Feedforward Neural Network Modeling of Fir Taper in Natural Forests of Greece

Spiros Kaloudis, Thomas Glezakos, Konstantinos P. Ferentinos, Theodore A. Tsiligiridis, Constantinos P. Yialouris

Informatics Laboratory, Agricultural University of Athens, Iera Odos 75, 11855, Athens, Greece.
*kaloudis@aua.gr, t_glezakos@yahoo.com, kpf3@cornell.edu, {tsili, yialouris}@aua.*gr

Abstract. *In this paper, a model of the taper of fir natural forests in three specific areas of Greece is developed. The modeling approach was that of feedforward neural networks, trained by the backpropagation training algorithm. Several one- and two-hidden-layer topologies were investigated and three final networks were trained and tested on real measured data. The obtained taper values were accurate enough so that neural networks could be considered as a useful alternative to the not so precise multivariable linear regression methodology used so far.*

Keywords. *Fir, neural networks, taper.*

1. Introduction

Forests management planning is a complicated task due to the high number of environmental and forest variables that participate in forest evolution, the long scheduling time and the contradictory human demands. Due to the complexity of forest management planning, forests managers need to have some insight information about all the factors that affect forest production. Factors with crucial role are the potential forest productivity and the usability of the produced timber for various uses.

The productivity of a forested area depends on several environmental factors, such as climate and topography, and tree species. This is expressed by the high of dominant trees at a reference age [6]. Other indices for site quality also apply, such us climatic characteristics and soil properties, but they are not considered credible for general use [16]. The usability of a tree log for special uses such as poles, depends on its shape. The shape of tree stem can be described by the change of its diameter over the length unit of the stem, which is called *taper*. The diameter of trees is affected by stand density; subsequently taper is also affected by stand density [3]. This is the reason why taper is considered of limited applicability as site quality index. However, due to the impact of stand density, taper can be manipulated by applying convenient silvicultural treatments to the stand. Thus, the improvement of timber usability can be achieved for each forest location, in respect both to environmental limitations and silvicultural treatments.

The purpose of this paper is to demonstrate a methodology that could be able to estimate the taper of a tree species (*fir* in our case), based on selected environmental factors and stand density. Two parameters were used as a metric of stand density: the crown closure and the basal area of the trees with diameter larger than 14cm. Relevant studies have been made for the prediction of the Site Quality Index or taper by environmental factors [5], [6], [8], [12], [21], [22]. Most of the undertaken studies use the multivariable linear regression as the basic correlation mechanism. Due to the fact that certain environmental variables like altitude have no linear effect on forest species, linear regression is not considered as the best method in this case. In addition, non-linear regression models are difficult to apply. In this work, feedforward neural networks were used to model the taper.

Neural networks (NNs) have been used to model a variety of biological and environmental processes (*e.g.*, [1], [9]-[12], [20]), but not in the specific area of taper modeling. Due to their capability to model highly non-linear processes, purely based on measured data, feedforward NNs were considered a proper alternative to the not so successful multivariable linear regression.

2. Materials and methods

The samples were collected from three different areas of the mainland of Greece, namely (from south to north) Parnitha, Karpenisi and Pertouli. These three areas are mountainous with varying topography. The first study area (Parnitha), according to Scaltsoyiannes et al. [19] and Mitsopoulos and Panetsos [19], is covered by *fir* forest (*Abies Cephalonica*). The forest was declared as a national park and consists of very old trees due to the reduction of cuttings. It suffers from severe attacks by parasite *Viscum album* [14]. The climate is hot-Mediterranean and the summer is hot with an annual number of biologically dry days between 40 and 75. The majority of rocks are limestone and flysch. The soil in most cases is shallow and only in a few places with low inclination and reach vegetation it is deep.

The second study area (Karpenisi) consists of a hybrid fir species (*Abies borissi-regis)* produced by fertilization between *Abies alba* and *Abies Cephalonica* [15], [19]. The forest is in good condition and is managed by the local forest service. The climate is wet and cold, the average annual rainfall is 1380mm and in the winter there is high snow accumulation. The dry season is short and appears annually from July to September. The most common types of rocks are flysch, psammite and limestone. The soil is of medium depth and also contains medium quantities of organic matter.

The forest in the third study area (Pertouli) consists of *Abies borissi-regis* and is one of the few forests that belong to the School of Forestry, University of Thessalonica. The forest is in good condition and managed regularly.

