ORIGINAL ARTICLE

Artificial neural network modeling for predicting final moisture content of individual Sugi (*Cryptomeria japonica*) samples during air-drying

**Ken Watanabe · Yasuhiro Matsushita ·
Isao Kobayashi · Naohiro Kuroda**

Abstract Sugi (*Cryptomeria japonica* D. Don) lumber is known to have a large variability in final moisture content (MC_f) and is difficult to dry. This study assessed the capability of artificial neural networks (ANNs) to predict the MC_f of individual wood samples. An ANN model was developed based on initial moisture content, basic density, annual ring orientation, annual ring width, heartwood ratio and lightness (L^* in the CIE $L^*a^*b^*$ system). The performance of the ANN model was compared with a principal component regression (PCR) model. The ANN model showed good agreement with the experimentally measured MC_f with a higher correlation coefficient (r) and a lower root mean square error (RMSE) than the PCR model, demonstrating the importance of nonlinearity of the variables and the higher capability of the ANN model than the PCR model. By adding redness (a^*) and yellowness (b^*) and drying time to the input variables of ANNs, r and RMSE values were improved to 0.98 and 1.2 % for the training data set, and 0.85 and 2.2 % for the testing data set, respectively. Although the developed ANNs are available under the limited conditions of this study, our results suggest that the ANNs proposed offer reliable models and powerful prediction capability for the MC_f, even though wood properties vary considerably and their complex interrelations are not fully elucidated.

K. Watanabe (✉) · I. Kobayashi · N. Kuroda
Forestry and Forest Products Research Institute, 1 Matsunosato,
Tsukuba, Ibaraki 305-8687, Japan
e-mail: kenwatanabe@ffpri.affrc.go.jp

Y. Matsushita
SET Software Co., Ltd, 4-1-29 Imaike, Chikusa-ku,
Nagoya, Aichi 464-0850, Japan

Introduction

In our previous paper [1], drying tests were conducted with Sugi (*Cryptomeria japonica* D. Don) blocks and the variability in final moisture content (MC_f) in relation to wood properties was investigated using principal component regression (PCR) analysis. The MC_f of the blocks was demonstrated to be affected by many factors including initial moisture content (MC_i), basic density (BD), annual ring orientation (ARO), heartwood ratio (HR) and CIE L^* color. However, the relationships between the MC_f and these wood properties could not be sufficiently described by the PCR, because there are complex interrelations between the MC_f and the wood properties and some of the interrelations are nonlinear, which cannot be modeled by the PCR. Therefore, a nonlinear approach is more useful in describing their relationships and predicting the MC_f of wood.

Artificial neural networks (ANNs) are a powerful nonlinear data modeling method, capable of finding complex nonlinear interrelations among many variables that produce outcomes. The concept of ANNs is inspired from the biological system of the brain comprising many neurons interconnected through synapses that process information. The background information on ANNs can be found in the literatures [2, 3]. The characteristic feature of ANNs is that they are not programmed; they are trained from a series of examples without needing to know beforehand the relations which may exist between the variables involved in the process, by adjusting the weight of the relations between

the variables. Thus, we hypothesize that ANNs may be applicable for describing the complex interrelations between MC_f and wood properties.

ANNs have been applied to a steadily increasing number of modeling tasks in diverse fields of engineering and science [3]. In the field of wood science, ANN modeling has been used to predict mechanical and physical properties of wood, such as bending strength and stiffness [4–6], fracture toughness [7], thermal conductivity [8], hygroscopic equilibrium moisture content [9], non-isothermal diffusion of moisture [10] and dielectric loss factor [11]. Few studies have attempted to model the moisture content of wood during drying process. Wu and Avramidis (2006) [12] applied ANN modeling to the prediction of timber kiln drying rates based on species, basic density and drying time. Accurate prediction of the experimental drying rate data was achieved with the developed ANN model, supporting the powerful predictive capacity of ANN modeling method. Ceylan (2008) [13] developed ANNs to predict the drying rate of the timber stack based on temperature and humidity inside the kiln and drying time, and showed that the drying rate was successfully predicted. In the above studies, however, the inherent variation in the drying characteristics between individual timbers was not taken into consideration, and it is unclear that ANNs would be applicable for predicting the moisture contents of each timber in the stack. If the MC_f of individual timbers can be predicted by ANNs prior to drying, it will be beneficial for sawmills to improve pre-sorting strategies and drying schedules.

