

# **Tensile Strength Estimation of Lumber**

By

Hossein S. Saboksayr

B. A. Sc. (Electrical Engineering) Sharif University of Technology, 1988

M. A. Sc. (Electrical Engineering) Concordia University, 1995

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES  
(DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING)

We accept this thesis as conforming  
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2001

© Hossein S. Saboksayr, 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical & Computer Engineering

The University of British Columbia  
Vancouver, Canada

Date August 31, 2001

## **Abstract**

The problem of wood tensile strength estimation of softwood lumber is studied in this thesis. The main contributions brought to this topic here are first, a set of knot geometry features that can be used in board strength estimation, and second a learning algorithm that selects the best set of features for the purpose of strength measurement.

The estimation problem is posed as an empirical learning problem that is based on the measured properties of wood. The process of producing the required database consisted of three distinct tasks: selecting and preparing the boards, measuring a set of properties of wood for every board, and estimating the measured strength of each board from the measured profiles.

A set of boards, providing a random sample of softwood lumber, already existed at UBC (from previous experiments). These boards were measured and used as the preliminary database. A second set of boards was selected randomly from the regular production of softwood lumber. These boards created the evaluation data set.

For the measurement task, all the boards were scanned using the available measurement machines. These machines were SOG and Microwave for grain angle measurement, X-ray for local density measurement, dynamic bending machine for the Modulus of Elasticity measurement, as well as the ultimate tensile strength tester for measuring the tensile strength of a board. The output profiles per board were saved in a data file (one data file per board per machine). The measured data files were stored in a database consisting of a structure of directories.

In the strength estimation task all the measured profiles of a board were mapped to specific features (usually statistical moments) and the features were then mapped to the strength of the board. One of the features of a board is the set of its knots. A conic model of a knot was chosen and the related mappings were developed such that the X-ray scanning could be used in order to detect the existence, location, and shape of knots in a board. Then geometrical features were proposed such that the set of knots of a board could be transformed into a set of features suitable for strength estimation methodology of this thesis.

Since specimens are costly to measure, means to reduce the number required were developed. To this end statistical learning theory was applied. This theory addresses the suitability of the learning model for the physical problem and the effectiveness of the features for the estimation problem. Based on this theory, the ASEC learning model was developed.

The learning problem for wood tensile strength estimation was divided into three problems: defining the most suitable feature set, measuring the suitability of a learning machine, and using the *a priori* knowledge about the dependence in the learning machine. A method for measuring the suitability of a regression estimator (VC-dimension) was developed in order to select the best model in a class of models. The ASEC learning model was developed in order to find the best set of new features from the given feature set by using the known dependencies.

Different learning machines were tested in order to determine what model is most suitable for tensile strength estimation of lumber. The validity of all the methods was demonstrated by analytical proof, by simulation, or by test on the database.



# Table of Contents

Abstract .....	ii
List of Tables .....	viii
List of Figures .....	ix
Nomenclature .....	xii
Acknowledgement .....	xvi
 Chapter 1. Introduction .....	 1
1.1. Objective and limitations .....	1
1.2. The paradigm .....	2
1.3. Specimens .....	2
1.4. An overview of the experimental work .....	4
1.5. An overview of the analysis .....	6
1.6. Review of other attempts to transform measured mechanical properties into strength .....	7
1.7. Economic importance of this knowledge .....	10
1.8. Why Geometrical features are important .....	10
1.9. Outline of this thesis .....	11
1.10. The Contributions of this thesis .....	12
 Chapter 2. Measurement means; practical constraints .....	 13
2.1. Measurement of the Physical Properties of Wood .....	14
2.1.1. Modulus of elasticity of wood .....	14
2.1.2. Moisture content .....	15
2.1.3. Dielectric Properties .....	16
2.1.4. Density of wood .....	17
2.2. Measuring density with an X-ray machine .....	17