2.1. Taper

Taper is defined as the amount of tree diameter change over the unit length of tree stem [13]. It expresses the tree stem completeness and thus provides an index for the stem usability to some special uses, such as poles. Taper is not constant across the whole length of the tree stem. Two types of taper can be distinguished:

i. The *absolute taper*, which is the difference between two diameters with a distance of one meter, and
ii. The *relevant taper*, which is the difference between breast height diameter (considered having a value equal to 100) and other diameters, which are measured in various distances from breast height and expressed as a percentage of the breast height diameter.

There are various ways to calculate taper [2], [13], [17]. In this study, taper is calculated by:

$$T = \frac{d_{1.30}}{h - 1.30} \tag{1}$$

where, T is taper, $d_{1.30}$ is the diameter at breast height (in cm) and h is the total tree height (in m).

In order to investigate the effect of taper calculation to the accuracy of the prediction, taper was calculated by two sets of trees. Specifically, it was calculated over either the entire number of measured trees of each data set, or over the five dominant trees.

2.3. Sampling methodology

Seventy eight sampling surfaces were collected in total (29 from Parnitha, 21 from Karpenisi and 28 from Pertouli). Each surface forms a square with a side of 15 meters (225m^2 area). The two sides were positioned in parallel to the contours and the other two vertically. In each sampling surface, only trees with a breast height diameter equal to or greater than 14cm were considered. For each tree the diameter at breast height and the total tree height were measured. The following factors were also measured: Altitude with GPS, aspect with a compass, slope with a clinometer, position on the incline in qualitative terms, incline shape in qualitative terms, soil depth in qualitative terms taken from maps and field observations, surface rock content measured as coverage percentage and taken from field observations, soil acidity (pH) measured by pH-meter, crown closure measured as coverage percentage, taken from field observations.

3. Feedforward Neural Network modeling

Given the fact that the traditional statistic methods are not well adapted to the problem, it was decided that the design and use of an appropriate NN with good generalization capabilities, could facilitate towards a solid solution. The program that was used for the design and implementation of the various networks was BrainMaker [4]. The feedforward type of NNs was used (multi-layer perceptron) while the backpropagation training algorithm [18] was used as the training methodology.

3.1. Initial data manipulation

The initially accumulated data, which contained 78 samples, were assembled out of the three study areas. All measured variables described in Section 2.3 comprised the NN inputs, while taper was the single NN output (Table 1). These variables were selected in regards to their relevance towards the value of taper, as well as their availability and ease of measurement.

Table 1. The 12 measured environmental variables that formed the NN inputs

Variable	Type of NN input	Min	Max	Number of NN inputs
Altitude	Number (meters)	750	1391	1
Aspect	Number (grads)	0	395	1
Slope	Number (degrees)	0	39	1
Soil acidity (pH)	Number	4.2	7.3	1
Surface rock content	Number (percent)	2	90	1
Basal area	Number (m^2)	0.279	2.576	1
Crown closure	Number (percent)	10	90	2
Position on the incline	Qualitative (Binary)	0	1	3
Incline shape	Qualitative (Binary)	0	1	3
Soil depth	Qualitative (Binary)	0	1	3
Location	Qualitative (Binary)	0	1	3
Tree species	Qualitative (Binary)	0	1	2
	Total number of NN inputs:			22

They can be organized into two categories: structural data, such as altitude or soil depth, which express constant variables for each location, and variable data, such as basal area or crown closure, which express properties at the level of the tree.

The initial data set was characterized by certain problems towards its usage as input to the NN. Firstly, the volume of the data set is considered to be minimal for the training and testing of a NN. In addition to that, there were 21 missing values for the crown closure variable, which were manipulated according to the procedures proposed in [4].

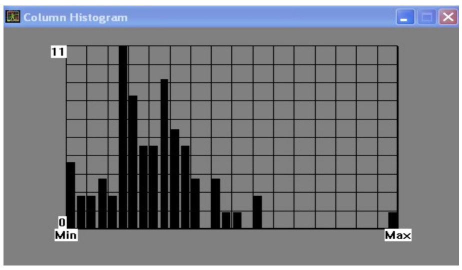
Finally, there was the need to analyze the available data in order to uncover possible standalone predictors of the taper variable. Each column of data was compared with the taper column in regards to correlation and Table 2 was produced.