This study was aimed at assessing the capability of ANNs to predict the MC_f of individual wood samples. An ANN model for MC_f was developed based on wood properties and compared with a PCR model employed in our previous study [1]. Furthermore, an additional ANN model for the prediction of moisture content during the drying process (MC_d) was developed, and the possibility of improving the capability of ANNs was evaluated.

Materials and methods

Data collection

This study incorporated the data obtained from the experimental work by Watanabe et al. (2012) [1]. The materials and methods of the work are briefly mentioned as follows. 79 small samples were cut from 10 green lumbers of Sugi. The MC_i, BD, ARO, HR, L^* and annual ring width (ARW) of the samples were measured. The samples were air-dried for 28 days in a conditioning room at a temperature of 20 ± 1 °C and a relative humidity of 44 ± 2 % RH. The weight of each sample was measured once in a day and the MC_d was determined by oven-dry method. The moisture

content of the samples took 8 days on average to reach air-dry moisture content of 15 %. Therefore, the MC_f of the samples was defined as the moisture content after a drying period of 8 days. In addition to the above wood properties, the CIE a^* and b^* color of the samples were additionally measured and included in the model construction.

ANN modeling

ANNs consist of many artificial neurons that process their inputs and send the output to one or many connected neurons until the information propagation is complete and the network produces an output. The ANNs are trained to learn the relationships in data by adjusting internal parameters until the predicted outcome is as close to the desired outcome as possible. The architecture of the most popular neural networks, called feedforward multilayer perceptron, is depicted in Fig. 1. The ANNs consist of the input layer, the one or more hidden layers and the output layer. The input layer receives the initial values of the variables; the output layer shows the results of the network for the input values; and the hidden layer is where data are being processed and makes it possible to model highly nonlinear relationships between input and output. To determine the optimum architecture and performance of the ANN, several parameters are adjusted, such as number of neuron layers, number of neurons in each layer, transfer functions, learning rule, learning coefficient ratio, number of learning cycles, and initialization of weights and biases [2, 3]. The parameters, which are varied based on the complexity of the problem, are determined by a designer, and the choice of specific parameters is more or less subjective [10, 14].

ANN model construction

Error-back propagation [15] is a well-known method to determine the weights systematically. However, this is known to involve several problems. The most important of these is the slow pace of learning from examples.
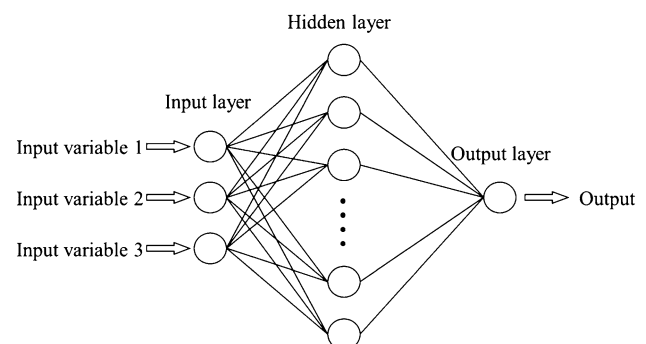

Fig. 1 General architecture of feedforward multilayer perceptron ANN

Moreover, the weights are computed by fixing the number of nodes in the hidden layer but the problem of arbitrariness of it could not be avoided.

As one of the approaches to improve these problems, cascade-correlation learning algorithm was developed by Fahlman and Lebiere (1991) [16] and showed significant improvements. Cascade-correlation is a method of incrementally adding processing elements. Instead of adjusting the weights in an ANN of fixed topology, cascade-correlation begins with a minimal network, then automatically trains and adds new hidden units one by one, creating a multi-layer structure. Once a new hidden unit has been added to the ANN, its input-side weights are frozen. This unit then becomes a permanent feature-detector in the ANN, available for producing outputs or for creating other, more complex feature detectors. NeuralWorks Predict (NWP) software (NeuralWare Inc., Pittsburgh, PA, USA) was used in this study which implements the cascade-correlation learning algorithm. NWP outperforms other neural network tools in that it also builds ANNs in the clever strategy of stopping rules against over-fitting on empirical data. Moreover, NWP undertakes some nonlinear transformation for input variables, and produces input neurons for each transformation in advance of learning process to avoid the complex representation of the model. Types of transformation used include linear (scaling), log, log–log, exponential, exponential of exponent, square-root, square, inverse, inverse of square-root, inverse of square, and so on, depending on the complexity of the problem [17]. NWP also uses a genetic algorithm to make a suitable choice of input variables from the set of all input variables and transformations of input variables [17], since it efficiently explores the large space of subsets of possible input variables.

