Support Vector Machines versus Artificial Neural Networks for wood Dielectric Loss Factor estimation

Lazaros Iliadis[1*], Stavros Tachos[2], Stavros Avramidis[3] Shawn Mansfield[3]

[1]Democritus University of Thrace , Department of Forestry & Management of the Environment & Natural Resources, 193 Pandazidou st., 68200 N Orestiada, Greece,
[2] Aristotle University of Thessaloniki, Greece [3] University of British Columbia, Department of Wood Science, Vancouver, Canada
*Email: [1] liliadis@fmenr.duth.gr [2] staxos@gmail.com [3] stavros.avramidis@ubc.ca 4

Abstract: This research effort aims in the estimation of Wood Loss Factor by employing Support Vector Machines. For this purpose experimental data for two different wood species were used. The estimation of the dielectric properties of wood was done by using various Kernel algorithms as a function of both ambient electro-thermal conditions applied during drying of wood and basic wood chemistry. Actually the best fit neural models that were developed in a previous effort of our research team were compared to the Kernels' approaches in order to determine the optimal ones.

Keywords: Wood Loss Factor, Gaussian Support Vector Machines, Polynomial Support Vector Machines, Artificial Neural Networks

1. Introduction

It has been shown in the literature that the knowledge of dielectric properties of wood can be used to determine its density and moisture content nondestructively. On the other hand this knowledge can lead to the detection of hand knots, spiral grain, and other defects [14]. From a theoretical perspective a better understanding of the molecular structure of wood and wood-water interactions can be obtained [9]. Various approaches towards this direction have been reported in the literature, namely the radio frequency vacuum drying [11][12][13] and the high frequency electric field heating, such as the veneer and finger-joint gluing and parallam manufacturing [19]. Thus, estimation of the dielectric constant (ε'), the loss tangent ($tan\delta$) and the loss factor (ε'') plays a catalytic role in the process of design, control, optimization and simulation.

Recently Koumoutsakos [11] has shown that ε'' is proportional to the thermal energy transferred to the wood, during radio frequency vacuum drying (RFV) process. RFV heating is a volumetric method where thermal energy is produced simultaneously through a pile of lumber which is placed inside an electromagnetic field. In this process the moisture starts its transportation from the center to the surface as soon as the wood is exposed to the field [1]. In fact, the electric power converted to thermal is given by the following formula 1.

$$PD = 5.56 \times 10^{-11} \, E^2 \, f \, \varepsilon'' \qquad (1)$$

It should be clarified that PD is the power density in W/m^3, E is the field strength in V/m and f is the frequency measured in Hz

Research efforts in the past have shown that there exists a strong relationship between the dielectric properties and wood attributes studied plus field frequency applied [18][19].

It is a fact that despite all conducted research so far, there is very little respective knowledge regarding the effect of wood chemical composition on the determination of ε''. Actually, Norimoto [15][16] have investigated the dielectric properties of some wood chemical constituents as a function of frequency and temperature, but no attempt was made towards the correlation of their percent content in the cell-wall composition to the gross wood ε'' values.

Another point that motivated this study is the fact that the data that have been used so far to construct such models have originated from assorted species, under different thermo-physical conditions and variable frequencies and thus, they are not suitable to be employed in drying modeling [2].

This research effort aims in the development of Support Vector Machines models (Gaussian, Fuzzy weighted and Polynomial) capable of determining reliably the value of ε'' based on ambient electro-thermal conditions and also on basic wood chemistry. On the other hand a comparative study is performed with a previous study of our research team that used Artificial Neural Networks (ANNs) for the same purpose [2]. This comparison aims in the determination of the optimal approaches that can be used to offer reliable and most of all cost and time effective approximation.

2. Materials and methods

2.1. Obtaining the experimental data

This study employees ε'' values and macro-physical data that were already published by Zhou and Avramidis [22]. According to Zhou and Avramidis all-sapwood and all-heartwood western hemlock [*Tsuga heterophylla* (Raf.) Sarg.], and all-heartwood western red cedar [*Thuja plicata* Donn] specimens were evaluated in the radial direction (thickness) and at various moisture contents and temperature levels were exposed to two levels of electric field voltage. After the exposure was terminated the ε'' values were calculated indirectly by performing heating studies at a 13.56 MHz fixed frequency with a laboratory size RFV dryer [22] [2].

