Different values of the parameters in Table 1 constitute separate networks that were tested. We tried network architectures with one, two, and three hidden layers and with a varying number of nodes per layer. An architecture of two hidden layers with eight nodes in the first hidden layer and four nodes in the second hidden layer was chosen because it produced the least root mean square error of the models tested.

TABLE 1. *Selected NeuralWorks Professional II/Plus network parameters.*

| Parameter | Value | Description |
|---|---|---|
| Network type | back-propagation | back-propagation of errors |
| | hetero-associative | different input and output variables |
| | min-max table | scaled inputs |
| | fully connected | all nodes connected in adjacent layers |
| Learning rule | delta rule | governs weight adjustments |
| Transfer function | sigmoid | smoothing function |
| Epoch size | 1 = standard for delta rule | no. inputs per weight update |
| No. hidden layers (HL) | 2 | |
| No. nodes/hidden layer | HL1 = 8, HL2 = 4 | |
| Momentum | 0.8 | modifies weights to deter convergent behavior |
| Learn counts (in thousands) | HL1 = 10, 30, 70, 150, 310 | the sequential number of inputs for which a learning coefficient applies |
| | HL2 = 10, 30, 70, 150, 310 | |
| | Output = 10, 30, 70, 150, 310 | |
| Learning coefficients | HL1: 0.9, 0.45, 0.225, 0.1125, 0.00001 | multipliers in the calculation of weights; values change after a set number of inputs (learn count) |
| | HL2: 0.6, 0.3, 0.15, 0.07, 0.00001 | |
| | Output: 0.15, 0.075, 0.01875, 0.00117, 0.00 | |

A network stopping criterion of 300,000 network cycles was determined during preliminary testing. One chip observation is processed by one network cycle. During preliminary tests, the neural networks were allowed to run with no stopping criterion until improvements in error were no longer made. This always occurred before 300,000 network cycles.

NeuralWorks randomly presented samples of the chip measurements to the networks for training. Sampling occurred without replacement until all 11,771 observations were presented at least once. Random sequences of the 11,771 observations were presented repeatedly until the stopping criterion was reached. Sample sizes consisted of approximately 17% of the chip measurements. Different sample sizes were tested, but those larger than 17% showed no improvement in root mean square error.

One hundred and forty-eight models were generated for each network architecture. The model with the least root mean square error (Fig. 2) was selected and output through a NeuralWorks utility in standard C language source code. Bias, root mean square error (RMSE), and index of fit were calculated for comparison with the two parametric models (Table 2). Bias was calculated as the average difference between observed and predicted values, and index of fit was calculated as one minus the quantity of the error sum of squares divided by the total sum of squares.

### *Direct parameter prediction*

Previous work by Koger (1994) demonstrated that extremely close approximations to the normalized weighted chip distributions could be obtained by nonlinear least squares fitting of the cumulative distribution of a three-parameter Weibull function. The Weibull cumulative distribution function was thus accepted as the appropriate model for predicting cumulative chip weight distributions. The Weibull distribution can assume many shapes, and because of this flexibility has been widely applied in forestry and other fields to model distributions. The following cumulative Weibull distribution function equation was employed in an SAS (SAS Institute Inc., Cary, NC) Guass-Newton *NONLIN* procedure to produce parameter estimates.

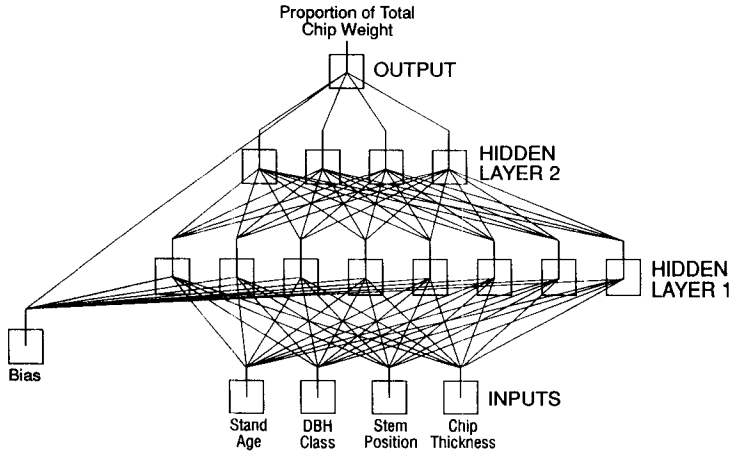


FIG. 2. Neural network architecture that possessed the least root mean square error.

$$F(t) = 1 - e^{-[(t-a)/b]^c} \quad (1)$$

where

F(t) = weight proportion of chips less than or equal to a thickness of t,

t = chip thickness in millimeters,

a = location parameter (the minimum chip thickness),

b = scale parameter,

c = shape parameter, and

e = base of the natural (Naperian) logarithm.

All a parameter estimates obtained the logical lower bound of 0. As a result, we elected to use the following two-parameter Weibull equation for both the direct and recovery modeling approaches.

$$F(t) = 1 - e^{-[t/b]^c} \quad (2)$$

After estimates of the b and c parameters were obtained, a regression analysis was performed to find the *best* equations to predict the model parameters from the 54 combinations of stand age (age), tree dbh class (dbh), and stem position (pos). The best parameter prediction equations developed were:

TABLE 2. *Bias, root mean square error, and index of fit by chip thickness class for the three modeling approaches.*

| Chip thickness class (mm) | No. obs. | Neural network | | Parameter recovery | | Direct parameter | |
|---|---|---|---|---|---|---|---|
| | | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| 0 | 8 | 0.002 | 0.004 | −0.001 | 0.007 | 0.002 | 0.004 |
| 1 | 389 | −0.001 | 0.011 | −0.032 | 0.037 | −0.012 | 0.015 |
| 2 | 3,184 | −0.002 | 0.026 | −0.043 | 0.052 | −0.003 | 0.030 |
| 3 | 3,039 | 0.005 | 0.056 | −0.010 | 0.060 | 0.024 | 0.069 |
| 4 | 2,393 | 0.000 | 0.072 | 0.038 | 0.088 | 0.030 | 0.088 |
| 5 | 1,467 | 0.000 | 0.077 | 0.064 | 0.108 | 0.008 | 0.089 |
| 6 | 688 | −0.009 | 0.071 | 0.042 | 0.088 | −0.038 | 0.087 |
| 7 | 356 | −0.005 | 0.063 | 0.012 | 0.069 | −0.062 | 0.091 |
| 8 | 160 | −0.004 | 0.057 | −0.013 | 0.063 | −0.061 | 0.085 |
| 9 | 47 | −0.002 | 0.042 | −0.023 | 0.049 | −0.050 | 0.065 |
| 10 | 40 | 0.000 | 0.000 | 0.014 | 0.016 | 0.001 | 0.002 |
| All | 11,771 | 0.000 | 0.057 | 0.003 | 0.073 | 0.007 | 0.070 |
| Index of Fit | | 0.957 | | 0.929 | | 0.936 | |