Two types of ANN models were constructed to model MC_f and MC_d, respectively. In the case of MC_f, the input variables are MC_i, BD, HR, ARO, ARW and L^* that were used to develop a PCR model for MC_f [1], whereas in the case of MC_d, the input variables are MC_i, BD, HR, ARO, ARW, L^*, a^*, b^* and drying time. The a^*, b^* and drying time were additionally included in the input variables of the ANN model for MC_d, because the inclusion of more input variables increases the data size, which may enhance the number of possible structures and model performance. The output layer in both cases consisted of only one variable (an output neuron) corresponding to MC_f or MC_d. By employing the genetic algorithm, the each input variable was transformed by scaling, hyperbolic-tangent or natural logarithm functions.

The required data set for training and testing of the model were obtained from the experimental results of a total of 79 samples above mentioned. 60 samples were randomly selected for ANN training data set, while the remaining 19 samples were used to test the generalization capability of ANNs. The ANNs were trained with the training data set and the optimum number of neurons in the input (data transformation) and hidden layer were determined. The ANNs were then tested with the testing data set which was not used in the training process.

Besides the ANN modeling, a PCR model for MC_f was developed with the training data set in the same manner as in our previous study [1]. The MC_f of the 19 samples in the testing data set was predicted using the PCR model developed.

The error measurements between the measured and the predicted values were performed in both the training and testing processes using Pearson's r correlation by the following Eq. (1):

$$\text{Pearson's } r \text{ correlation} = \frac{\sum_{i=1}^{N}(xp - \overline{xp})(xm - \overline{xm})}{\sqrt{\sum_{i=1}^{N}(xp - \overline{xp})^2}\sqrt{\sum_{i=1}^{N}(xm - \overline{xm})^2}}$$

(1)

where N is the number of data sets, x_p is the predicted value, x_m is the measured value, $\overline{xp}$ and $\overline{xm}$ are the average values of each variable, respectively.

In order to show the degree of contribution of the input variables to the determination of the network output, a sensitivity analysis was performed with NWP, which computes partial derivatives of the output variable with respect to each of the input variables. The sensitivity analysis produces a quantitative measure of the variation in the MC_f calculated by the network, when each variable changes. The normalized sensitivity for each input variable was calculated according to Eq. (2):

$$\text{Normalized sensitivity} = \frac{\sum_{i=1}^{N}\left(\frac{\partial y_i}{\partial x_i}\right)^2}{\sigma^2 N}$$

(2)

where σ^2 is the variance of the partial derivatives for each input variable, x_i and y_i are the input and output vectors for each data set. High values of this sensitivity indicate that a slight variation of the variable produces considerable changes in the output MC_f, and vice versa. Furthermore, the average value of sensitivity for each input variable was calculated according to Eq. (3):

$$\text{Average value of sensitivity} = \frac{\sum_{i=1}^{N}\frac{\partial y_i}{\partial x_i}}{N}$$

(3)

which indicates a positive relationship between input and output variables for its positive sign, while a negative sign indicates an inverse relationship. This is a standard diagnostic procedure commonly used to gain insight into a multilayer neural network solution [17].

Results and discussion

A summary of the input and output variables for the training data set and the testing data set is listed in Table 1. There was a large variation in each variable. The testing data set was almost within the range of the training data set, but the maximum values of BD, ARW and b^* in the testing data set exceeded the range of the training data set. This may lead to an error in the prediction of MC_f and MC_d, because the developed models cannot extrapolate beyond the range of the data used for training (clipping to maximum and minimum values of the field).