Also the above described wood species were analyzed by Zhou and Avramidis [22] regarding their chemical composition as follows: air dried wood samples were ground in a Wiley Mill to pass a 40-mesh screen and extracted in a Soxhlet apparatus with acetone for 12 hours, and extractives determined gravimetrically [2]. According to Zhou and Avramidis [22], the content of Lignin was determined using a modified Klason approach derived from the TAPPI standard method T222 om-98. For more details refer to [22].

2.2. 5-Fold-Cross Validation

Cross-validation is a methods for estimating generalization error based on "resampling" [21] [7] [8] [17]. The resulting estimates of generalization error are often used for choosing among various models. In k-fold cross-validation, the data are divided into k subsets of (approximately) equal size. The model is trained k times, each time leaving out one of the subsets from training, but using only the omitted subset to compute whatever error criterion interests you. If k equals the sample size, this is called "leave-one-out" cross-validation. The division of the data set was done by the use of MATLAB's *crossvalind* function, which is included in the Bioinformatics Toolbox.

2.3. ε-SV Regression (ε-SVR)

Support Vector Machines (SVM) are keeping the training error foxed while at the same time they are minimizing the confidence interval [10]. Their primary target was pattern recognition [3]. Now they are used for both regression and classification. Vapnik [20] [10] introduced the following loss function that ignores errors less than a predefined value $\varepsilon > 0$

$$\left| y - f\left(\vec{x}\right) \right|_{\varepsilon} = \max\left\{ 0 \ , \ \left| y - f\left(\vec{x}\right) \right| - \varepsilon \right\} \tag{2}$$

It is a fact that the ε-SVR algorithm offers in many cases the optimal function of the form: $f\left(\vec{x}\right) = k\left(\vec{w}, \vec{x}\right) + b \qquad \vec{w}, \vec{x} \in R^{N}, \ b \in R$ (3). The whole idea is based on the determination of the function with the minimum testing error. The problem in this case is that the minimization of the above function is not possible due to the fact that the probability distribution P is unknown. Consequently, the actual solution can be reached by minimizing the following normalized risk function 4.

$$\frac{1}{2}\left\| \vec{w} \right\|^2 + C_{SVR} \cdot R_{emp}^{\varepsilon}\left[f \right] \tag{4}$$

$R_{emp}^{\varepsilon}\left[f \right]$ is the function of empirical risk $R_{emp}\left[f \right] = \dfrac{1}{p}\sum\limits_{i=1}^{p} L\left(f, \vec{x}_i, y_i \right)$ (5) whereas the loss function is $L(f, \vec{x}_i, y_i) = \left| y_i - f\left(\vec{x}_i\right) \right|_{\varepsilon}$ (6) and it ignores errors less than ε. Also it is clear that $\left\| \vec{w} \right\|^2$ is related to the complexity of the model, whereas C_{SVR} is a constant value determining the point that relates $R_{emp}^{\varepsilon}\left[f \right]$ to $\left\| \vec{w} \right\|^2$ [6].

Obviously, the objective which is the Minimization of the above risk function 4 can be phased as a specific optimization problem which phases the constraints described below: The aim is the minimization of the following function:

$$\tau\left(\vec{w}, \vec{\xi}, \vec{\xi^*} \right) = \frac{1}{2}\left\| \vec{w} \right\|^2 + C_{SVR} \cdot \frac{1}{p}\sum\limits_{i=1}^{p}\left(\xi_i + \xi_i^* \right) \tag{7}$$

The constraints in this case are: $\left(k\left(\vec{w},\vec{x}_i\right)+b\right)-y_i \leq \varepsilon+\xi_i$ (8) and $y_i-\left(k\left(w\cdot x_i\right)+b\right)\leq \varepsilon+\xi_i^*$ (9) and of course $\xi_i,\xi_i^* \geq 0$ (10). Obviously (and this can be seen in the following figure 1) ξ_i^*,ξ_i are the distances of the training data set points from the e-zone. All of the errors that are smaller than ε are ignored. As it is clearly shown, ξ_i stands for the distance of a point that can be found above the e-tube zone and ξ_i^* stands for the distance of a point that is located below the e-tube zone. It is well known that in the case of the application of Lagrange multipliers the problem would correspond to a double optimization one as follows: Maximize

$$W\left(a,a^*\right)=-\varepsilon\sum_{i=1}^{p}\left(a_i^*+a_i\right)+\sum_{i=1}^{p}\left(a_i^*-a_i\right)y_i-\frac{1}{2}\sum_{i,j=1}^{p}\left(a_i^*-a_i\right)\left(a_j^*-a_j\right)k\left(x_i,x_j\right) \quad (11)$$