2.3. Measuring Grain Angle with SOG and Microwave Machines .....	22
2.3.1. Grain angle via an SOG machine .....	22
2.3.2. Grain angle measurement via microwave .....	25
2.4. Modulus of Elasticity using the Continuous Lumber Tester .....	27
2.4.1. Calibration .....	28
2.5. Ultimate Tensile Strength Tester .....	36
 Chapter 3. Modeling wood failure .....	37
3.1. Tensile strength of clear wood as a random variable .....	38
3.2. The effect of grain angle on strength .....	39
3.3. The effect of density variation on strength .....	40
3.4. The effect of holes on strength .....	41
3.5. The effect of knots on strength .....	41
3.6. Other factors .....	42
3.7. A 2x4 board as a system .....	43
3.8. Proposed approach to modeling strength as a continuous function of wood properties .....	44
 Chapter 4. Characterizing a board; features .....	47
4.1. A general view on defining the features .....	49
4.2. Properties of clear wood .....	51
4.3. Characterizing the defects .....	55
4.4. Geometrical features .....	57
4.5. Characterizing feature sets are not unique .....	59
4.6. Board model .....	60
 Chapter 5. Extracting features from measured profiles .....	62
5.1. The image of a knot in an X-ray image; the shadow image .....	62

5.2. Thresholding; classifying a pixel into knot or clear wood classes .....	64
5.2.1. Bayesian classification .....	66
5.2.2. Edge detection .....	68
5.2.3. Experimental data .....	70
5.3. Indexing the pixels of a knot .....	74
5.4. Knot pattern .....	75
5.4.1. Knot model parameters .....	78
5.4.2. Measuring and calculating the features of the pattern .....	78
5.4.3. Matching patterns .....	82
5.5. Geometrical features of a knot .....	82
5.6. X-ray image analysis; extraction of board image and defect detection.....	83
5.6.1. Edges of the board .....	84
5.6.2. Knot detection .....	85
5.6.3. Three dimensional knot shape detection (simulation) .....	90
5.7. Knot detection for real data .....	96
5.8. X-ray and SOG interrelation .....	100
5.9. Sufficiency of sampling rate; the problem of resolution .....	103
5.10. Preparation of specimens and measurement process .....	105
5.11. Specimen Characterization .....	112
 Chapter 6. Strength Estimation; Empirical Learning for Small Sample Size .....	 120
6.1. Strength estimation as an empirical learning problem .....	122
6.2. High dimensionality of the feature set .....	130
6.3. The relationship between estimation accuracy, learning capacity, and sample size .....	132
6.4. Extending the learning-capacity (VC-dimension) to regression estimators .....	135
6.4.1. The equivalence of classifiers and regression estimators .....	135
6.4.1.1. Learning problem statement .....	135
6.4.1.2. Constructing a classifier by using a regression estimator .....	137

6.4.1.3. Constructing a regression estimator by using a classifier	138
6.4.2. The VC-dimension of a regression estimator	142
6.4.3. Corollary	143
6.4.4. Measuring the VC-dimension at a point	143
6.5. Strength estimation by using existing methods	149
6.5.1. Linear regression	149
6.5.2. K-Nearest Neighbor and radial basis functions	150
6.5.3. Multi-layer neural networks	151
6.5.4. SPORE	153
6.6. Alternating Space Expansion and Contraction (ASEC)	154
6.7. Estimation accuracy of different learning machines	163
6.8. Discussion of the results	165
 Chapter 7. Conclusions and Future Directions	 167
7.1. Conclusions	167
7.2. Future Directions	168
 Appendix: Board failure; initiation and propagation of a fracture	 170
 References	 174

List of Tables

Table 1: The average output, average MOE, and transformation coefficients for selected specimens. .... 31

Table 2: The number of output signals of measurement machines ..... 108

Table 3: Specimen file names ..... 109

Table 4: The measured VC dimension of a linear regression estimator. .... 146

Table 5: The parallel feature sets. .... 157

Table 6: The computed after each iteration of ASEC ..... 162

Table 7: The accuracy of estimation by using different learning machines. .... 164

Table 8: The accuracy of the estimation for different feature sets. .... 165

## List of Figures

Figure 1.1. Wood strength estimation problem paradigm; different stages of the learning problem.	2
Figure 1.2. Different Specimen sets that were produced during the course of this research.	4
Figure 2.1. Three-point and four-point bending test.	15
Figure 2.2. The mechanism of a Continuous Lumber Tester	15
Figure 2.3. A simple diagram of the X-ray system and the shadow image	18
Figure 2.4. The profile of one cross-section of the board.	19
Figure 2.5. The shadow of an X-ray image.	20
Figure 2.6. The shadow image of a knot.	20
Figure 2.7. The contour map of an X-ray image.	22
Figure 2.8. The rotating head for grain angle measurement.	23
Figure 2.9. A measured grain angle profile of a board.	24
Figure 2.10. Microwave System for measuring wood grain angle	25
Figure 2.11. The arrangement of the probes (left) and the structure of a probe (right).	27
Figure 2.12. Typical output profiles of a dynamic bending machine.	28
Figure 2.13. The extracted and adjusted dynamic bending machine output profile for specimen #0760	29
Figure 2.14. Measured static MOE of specimen #0760	30
Figure 2.15. Linear relationship between the average measured dynamic bending machine output	32
Figure 2.16. Measured MOE profiles by and static bending methods.	34
Figure 2.17. The repeatability of the measured profiles for machine.	35
Figure 2.18. The repeatability of the measured profiles for machine (continued).	36
Figure 3.1. Strength estimation as a learning problem.	45
Figure 4.1. Steps of strength estimation method.	50
Figure 4.2. The histogram representation of density distribution of a board	53
Figure 4.3. Grain angle profile of a board and its distribution.	54
Figure 4.4. The MOE profile of a board and its distribution.	55