Prediction of MC_f

Figure 2 shows the plots of the experimentally measured versus predicted MC_f in the training process. The ANN model yielded a correlation coefficient (r) of 0.91 and an

Table 1 Range and standard deviation of input and output variables for training data set and testing data set, respectively

Variables	Training data set ($n = 60$)				Testing data set ($n = 19$)			
	Mean	SD	Max	Min	Mean	SD	Max	Min
MC_i (%)	80.3	40.4	284.4	39.8	80.2	26.5	144.1	50.7
BD (kg/m^3)	313.2	35.2	463.1	263.7	317.0	53.1	515.6	264.3
ARO (°)	33	17	65	0	30	20	57	0
ARW (mm)	4.5	1.4	7.9	1.9	4.6	1.9	8.2	1.6
HR	0.8	0.3	1.0	0.0	0.7	0.4	1.0	0.0
L^*	33.8	8.7	52.3	11.9	36.4	8.0	51.0	21.1
a^*	14.3	3.4	22.3	7.9	14.2	2.9	20.0	11.2
b^*	28.9	4.1	34.6	14.9	30.2	3.3	35.9	24.7
MC_f (%)	15.8	6.5	36.2	8.8	13.5	4.1	22.5	9.0

SD standard deviation, *BD* basic density, *ARO* annual ring orientation, *ARW* annual ring width, *HR* heartwood ratio

RMSE of 2.8 %, while the PCR model yielded an r of 0.76 and an RMSE of 4.2 %. The r value in the PCR model is comparable with the one obtained in our previous study ($r = 0.74$) where a PCR model was developed and validated with the entire data set [1]. In both the models, the predicted MC_f was in good agreement with the measured one. It should be emphasized that the correlations produced by the models were much higher than the correlations between the MC_f and the input variables. When the entire data set was analyzed, significant r correlations between MC_f and MC_i ($r = 0.33$, $P < 0.01$), HR ($r = 0.38$, $P < 0.001$), and L^* ($r = -0.60$, $P < 0.001$) were identified, whereas no significant correlations were observed between MC_f and BD, ARO, and ARW [1]. Hence, both the ANN and PCR models had much better predictive ability for MC_f than the traditional simple linear regression.

When the MC_f of the testing data set was predicted (Fig. 3), the values of r and RMSE were 0.80 and 2.8 % (ANN model), and 0.66 and 3.8 % (PCR model). Compared with the statistics in Fig. 2, the correlation coefficients in the testing data set were lower than that in the
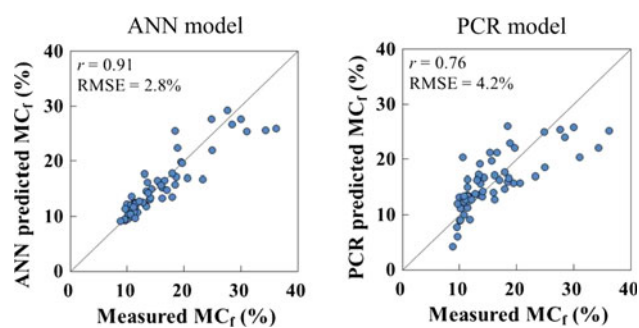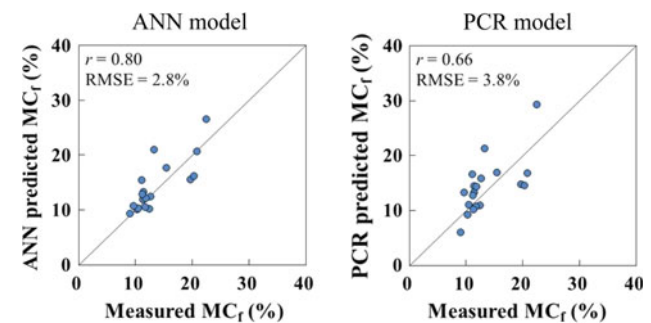

Fig. 3 Plots of the experimentally measured versus predicted MC_f in the testing process using the ANN model and the PCR model. The *Solid line*, a one–one relationship between measured and predicted values, *RMSE* root mean square error

Fig. 2 Plots of the experimentally measured versus predicted MC_f in the training process using the ANN model and the PCR model. The *Solid line*, a one–one relationship between measured and predicted values, *RMSE* root mean square error
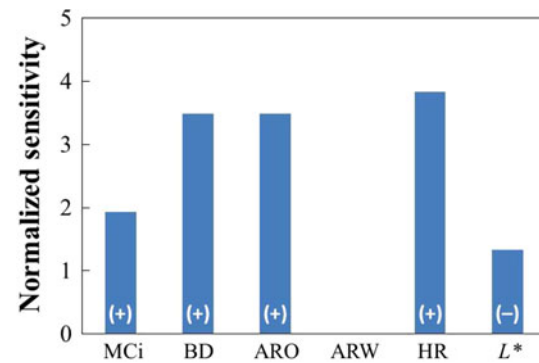