subject to $\sum_{i=1}^{p}\left(a_i-a_i^*\right)=0$ (12) and $a_i^{(*)} \in \left[0,\dfrac{C_{SVR}}{p}\right]$ (13).
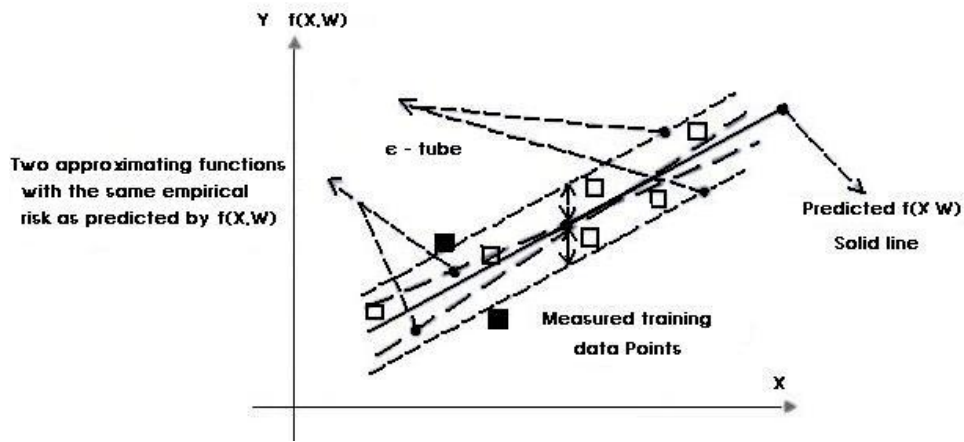

Figure 1. ε-SVR for Regression

Based on Vapnik and Kecman [20][10], the optimization problem can be solved by the use of the following linear extension of the kernel functions (14) and (15)

$$w=\sum_{i=1}^{p}\left(a_i^*-a_i\right)\vec{x}_i \qquad (14)$$

$$f\left(\vec{x},a,a^*\right)=\sum_{i=1}^{p}\left(a_i^*-a_i\right)k\left(\vec{x},\vec{x}_i\right)+b \qquad (15)$$

where b is estimated by the following function 16

$$b = average_i \left\{ \varepsilon \cdot sign\left(a_i - a_i^*\right) + y_k - \sum_j \left(a_j - a_j^*\right) k\left(\vec{x_j}, \vec{x_i}\right) \right\} \tag{16}$$

The training set vectors $\vec{x_i}$ related to nonzero values of $\left(a_i - a_i^*\right)$ are called SVM. The following function 17 presents the Radial Basis function (RBF) kernel

$$f\left(\vec{x}, a, a^*\right) = \sum_{i=1}^{p} \left(a_i^* - a_i\right) \exp\left\{ \frac{-\left\|\vec{x} - \vec{x_i}\right\|^2}{2 \cdot \sigma_{RBF}^2} \right\} + b \tag{17}$$

The involved parameters $\left\{\sigma_{RBF}, \gamma, \varepsilon\right\}$ are very significant and they determine to a great level of extend the good fit of the ε-SVR. It should be specified that σ_{RBF} is the RBF kernel's standard deviation, γ has a constant value that determines the point where the empiric error is related to complexity and finally the parameter ε is the width of the ε-zone. In statistics, when we consider the Gaussian probability density function σ_{RBF} is called the *standard deviation*, and the square of it, the variance.

2.4. Fuzzy weighted Support Vector Regression

The fuzzy weighted SVR with a fuzzy partition first employs the fuzzy c-mean clustering algorithm to split training data into several training subsets. Then, the local-regression models (LRMs) are independently obtained by the SVR approach for each training subset. Finally, those LRMs are combined by a fuzzy weighted mechanism to form the output. Experimental results show that the proposed approach needs less computational time than the local SVR approach and can have more accurate results than the local/global SVR approaches does [5].

3. Application

In this section of the paper, the application results of the Global SVR approach for all existing data vectors will be presented thoroughly. The regression was done for different kernels, namely the Gaussian (RBF) the Fuzzy weight SVM and the Polynomial one. In order to perform Global SVR a global regression model was developed.

The regression was performed by the use of the LIBSVM v2.9 (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) [4] which is encoded in C++ and offers a Matlab Interface. In this specific application the RBF-Kernel was applied.