Figure 4.5. The local density profile of the board generated by a moving window.	57
Figure 4.6. The density profile of a knot as compared with the density of clear wood of a board.	58
Figure 4.7. Conic model of a knot.	58
Figure 4.8. The model of a board that consists of the clear wood and defects.	60
Figure 5.1. The X-ray image of a knot	63
Figure 5.2. Contour map of an X-ray image.	64
Figure 5.3. Probability distribution of clear wood density and knot wood density.	66
Figure 5.4. Density distribution of X-ray image pixels of a specimen	73
Figure 5.5. Sampling pixels of a knot from an X-ray image.	74
Figure 5.6. A search for neighboring knot pixels.	75
Figure 5.7. A knot pattern (from its image)	76
Figure 5.8. Two orthographic projections of a knot.	77
Figure 5.9. The pattern of a knot.	78
Figure 5.10. The approximate image of the knot.	79
Figure 5.11. The four points of the pattern	80
Figure 5.12. The flow chart for knot pattern analysis.	81
Figure 5.13. The output of the X-ray system with and without the sample board.	84
Figure 5.14. Output of the X-ray imaging machine.	85
Figure 5.15. The edge of all knots of a board detected by fixed Bayesian based thresholding.	86
Figure 5.16. The edge of all knots of a board ...	86
Figure 5.17. Knot detection by using Bayesian thresholding. Specimen 130.	87
Figure 5.18. Knot detection by using Bayesian thresholding. Specimen 70.	87
Figure 5.19. The top view X-ray image and discretized X-ray image of a knot.	88
Figure 5.20. The side view X-ray image and discretized X-ray image of a knot.	89
Figure 5.21. The diagram of a knot cone and its parameters.	92
Figure 5.22. Contour diagrams showing the simulated top X-ray image ...	94
Figure 5.23. Estimation accuracy for the knot diameters.	95
Figure 5.24. The simulated knot pattern that is partially stretching outside the board.	96

Figure 5.25. The result of applying the knot detection algorithm to the X-ray image.	97
Figure 5.26. The simulated X-ray images.	98
Figure 5.27. The location and size of knots in CCD camera images.	99
Figure 5.28. Most of the knots show a grain variation in the grain angle profile.	101
Figure 5.29. Visual image, X-ray image, SOG profile, and MOE profile.	102
Figure 5.30. The measured signal and its Fourier-transform.	103
Figure 5.31. The measured signal and its Fourier-transform of a typical SOG profile.	104
Figure 5.32. The database structure.	111
Figure 5.33. The coordinate of a knot in a board	113
Figure 5.34. The general flow chart of geometrical feature extraction program	115
Figure 5.35. The flow chart of the knot extraction routine	116
Figure 5.36. The flow chart of knot extraction routine 2	117
Figure 5.37. The flow chart of the statistical feature extraction program	118
Figure 6.1. The learning problem	123
Figure 6.2. Distance of all other boards from a sample board.	131
Figure 6.3. The block diagram for estimating the VC-dimension	147
Figure 6.4. A multi-layer neural network.	152
Figure 6.5. The cascade form of SPORE-1.	153
Figure 6.6. Two iterations of ASEC	155
Figure 6.7. Estimation accuracy from the contracted space (top) and the empirical risk (bottom).	158
Figure 6.8. The curve of the upper bound of the risk functional for the given number of features.	159
Figure 6.9. Replacement of close features in the trajectory table.	161
Figure 6.10. The flow chart of one cycle of the ASEC	163