Fig. 4 Results of sensitivity analysis in the ANN model for predicting MC_f. The sign of the average sensitivity for each input variable is shown in *brackets*. *BD* basic density, *ARO* annual ring orientation, *ARW* annual ring width, *HR* heartwood ratio
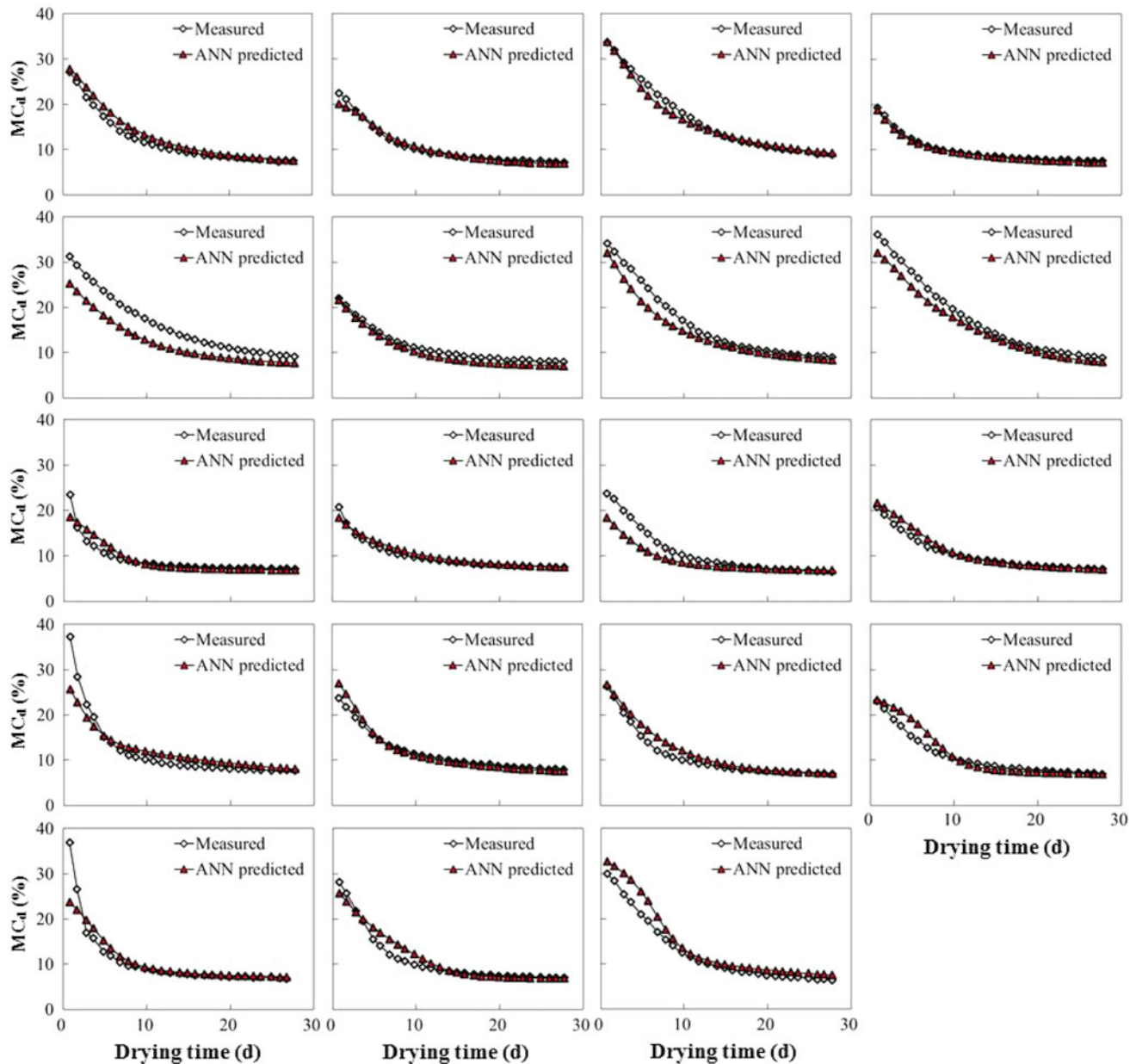
Fig. 5 Comparison between measured and ANN predicted MC_d for each sample in the testing data set

training data set. However, the performance of the models was still moderate to high.