3.1. Performing trial and error for the Gaussian Kernel

The initial data set comprised of one hundred forty-four (144) sets of experimental data, where the ε'' was measured under different temperature and moisture conditions and for various types of chemical composition.

During Training and Testing, several hundreds of experiments were performed in order to determine the optimal parameters' values.

More specifically in the case of the Gaussian (RBF) Kernel, 400 experimentation cases (applications of the 5-fold cross validation) were constructed with the integer value of σ_{RBF} ranging from 1 to 20 and with the value of the parameter γ ranging from 1 to 14. In the same case the e parameter took real values from the following set {0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.1, 0.12, 0.14, 0.16, 0.18, 0.20, 0.22} This means that totally 400*14=5600 trials were performed during the 5-fold cross validation process.

The Root Mean Square Errors (RMSE) produced in each Training or Testing cycle were stored and then they were averaged each time 5-fold cross validation was employed. In each trial the initial data set was divided into five subsets (4 training and one testing) where five training and testing cycles were performed (each time changing the testing set in a round robin manner). Totally 5600 Average Root Mean Square Errors were obtained (ARMSE) one for each 5-fold cross validation cycle and for each unique combination of the parameters, together with the corresponding % Mean Average Percent Error (MAPE).

In the case of the Gaussian Kernel the optimal SVM model was produced for the $\sigma_{RBF} = 2$ and for the value of the parameter $\gamma=12$ in the case with e=0.04 as it is shown in the following table 1.

Table 1: Performance of the Optimal Gaussian Kernel

Optimal Gaussian Kernel SVM (5-fold cross validation)

Value of σ_{RBF}	Value of γ	Value of e	Mean RMSE Training	Mean RMSE Testing	MAPE Training	MAPE Testing
2	12	0.04	0.00009664	0.0001501	2.4824596	3.7328347

The following figures 2 and 3 present the evolution of the Mean RMSE and of the MAPE respectively, according to the values of the e parameter for the Gaussian Kernel. It is clearly shown that the best model should have an e value equal to 0.04.
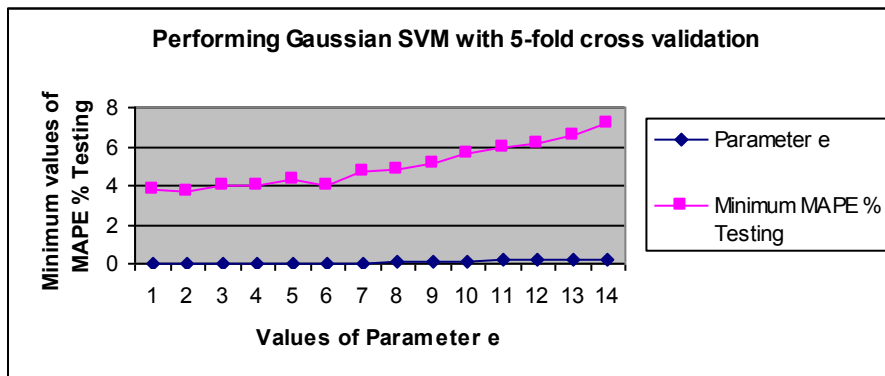

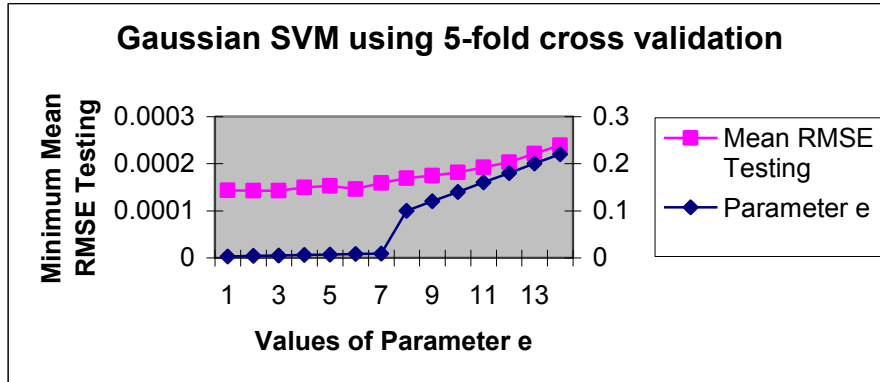**Figure 2. Evolution of the Mean RMSE for different values of the parameter**