The strength of correlation reveals the intensity of the connection between two variables, as well as its direction. The results clearly show a strong correlation of positive strength between the requested outputs taper and slope as well as soil acidity (pH). The rest of the correlation values are not that strong and probably indicate that it should take the combination of such inputs to render a recognizable effect as a strong predictor for taper. The role of the NN to be designed is to uncover such relationships and take control of them so as to produce a system with good generalization capabilities.

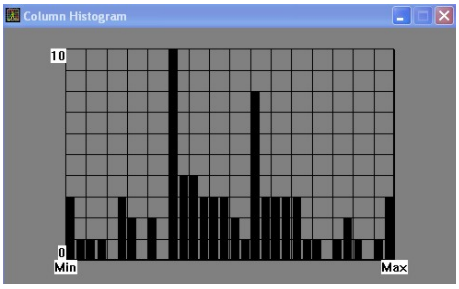
Table 2. Strength of input/output data correlations

Variable	Strength
Altitude – Taper	0.35
Slope – Taper	0.47
Soil acidity (pH) – Taper	0.41
Surface rock content – Taper	0.27
Basal area – Taper	0.30

The next step in data manipulation was to examine the distribution of data for each variable and eliminate the potential problem arisen by outlier values. Such values may turn the network's attention to certain isolated cases which, although they may exist, they are not so common in the data set. If these extreme values were to bear the same importance as the more common values, the network would have more trouble in distinguishing between them and the more common ones and setting them aside. For example, the data distribution histogram for the variable basal area (Fig. 1a) clearly shows that there is an extreme lonely case at the far right which should be eliminated before the variable would be fed into the NN.



(a)



(b)

Figure 1. Distribution histogram of basal area variable in the training data set, (a) initially and (b) after appropriate data manipulation

In order to improve the performance of the network by giving to the NN the inherent ability to look more closely at the most typical values, the maximum range should be altered for this variable, resulting in a more normally spread out distribution histogram (Fig. 1b). This does not mean that the extreme values are swept away from the network's consideration, but rather that the system will give more attention looking closer at the typical range, without of course ignoring the outliers.

Finally, shuffling of the available data samples was performed in order to minimize the bias of the system to be produced. The initial data set was comprised of records grouped by the location of the study areas.

3.2. Data sets formulation and NN design and training

The need to overcome the scarcity of data led to the decision to develop a large number of networks and evaluate them initially by their scores towards a small number of samples which the networks had not "seen" before. The networks which best performed during this procedure would furthermore undergo training using cross-validation.

The initial randomized and manipulated data set of the 78 samples was split into two parts. The first segment of data, comprising of 73 samples, was considered as the training set and would be further split into two sections in a ratio of 9:1, forming the training and cross-validation data sets for the NNs respectively. The second segment of the initial data set, including only 5 samples, formed the testing set to be used for the evaluation of the performance of the trained networks.

3.2.1. Design of NN topologies

In order to select the best performing network topology, the common trial-and-error approach was used, evaluating the performance of each NN topology in regards to RMS and average errors and the achieved R-squared. The networks were trained with training tolerance and cross-validation tolerance equal to 0.1, a constant learning rate of 1 and a basic smooth factor equal to 0.9. Several feedforward NN topologies ranging from one hidden layer (1-HL) containing only 5 neurons to two hidden layers (2-HL) with 5 and 22 hidden neurons respectively were trained and investigated. The most effective network according to both average and RMS errors, but also one with the best R-squared, had two hidden layers with 7 and 11 hidden nodes respectively. Table 3 shows the 5 best performing NN topologies sorted by their RMS error values.

The criterion for the number of hidden layers and neurons for the next network which participated in our research, took under consideration the number of inputs and patterns of our initial data set, as well as the number of data samples contained in the initial training set. This network had only one hidden layer with 9 hidden nodes, a value given by the following equation [4]:

$$HN = \frac{\frac{(Inputs + Outputs)}{2} + (0.1 * NoTS)}{2} = 9 \qquad (2)$$

Where,
HN = Hidden Neurons,
$NoTS$ = Number of Training Samples.

Lastly, the third NN contained again only one hidden layer with 19 neurons, which was the output of another trial-an-error procedure, based on the raw available data, before any special data manipulation.