In both the training and testing data sets, the ANN model consistently gave higher r values and lower RMSE values than the PCR model (Figs. 2 and 3). Thus, the predictive ability of the ANN model was demonstrated to be greater than that of the PCR model. The architecture of the ANN model for MC_f was consisted of 7 input neurons, 10 neurons having a hyperbolic-tangent transfer function in the hidden layer and 1 output neuron with a logistic transfer function. The number of neurons in the hidden layer was much higher than those reported by Wu and Avramidis

(2006) [12] and Ceylan (2008) [13], who predicted the drying rate of timber stacks using the ANN models with 4–5 neurons in the hidden layer. In general, the more neurons the hidden layer contains, the higher the nonlinearity of the ANNs is. Therefore, the 10 neurons in the hidden layer of the present ANN model indicate the highly nonlinear relationships between the input wood properties and the output MC_f, which may probably result in the higher predictive ability of the ANN model than the PCR model.

To estimate the relative importance of the individual inputs to model predictions, sensitivity analysis was
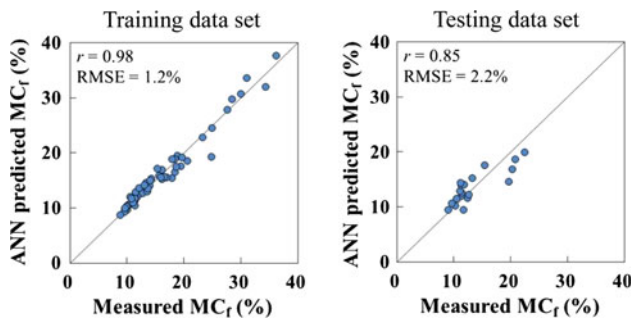
Fig. 6 Plots of measured MC_f versus ANN predicted MC_f for training data set and testing data set, respectively. The predicted MC_f was obtained from the drying curves predicted by the ANN model for MC_d. The *Solid line*, a one–one relationship between measured and predicted values, *RMSE* root mean square error

conducted. Figure 4 shows the normalized sensitivity for each input variable. The sensitivity value of ARW was 0, which means that the ARW was eliminated substantially from the input variables by the genetic algorithm. The sensitivity analysis revealed that all the variables except ARW had an influence on MC_f. The MC_i, BD, ARO and HR had a positive influence on the MC_f, while the L^* had a negative influence on the MC_f. This finding is in consistent with the results of the PCR analysis reported in our previous paper [1]. It is apparent from the figure that HR was the variable presenting a higher influence on MC_f, followed by BD, ARO, MC_i and L^* in decreasing order. This order is quite different from the one obtained from the PCR analysis [1]. Because the ANN model could describe the non-linear relationships between the wood properties and the MC_f more fully than the PCR model, the results of the sensitivity analysis are considered to be more reasonable than those of the PCR analysis.

Prediction of MC_d

The ANN model for MC_d was developed based on MC_i, BD, HR, ARO, ARW, L^*, a^*, b^* and drying time. The architecture of the ANN for MC_d was consisted of 9 input neurons, 20 neurons having a hyperbolic-tangent transfer function in the hidden layer and 1 output neuron with a logistic transfer function. Figure 5 shows the comparison between the measured and the ANN predicted MC_d for each sample in the testing data set. The predicted drying curves were roughly in agreement with the measured ones, in spite that the MC_i of the testing samples varied, ranging from 50.7 to 144.1 %. In around half of the samples, the ANN predicted drying curves fitted well to the measured ones, whereas in other samples, the predicted MC values in the early stage of drying did not closely follow the experimental ones. This discrepancy can be partially

explained by the insufficient number of data sets in the early stage of drying where the MC_d varied widely.

To further validate the ANN model for MC_d, the MC_f values were read from the drying curves obtained, and the measured and predicted MC_f values were compared. Figure 6 shows the plots of the measured MC_f versus ANN predicted MC_f for the training and testing data sets, respectively. The relationships between the two were good with an r of 0.98 and an RMSE of 1.2 % in the training data set and an r of 0.85 and an RMSE of 2.2 % in the testing data set. Compared with the results of the ANN model for MC_f (Figs. 2 and 3), higher correlations and lower RMSEs were found in both the training and testing data sets, which could be attributed to the additional input variables and the consequent larger data size in constructing the ANN model for MC_d. These results demonstrate that the capability of predicting the MC_f could be improved by the ANN model developed based on MC_i, BD, HR, ARO, ARW, L^*, a^*, b^* and drying time.