Figure 3. Evolution of the MAPE for different values of the e parameter

For the case of the Polynomial Kernel, the σ_{RBF} again took integer values from 1 to 20 and the parameter γ from 1 to 14. However the e parameter took real values from the following set {0.005, 0.1, 0.15, 0.20, 0.25}. The performance of the best fit Polynomial Kernel is shown in the following table 2 and it was achieved for $\sigma_{RBF=}3$ $\gamma=18$ and e=0.1

Table 2: Performance of the Optimal Polynomial Kernel

Optimal Polynomial Kernel SVM (5-fold cross validation)

Value of σ_{RBF}	Value of γ	Value of e	Mean RMSE Training	Mean RMSE Testing	MAPE Training	MAPE Testing
3	18	0.1	0.000485	0.000549	11.74819	14.01062

For the Fuzzy weight Support Vector Machine the σ_{RBF} again took integer values from 1 to 20 and the parameter γ from 1 to 14. Also the e parameter took real values from the following set {0.005, 0.1, 0.15, 0.20, 0.25}. The performance of the best fit Fuzzy weight SVM is shown in the following table 3 and it was achieved for $\sigma_{RBF=}3$ $\gamma=16$ and e=0.05

Table 3: Performance of the Optimal Fuzzy weight SVM

Optimal Fuzzy weight SVM (5-fold cross validation)

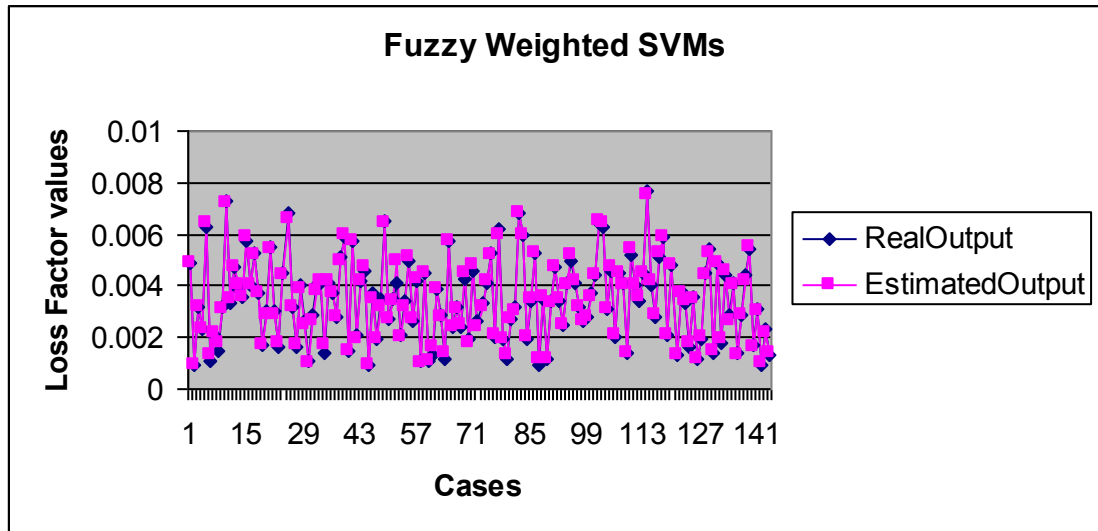| Value of σ_{RBF}	Value of γ	Value of e	Mean RMSE Training	Mean RMSE Testing	MAPE Training	MAPE Testing
3	16	0.05	0.000488	0.000542	12.13017	14.12399

Figure 4. Actual versus estimated values of Loss Factor for Fuzzy Weighted SVMs

4. Discussion – Comparative analysis

The results of the e-SVM regression were compared to the results of the Artificial Neural Network that was developed by our research team in a previous study [2] and also the results of the three SVM kernels were compared to each other.

In [2] the optimal ANN that estimated wood loss factor for the same input variables was found to be a multi layer Back propagation one, with Sigmoid transfer function. It was found to have eleven (11) processing elements in the input layer, nine (9) neurons in the hidden layer and of course one neuron in the output layer [2].

Table 4: Performance of the optimal ANN for the estimation of loss factor in wood

Optimal ANN RMSE Training	Optimal ANN R^2 Training	Optimal ANN RMSE Testing	Optimal ANN R^2 Testing
0.017	R^2=0.9989	0.0382	R^2=0.9945

The above table 4 presents the performance of the best fit ANN. It is clearly shown that the Gaussian Kernel approach has by far much lower RMS error than the Polynomial and the Fuzzy weight SVM and also than the best fit ANN. Thus it can be considered the most reliable modeling approach.