3.2.2. NN training and cross-validation

Initially, we tried to fully train the networks without cross-validation, that is, by using the entire training set of 73 samples. However, after convergence, their evaluation was lower than expected, with success ranging from 62.5% to 67.5%. This was a clear indication that the training set was rather small, thus cross-validation was used for early stopping of the training process so that overfitting was avoided.

Three pairs of networks were trained with two randomly generated training and cross-validation sets from the initial set of available data. The entire NN design and training process was repeated five times with different randomly generated training/testing sets combinations, in order to better prove the value of the results.

Table 3. Best performing NN topologies sorted by their RMS error values.

Ranking	1-HL	2-HL	Training iterations	Avg error	RMS error	R^2 Taper
1	7	11	285	0.0183	0.0240	0.9540
2	36	7	672	0.0235	0.0336	0.9322
3	22	8	145	0.0260	0.0317	0.9224
4	33	10	97	0.0281	0.0358	0.9581
5	38	6	346	0.0289	0.0331	0.9439

4. Testing results

In this section, the results of the application of the three final NN topologies to the four testing data sets are presented. These three network topologies were trained each time with two different randomly generated training sets. Training was performed with and without cross-validation for early stopping. Thus, in total, 12 different configurations were produced, as described in Table 4.

As described in Section 2.2, taper values were calculated over either the entire number of measured trees of each data set, or over the five dominant trees. These two approaches are denoted as Taper-All and Taper-5, respectively. Figures 2–4 summarize the average performance of all 12 configurations over the five randomly generated testing sets.

Table 4. The various network configurations considered for testing. Configurations ending with "1" used cross-validation during training for early stopping, while those ending with "0" did not.

Configurations	NN topology	Training data set no.
CONF10, CONF11	1-HL, 9 nodes	1
CONF20, CONF21	1-HL, 9 nodes	2
CONF30, CONF31	1-HL, 19 nodes	1
CONF40, CONF41	1-HL, 19 nodes	2
CONF50, CONF51	2-HL, 7 – 11 nodes	1
CONF60, CONF61	2-HL, 7 – 11 nodes	2

Generally, RMS of Taper-5 has less variation among the various configurations, compared to Taper-All. Its best performance is achieved with configuration 11 (CONF11), with RMS value of 0.369 (cm/m) (Fig. 2). Taper-All has the best performance with configuration 51 (CONF51), with RMS value of 0.350 (cm/m) (Fig. 2).

The observed and predicted values of Taper-5 with CONF51, averaged over the five testing sets, are given in Fig 3. The correlation coefficient (R) between the observed and predicted values is 0.559, while the corresponding value for Taper-All is 0.701. In Fig. 4, the observed and predicted values of Taper-All with configuration CONF51 are shown, again averaged over the five testing sets. Among all examined cases, the predicted values of Taper-All using the configuration CONF51 achieved the best fit to the observed values.

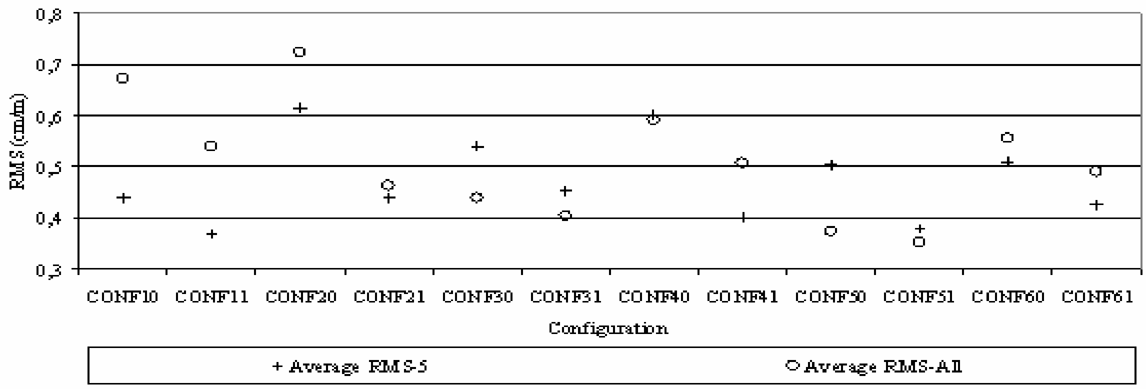


Figure 2. Average RMS of observed and predicted values of Taper-5 and Taper-All, for the five tested data sets, with the various configurations of NN.