Overall, the ANN models had good predictive ability for MC_f, and therefore it can be suggested that the ANNs proposed offer reliable models and good prediction capability, even though wood properties vary considerably and their complex interrelations are not fully elucidated. However, the predictive ability of the ANN models in the testing process was poorer than that in the training process. This implies that the volume of data was not sufficiently large to guarantee the generalization of the ANNs, and new data sets may be required for further improvement.

Conclusions

The capability of ANNs to predict the MC_f of the individual small Sugi samples during air-drying was evaluated, comparing with the PCR model employed in our previous study. Our results showed that the proposed ANNs could model highly nonlinear relationships between the inherent wood properties and the MC_f, and successfully predicted the MC_f more accurately than the PCR model. These results suggest that the ANNs proposed offer reliable models and good prediction capability, even though wood properties vary considerably and their complex interrelations are not fully elucidated. It should be noted that the developed ANNs are available under the limited conditions of this study, although if this ANN approach can be scaled up to a real-size timber, it will allow sawmills to refine pre-sorting strategies by predicting the MC_f of individual timbers.

the wood processing department of Forestry and Forest Products Research Institute, for his assistance in the experiments.

References

1. Watanabe K, Kobayashi I, Kuroda N (2012) Investigation of wood properties that influence the final moisture content of air-dried sugi (*Cryptomeria japonica*) using principal component regression analysis. J Wood Sci Online First. doi:10.1007/s10086-012-1283-5
2. Picton P (2000) Neural Networks, 2nd edn. Palgrave, New York
3. Samarasinghe S (2006) Neural networks for applied sciences and engineering. From fundamentals to complex pattern recognition. CRC Press, Boca Raton
4. Mansfield SD, Iliadis L, Avramidis S (2007) Neural network prediction of bending strength and stiffness in western hemlock (*Tsuga heterophylla* Raf.). Holzforschung 61:707–716
5. Mansfield SD, Kang K-Y, Iliadis L, Tachos S, Avramidis S (2011) Predicting the strength of *Populus* spp. clones using artificial neural networks and ε-regression support vector machines (ε-rSVM). Holzforschung 65:855–863
6. Esteban LG, Fernández FG, Palacios P (2009) MOE prediction in *Abies pinsapo* Boiss. timber: application of an artificial neural network using non-destructive testing. Comp Struc 87:1360–1365
7. Samarasinghe S, Kulasiri D, Jamieson T (2007) Neural networks for predicting fracture toughness of individual wood samples. Silva Fennica 41:105–122
8. Avramidis S, Iliadis L (2005) Predicting wood thermal conductivity using artificial neural networks. Wood Fiber Sci 37:682–690
9. Avramidis S, Iliadis L (2005) Wood-water sorption isotherm prediction with artificial neural networks: a preliminary study. Holzforschung 59:336–341
10. Avramidis S, Wu H (2007) Artificial neural network and mathematical modeling comparative analysis of nonisothermal diffusion of moisture in wood. Holz Roh Werkstoff 65:89–93
11. Avramidis S, Iliadis L, Mansfield SD (2006) Wood dielectric loss factor prediction with artificial neural networks. Wood Sci Technol 40:563–574
12. Wu H, Avramidis S (2006) Prediction of timber kiln drying rates by neural networks. Dry Technol 24:1541–1545
13. Ceylan I (2008) Determination of drying characteristics of timber by using artificial neural networks and mathematical models. Dry Technol 26:1469–1476
14. Özşahin Ş (2012) The use of an artificial neural network for modeling the moisture absorption and thickness swelling of oriented strand board. BioResources 7:1053–1067
15. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds) Parallel distributed processing, vol 1. MIT Press, Cambridge
16. Fahlman SE, Lebiere C (1991) The cascade-correlation learning architecture. August 29, CMU-CS-90-100, School of Computer Science, CMU, Pittsburgh, PA
17. NeuralWare (2009) NeuralWorks Predict® User Guide. The complete solution for neural data modeling. NeuralWare, Pittsburgh, PA, USA