# Nomenclature

A	Minimum of a bounded function
a	Fracture surface
$a_i$	Area of either half of a knot pattern
B	Maximum of a bounded function
CART	Classification And Regression Tree. A learning method
CCD camera	A Charge-Coupled-Device digital camera
CLT™	Continuous Lumber Tester, a machine for measuring the dynamic modulus of elasticity of a board
$C(j i)$	The cost of misclassifying an element of class 'i' into class 'j'
c	Speed of light
D	Deflection (m) of a board in a bending test
d	Diameter of a knot at the center of the knot
$d_i$	Diameter of a knot at either end of it
$d(.,.)$	Distance between two points
E	Modulus of elasticity Also, energy of matter
$E(.)$	Expected value of a random variable
F	The feature space that represents $\hat{X}$
<i>feature</i>	A property of a profile, e.g., the average of a profile
$\hat{F}$	The physically meaningful set of features.
$F(.)$	The cumulative distribution function
$f(.)$	An approximating function
$f_{ij}$	i-th feature that is selected in j-th iteration of subset selection in ASEC
$g(.)$	A function
h	A measure of the degree of freedom of a data set Also, the learning capacity of a machine
$h(.)$	An element of the approximating function set, in a dictionary based learning machine
I	Intensity of an electromagnetic wave Also, moment of inertia ( $m^3$ ) of an object Also, size of a subset that is selected in an iteration of ASEC
$I_0$	The initial intensity of electromagnetic wave before attenuating in matter
KAR	Knot Area Ratio, the worst ratio of a knot cross-section to the board cross-section
$K_I$	The stress intensity factor for opening mode of failure

$K_{Ic}$	The critical stress intensity factor for opening mode of failure
$K_{II}$	The stress intensity factor for sliding mode of failure
$K_{IIc}$	The critical stress intensity factor for sliding mode of failure
$K_{III}$	The stress intensity factor for tearing mode of failure
$L$	The span ( $m$ ) of the bending test
$L(.)$	A loss function, based on which risk functional is defined
$\lambda$	Total number of specimens in a training set
$\lambda_i$	Distance of the center of a knot pattern from either end
$M(.)$	The Fourier transform of the MOE profile
MEL	Machine Evaluated Lumber. A lumber-grading standard
MOE	Young's Modulus of Elasticity
MOR	Modulus Of Rupture. The bending stress level that causes board to break
MSR	Machine Stress Rated; A lumber-grading standard
$m$	Mass of matter Also, the mean of a probability distribution function
$m(.)$	The measured MOE profile
$m'(.)$	The measured MOE profile with zero mean
$m_i$	Mean of the probability distribution function of class ' $i$ '
$N$	Number of pixels in a class
$n$	Number of samples or specimens
$n_i$	Number of samples or specimens in class ' $i$ '
$o$	The center of a knot pattern
$P$	The load ( $N$ ) in deflection of a board.
<i>profile</i>	A measured property in the form of a one dimensional signal or a two dimensional image
$p_i$	Any point on the knot pattern
$p_n$	A knot pixel in a generated cube
$p_c$	Projection of a knot pixel on the central line of a generated cube
$p_i(x)$	The probability distribution of ' $x$ ' if it belongs to class ' $i$ '
psi	Pound per Square Inch
$Q(.,.)$	A loss function
$q_i$	Probability of class ' $i$ ' happening
$R$	The complete space of physical features
$R_i$	A region in the measured variable $X$
$R(.)$	The risk functional, the minimization of which is the goal of training a learning machine
$R_{emp}(.)$	Empirical risk functional, by minimizing which a learning machine is trained
$R_{emp}^c(.)$	Empirical risk functional of classifiers

$R_{emp}^r(.)$	Empirical risk functional of regression estimators
$R_{emp}^{cr}(.)$	Empirical risk functional of regression estimators that are made up of a group of classifiers
$r$	Distance (in a two dimensional polar coordinate) of a material element around the tip of a fracture
$r^2$	Coefficient of determination ( $r^2(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$ ). An estimation accuracy measure
$r(.)$	Penalization function
$S$	The tensile strength of a board
$S_{Ci}$	The set of samples that belong in class $i$
$S_i$	A set of approximating functions with fixed complexity
$\hat{S}$	Estimated tensile strength
$S(.)$	The Fourier transform of the SOG profile
SOG	Slope Of the Grain. Slope of the grain of wood
SPORE	A learning method for high dimensional feature space
SPF	Spruce, Pine, and Fir
$s$	standard deviation of a distribution
$s(.)$	The measured SOG profile
$T_i$	An iteration of subset selection in ASEC
$t$	Thickness of a specimen
$U_p$	Potential energy in a board due to stress
$U_s$	Surface energy of a fracture
$w$	The set of parameters of an approximating function
$w^*$	Optimum set of parameters of an approximating function
$X$	The measurable space of physical features Also, the measured density (the value) of a pixel of an X-ray image
$\hat{X}$	The space of all the measured profiles
$X_0$	A single point in the feature space
$x$	The input to a learning machine
$Y$	The vector of outputs
$y$	The output of a learning machine
$\hat{y}$	A vector
$z_i$	A sample point
$z$	The sample space, the set of all sample points
$Z_i^{2\lambda}$	A set of $2\lambda$ specimens
$\Phi(.)$	Penalization functional, which defines the complexity of an approximating function