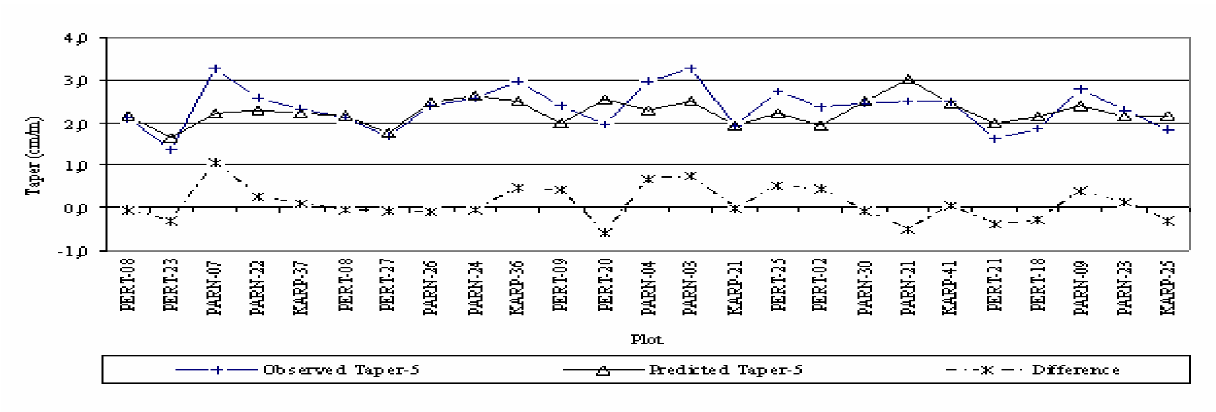


Figure 3. Observed and predicted values of Taper-5, averaged over the five testing sets, based on NN configuration 11 (CONF11), and the corresponding errors.

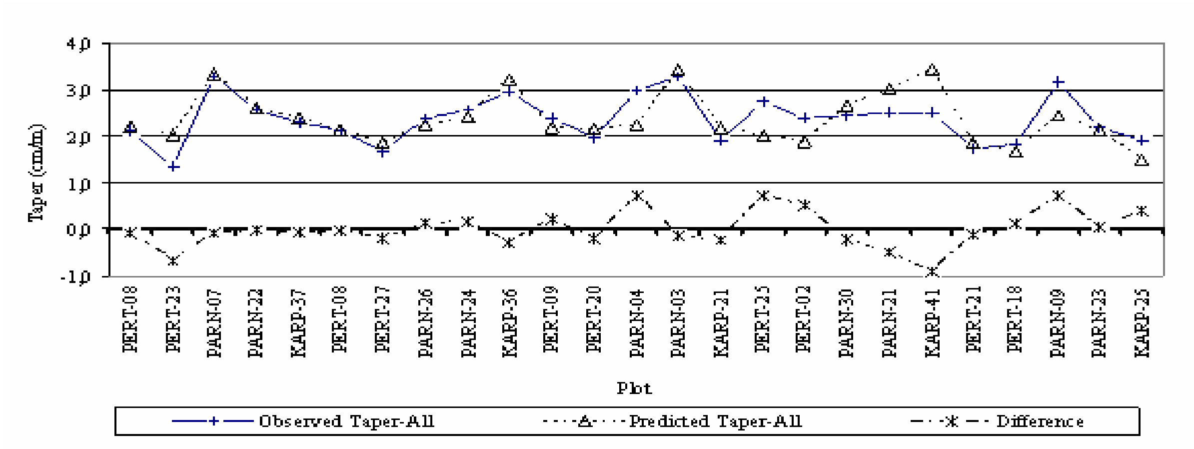


Figure 4. Observed and predicted values of Taper-All, averaged over the five testing sets, based on NN configuration 51 (CONF51), and the corresponding errors.

5. Conclusions

In this work, a model of the taper of fir natural forests in three specific areas of Greece was developed. Taper values for training and testing purposes were calculated over either the entire number of measured trees of each data set, or over the five dominant trees. After some data manipulation and preliminary training explorations over the best possible network topology, 12 network topology/training approach configurations were considered for the final testing of the developed modeling methodology. Among all these configurations, a 2-hidden-layer NN with 7 and 11 nodes in the first and second hidden layer respectively, trained using cross-validation for early training stopping to avoid overfitting, achieved the best performance.