This research effort is quite new and processes a high degree of innovation due to the fact that similar research has not been performed in wood science literature before. It is the final part of a wider effort to produce rational and useful Soft Computing regression models towards the estimation of wood dielectric properties.

It has been clearly shown that all three Kernel approaches and the ANN model are reliable and they can be used alternatively in wood industry applications. However the Radial Basis function (Gaussian) Kernel has proven to produce a much more promising model. Further experiments will be performed to produce more data vectors and complimentary modeling efforts will be performed in the future.

References:

1. Avramidis S., (1999). Radio Frequency Vacuum Drying of wood. Proceedings of the International conference of cost Action E1 5 Wood Drying. Theme State of the Art for Kiln Drying

2. Avramidis S., Iliadis L., Mansfield S. (2006) Wood Dielectric Loss Factor Prediction with Artificial Neural Networks Wood Science and Technology ISSN: 0043-7719 Volume 40, Number 7, pp. 563-574. Journal of the International Academy of wood Science. Springer Berlin Heidelberg

3. Boser B., Guyon I., and Vapnik V.: A training algorithm for optimal margin classifiers In Fifth Annual Workshop on Computational Learning Theory, pages 144--152, Pittsburgh, ACM (1992)

4. Chang C.C., Lin C. J.(2009). LIBSVM A Library for Support Vector Machines, on line at:
http://www.csie.ntu.edu.tw/~cjlin/libsvm/

5. Chuang C. C. (2007). Fuzzy Weighted Support Vector Regression with a Fuzzy Partition Systems,Man, and cybernetics, Part B: Cybernetics, IEEE Transaction on. Vol 37, Issue 3. pp. 630-640

6. Davy M. (2005). An Introduction to Support Vector Machines and other kernel algorithms, Proc. of the 9th Engineering Applications of Neural Networks Conference, Lille, France

7. Efron, B. and Tibshirani, R.J. (1993), An Introduction to the Bootstrap, London: Chapman & Hall

8. Hjorth, J.S.U. (1994), Computer Intensive Statistical Methods Validation, Model Selection, and Bootstrap, London: Chapman & Hall

9. Kabir M. F., Daud W. M., Khalid K. b., Sidek H. A., (2001). Temperature dependence of the dielectric properties of rubber wood. Wood and Fiber Science. V. 33(2)

10. Kecman V., (2001). Learning and Soft Computing. MIT Press, USA

11. Koumoutsakos, A., Avramidis, S., and S. Hatzikiriakos. 2001a. Radio frequency vacuum drying Part I. Theoretical model. Drying Technol. 19(1):65-84.

12. Koumoutsakos, A., Avramidis, S., and S. Hatzikiriakos. 2001b. Radio frequency vacuum drying Part II. Experimental model evaluation Drying Technol. 19(1): 85-98.

13. Koumoutsakos, A., Avramidis, S., and S. Hatzikiriakos. 2003. Radio frequency vacuum drying Part III. Two dimensional Model, Optimization and

Validation. Drying Technol. 21(8): 1399-1410

14. Martin P., R. Collet, P. Barthelemy, G. Roussy. (1987). Evaluation of wood characteristics: Internal scanning of the material by microwaves. Wood Sci. Technol. 21:361–371

15. Norimoto, M. 1976. Dielectric properties of wood. Wood Research. No. 59/60: 106 - 152.

16. Norimoto, M., Hayashi, S. and T. Yamada. 1978. Anisotropy of dielectric constants in coniferous wood. Holzforschung 32: 167-172.

17. Shao, J. and Tu, D. (1995), The Jackknife and Bootstrap, New York: Springer-Verlag

18. Siau, J.F. 1995. Wood: influence of moisture on physical properties. Department of Wood Science and Forest Products, VPI&SU

19. Torgovnikov, G. I. 1993. Dielectric properties of wood and wood-based material. Springer-Verlag, Berlin Heidelberg, New York

20. Vapnik V.N., S. Golowich, Smola A (1997).: Support Vector method for function approximation regression estimation and signal processing, In: Advances in Neural Information Processing Systems, vol. 9, Cambridge, MA:MIT Press, 281–287 (1997)

21. Weiss, S.M. and Kulikowski, C.A. (1991), Computer Systems That Learn, Morgan Kaufmann

22. Zhou, B., and S. Avramidis. 1999. On the loss factor of two B.C. softwoods. Wood Sci. and Technol., 33(4):299-310