$\Delta$	The difference between two consecutive threshold levels
$\Delta(.)$	The difference between practical and ideal risks
$\Gamma_T^B$	A bootstrap sample from a training set
$\Lambda$	Population of all the image pixels
$\phi(.)$	The upper bound error for empirical risk functional
$\eta$	A real number between zero and one
$\alpha$	The angle between two vectors Also, a set of parameters of a learning machine
$\alpha_\lambda$	Optimum set of parameters resulting from a training set of size $\lambda$
$\alpha_0$	Optimum set of parameters resulting from minimizing the risk functional
$\mu$	The location parameter in an approximating function Also, mass absorption coefficient ( $1/cm$ )
$\sigma$	The scale factor in an approximating function Also, the stress level in a board Also, the square root of variance of a distribution
$\sigma_c$	The critical stress level
$\lambda$	The shape factor in an approximating function Also, the regularization parameter in an empirical risk functional
$\gamma_s$	Elastic surface energy per unit thickness
$\theta$	Angle (in a two dimensional polar coordinate) of an element around the tip of a fracture
$\rho$	Mass density ( $g/cm^3$ )
$\varepsilon$	Total permittivity of matter Also error in strength estimation
$\varepsilon(\lambda)$	Error in empirical risk functional due to the size of the learning set
$\varepsilon'$	The permittivity of matter
$\varepsilon''$	The loss factor in permittivity of matter
$\pi_i$	Class (clear wood or knot wood) of an image pixel
$\omega_i$	The output of a classifier
$\xi(Z_i^{2\lambda})$	Maximum variation in empirical risks of $\lambda$ samples, in a set of size $2\lambda$
$v^\lambda$	Empirical risk of a set of size $\lambda$ , in VC dimension detection process

## **Acknowledgements**

I would like to acknowledge and express my appreciation for the support, patience, and continuous encouragement of my supervisor Peter Lawrence, which has contributed a great deal to the completion of this thesis. I would also like to acknowledge the help and cooperation of Frank Lam (Co-supervisor), David Barrett, Gary Schajer, Greg Grudic, Jacek Biernacki, Carl Flatman, Wilson Lau, Glen Trarup, and Bob Myronuk,

All the required measurements were done in Wood Products Laboratory (Wood Science Department of UBC) as well as our industrial collaborators. The cooperation and help from the laboratory coordinators and researchers is also acknowledged.

I also acknowledge my friendly roommates (in CICS), Mr. Philip K. Liu and Mr. Mohamed K. Cherif. The relaxed atmosphere of my workplace has definitely had a positive effect on my work.

## **To my parents**

for their unlimited love and support

# ***Chapter 1. Introduction***

## **1.1. Objective and limitations**

The goal of this research is to find a suitable method for estimating a board's tensile strength by measuring the important-known material properties of wood. As will be shown, this problem inherently leads to a series of studies of wood structure, wood properties measurement, feature definition, learning systems, and estimation accuracy analysis. This thesis is an attempt to tackle this problem on all levels to produce a coherent analysis of the problem.

From a practical point of view, the problem of strength estimation can be divided into the following problems: collecting a set of specimens, selecting measurement means, defining and extracting features from the measured signals, and estimating the strength by using the features of a specimen.

Strength estimation can be cast as a learning problem. To do so, one needs different parallel measurement profiles of different wood properties and must define a variety of different features, in order to reduce the chance of having any un-modeled factors. Therefore, a high dimensional feature space is needed for strength estimation. In the end, the problem of wood strength estimation is transformed into a learning problem in a high dimensional (more than 20 dimensions) feature space.

The practical restrictions can be considered as conditions of the learning problem. There are potentially three major restrictions to the application of the resulting strength-estimating model. The first restriction is, that because of measurement difficulty, the number of specimens for the database will be limited, which could make the database unsuitable for some learning systems. The second restriction is that properties of each species can vary depending on the region of harvest, which means that any single model cannot be claimed as a general model of wood strength. Finally, the production lines may have different measurement means that have demonstrated reliability over time.