Linear models of taper use a large number of variables and thus loose their efficacy or they obtain low accuracy. In addition, any appropriate non-linear models are very hard to manipulate. Therefore, the accuracy achieved by the proposed neural network approach, make it a useful modeling alternative.

Future work includes the introduction of different environmental variables and the consideration of a larger amount of measured data samples for further improvement of the developed neural network model.

6. References

[1] Altendorf, C.T., Elliott, R.L., Stevens, E.W., Stone, M.L. Development and Validation of a Neural Network Model for Soil Water Content Prediction with Comparison to Re-

gression Techniques. Trans. ASAE 1999; 42: 691-699

[2] Anuchin, N. Forest Mensuration. 2nd Edition. Israel Program for Scientific Translations. Jerusalem 1970.

[3] Assmman, E. Waldertragskunde. 1961.

[4] BrainMaker Professional, Neural Network Simulation Software. User's Guide and Reference Manual, Part1 & Part2.

[5] Corona, P., Scotti, R., Tarchiani, N. Relationship Between Environmental Factors and Site Index in Douglas – Fir Plantations in Central Italy. Forest Ecology and Management 1998; 110: 195-207

[6] Curt, T., Bouchaud, M., Agrech, G. Predicting Site Index of Douglas – Fir Plantations from Ecological Variables in the Massif Central Area of France. Forest Ecology and Management 2001; 149: 61-74.

[7] Davis, L.S., Johnson, K.N. Forest Management. McGraw-Hill; 1987.

[8] Economou, A. Site Quality Evaluation of Planted and Naturally Regenerated Black Pine (*Pinus nigra*, (Arn) var. *pallasiana* (Lamb*)* in Arcadia and Central Pindos Range, Greece. PhD dissertation; 1987.

[9] Ferentinos, K.P., Albright, L.D. Predictive Neural Network Modeling of pH and Electrical Conductivity in Deep-Trough Hydroponics. Trans. ASAE 2002; 45: 2007-2015

[10] Ferentinos, K.P.: Biological Engineering Applications of Feedforward Neural Networks Designed and Parameterized by Genetic Algorithms. Neural Networks 2005; 18: 934-950

[11] Hong, F., Tan, J., McCall, D.G. Application of Neural Network and Time Series Techniques in Wool Growth Modeling. Trans. ASAE 2000; 43: 139-144

[12] Kaloudis, S., Touratzi, T., Didaskalou, K., Galanopoulou, S., Kerkides, P. Taper Correlation with Environmental Factors and Stand Density in Fir Forest of Parnitha. In: Proceedings of 1st pan-Hellenic environmental congress "Modern Environmental Matters" (in Greek). Orestiada 2004; p. 60-68.

[13] Matis, K. Forest Biometrics II, vol. A, (in Greek). Thessaloniki: 1994.

[14] Ministry of Agriculture of Greece, General Secretariat of Forests and Natural Environment. Natural Forest of Parnitha, Research on structure and evolution of fir forest. (in Greek) Athens, Greece: 1996.

[15] Mitsopoulos, D., Panetsos, C. Origin of Variation in Fir Forests of Greece. Silvae Genetica 1984; 36: 1-15

[16] Papamihos, N. Lecture Notes in Forest Edaphology (in Greek). Thessaloniki: 1979.

[17] Prodan, M. Sauerlander's Verlag. Frankfurt: (1965)

[18] Rumelhart, D.E., Hinton, G.E., Williams, R.J. Learning Representations by Back-Propagating Errors. Nature 1986; 323: 533-536

[19] Scaltsoyiannes, A., Tsaktsira, M., Drouzas, D. Allozyme Differentation in the Mediterranean Firs (Abies, Pinaceae): A First Comparative Study With Phylogenetic Implications. Plant Systematics and Evolution 1998; 216: 289-307

[20] Seginer, I., Boulard, T., Bailey, B.J. Neural Network Models of the Greenhouse Climate. J. of Agr. Eng. Res. 1994; 59: 203-216

[21] Tyler, A.L., Macmillan, D.C., Dutch, J. Predicting the Yield of Douglas Fir from Site Factors on Better Quality Sites in Scotland. Ann. Sci. For. 1995; 53: 619-634.

[22] Wang, G.G. Is Height of Dominant Trees at a Reference Diameter an Adequate Measure of Site Quality?. Forest Ecology and Management 1998; 112: 49-54