Therefore, many asymptotic learning methods may not be suitable for strength estimation and various learning methods should be explored in order to get the best result. Also, this research does not claim generality for the produced model to many species and wood harvesting regions. However, since the database of this thesis consists of softwood lumber produced here in British Columbia (from here on B.C.) the strength estimating model is suitable for B.C.

## 1.2. The paradigm

The problem of wood strength estimation can be summarized as shown in Figure 1.1. The problem can be divided into four major tasks; specimen selection, measurement, feature extraction, and model development. Each part of the problem is briefly explained in the following.

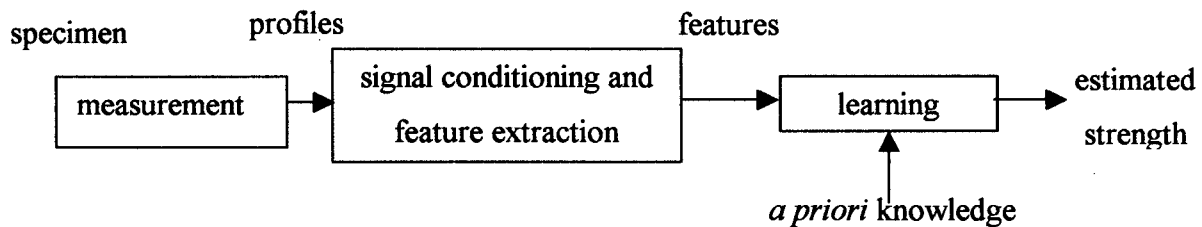


Figure 1.1. Wood strength estimation problem paradigm; different stages of the learning problem.

## 1.3. Specimens

A model is only representative of the database used to formulate it. For this study, specimens were selected at random from normal production in a regular mill. In particular, four bundles of regular production from two different mills (two bundles from each) were shipped to the laboratory for measurement and destructive testing. (The boards with obvious fractures were removed.) Each board was assigned a number that



was written on the cross-section of one of its ends. This number was used to code all the data files that would be related to the same board and was kept constant throughout the project. Also, the same number was used to reference the specimen in the final feature set.

During the course of this research, two independent data sets were produced for the pilot study and the evaluation study, respectively. The pilot study data set consisted of 800 boards and was further divided into two separate sets: set one was used to develop a set of features and the boards in this set were destructively tested for tensile strength; set two was used as a repetition of the first set as well as for comparing the measurement machines in the laboratory with the measurement machines used in the industry. A description of all the measurement machines and their limitations will be presented later in the related sections.

In the evaluation study, presented in Sections 6.5 to 6.8, 1100 boards were selected from local mills. Similar to the boards in the pilot study, the boards in the evaluation study were a random sample of the produced boards. All the measurements were done using industrial machines and the tensile strength of the specimens were measured at the University of British Columbia (from here on, UBC). The following diagram shows the specimen sets and the related measurement machines.

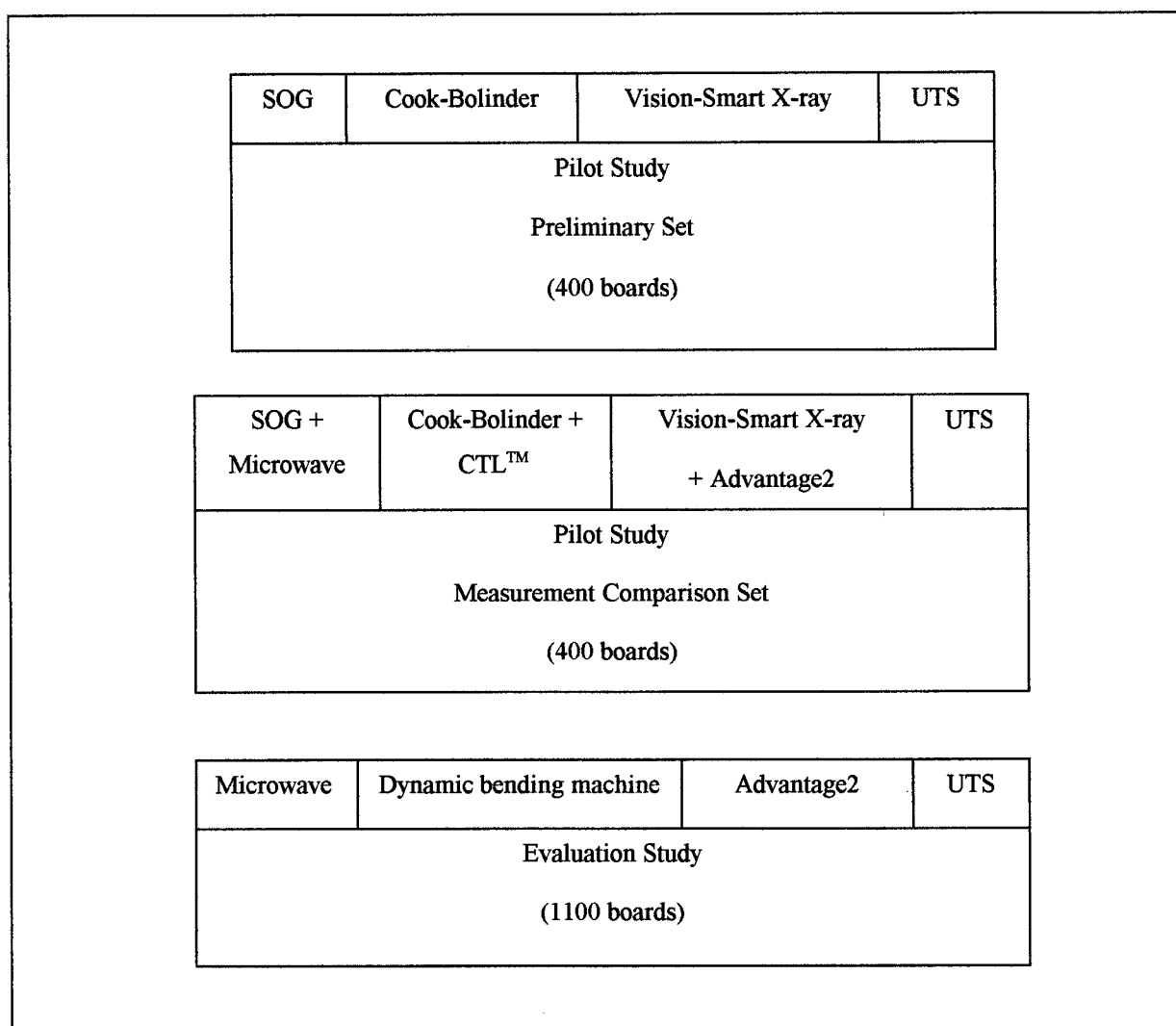


Figure 1.2. Different Specimen sets that were produced during the course of this research. The SOG and Microwave machines measure the local grain angle of wood. Cook-Bolinder and CTL™ machines measure the flatwise MOE of the board. Vision-Smart X-ray and Advantage2 measure the local density. UTS is for ultimate tensile strength measurement.

#### 1.4. An overview of the experimental work

The measurement means will be discussed in detail in Chapter 2. Briefly, they consisted of the following systems.

For X-ray, a Vision Smart X-ray machine (at UBC Wood Products Laboratory) and a Newnes Advantage2 were used. The X-ray machines ([1]) produced two-dimensional

density images of each board. Two images of the same board, one from the top and one from the side, were produced. The Vision Smart X-ray machine was used for the pilot study and was set in the Wood Products Laboratory of UBC. The Newnes Advantage2 X-ray machine was used for the evaluation study. The data produced from the two machines were similar, but the Newnes Advantage2 database was used in the evaluation study because of the higher resolution and because its conveyor had a more consistent speed.

For SOG, the capacitance based heads, at UBC Wood Products Laboratory, and the microwave machine were used. These machines measured the local grain angle of wood along the board. The grain angle measurements were very similar between the two machines therefore the microwave measurements were used for strength estimation in evaluation study.

For MOE, Cook-Bolinder MOE tester (at UBC Wood Products Laboratory) and Metriguard CLT™([4]) machine (at our industrial collaborator), also called dynamic bending machine here, were used. Cook-Bolinder MOE tester measured the flatwise Modulus of Elasticity (MOE) profile of a board in a three-point test (see Figure 2.1) and measured the MOE profile of one side of the board in one scan. A Metriguard CLT™ machine is commonly used for wood grading and produces the two MOE profiles by one pass of the specimen through the machine (see Section 2.1.1). As will be shown in Section 2.4.1, the CLT™ machine is repeatable and its output is highly correlated to static MOE (see Section 2.1.1). Therefore, in the evaluation data set, only the CLT™ profile was measured and used for strength estimation.

The ultimate strength tester (at UBC Wood Products Laboratory) was used to get the tensile strength of each board in both the pilot study and evaluation study. In the UBC Wood Products Laboratory, all of the different types of measurements were carried out independently. Once the measurement of one profile of all the boards was finished, the laboratory was prepared for the next measurement.

For each measurement, as each board was fed into the measurement machine its number was registered in a table. The measured profile of each board was saved in a different data file. Once an experiment (e.g. an X-ray measurement) was over, all the data files were renamed so that the file name represented the name of the measurement (e.g. X-ray)

and the board number (e.g. 0114). The calibration files (if any) were kept with the data files in a directory to be used in a feature extraction program.

### **1.5. An overview of the analysis**

As was stated above, the selected specimens limited the application of the model. Since the sample boards were from BC, the strength estimation model is suitable only for softwood production in local industry. The limitations of the model notwithstanding, the best choice of measured properties for model development is that they produce the minimum amount of uncertainty for board characterization. Measurement selection and feature extraction can be viewed as attempts at specimen characterization, such that boards with different features are mapped to different points in the feature space. Since the strength of a board is a random variable in this space, good feature definition minimizes the variance of this random variable at a given point.

Each specimen is characterized by a set of features, which are the statistical or the geometrical transformations of the measured profiles. The total of 64 features will be used to characterize each specimen of the database of this thesis. Each one-dimensional profile will be transformed into eight statistical features and the combination of the two X-ray images will be transformed into 16 features.

There will be three MOE profiles. Of these three, two profiles are the measured profiles and the third profile is the average profile of the two measured profiles (at every point along the board). Therefore, there will be 24 features from the measured MOE profiles.

The measured grain angle profiles will be grain angle, grain alignment, longitudinal signal amplitude, transverse signal amplitude, longitudinal signal phase, and transverse signal phase. However, the three signals; grain angle, grain alignment, and longitudinal signal amplitude were used in the strength estimation. Therefore, there will also be 24 features from the measured grain angle profiles. Each of these profiles will be transformed into: minimum, maximum, average, standard deviation; variance, kurtosis, skewness; and the standard deviation of the absolute value.

The X-ray images of the board will be transformed into the board three-dimensional model. The clear wood features: average, standard deviation, and variance of clear wood

density (as a random variable); the board volume; and the board density (sum of clear wood and non-clear wood density). Also, the knots will be characterized by: the number of knots; the ratio of knot volume to board volume; average and standard deviation of neighboring knot distances; average and standard deviation of knot three-dimensional shape from the main axis of the board; average and standard deviation of the standard-deviation of all the knot pixels from the main axis of the board; and average and maximum of knot area ratio.

The learning model is a regression-like estimator that relates measured features to the strength of the board. The selected features are given in Section 5.11. At this level, the concept of learning capacity (or VC dimension) of the learning system is important (see Section 6.3). Briefly, the limited number of measured specimens limits the accuracy of the learning system. In other words, there should be a search among different methods of learning to find the best model for a limited-size training set. Statistical learning theory was used to guide this search.

A group of learning systems operating on different concepts was tested to find which method works best. A learning method was suggested whose structure is compatible with the constraints of this learning problem. Finally, the performance of different learning models was compared.

## **1.6. Review of other attempts to transform measured mechanical properties into strength**

Since wood is a traditional structural material, there are many heuristic and scientific methods developed for estimating its strength. Among these, the weight of a piece of wood, which was used to represent the density of wood, and the detection of the presence of knots are the oldest. Scientific methods were later developed for finding the relationship between wood strength and one or more measured properties of wood. Since an estimation method can almost be identically used for tensile or the bending strength of wood, we will not distinguish between those. In the following discussion, different approaches to wood strength estimation are summarized.

The modulus of elasticity of wood is widely considered as the best single feature for estimating the strength of wood ([5],[6],[7]). In this method, the MOE (modulus of elasticity) profile of a board is measured and a regression estimator maps the average of MOE profile to the strength of wood. This method is the basis of an industry standard for machine stress grading. Special measurement machines were developed to measure this profile while the board passes through the machine at high speed. The coefficient of determination ( $r^2(x,y) = \frac{\text{COV}(x,y)}{\sigma_x \sigma_y}$ ) between strength and the measured feature is  $r^2 \approx 0.5$ .

The existing technology allows only the flatwise MOE profile of a board to be measured, that is, the MOE profile is measured by placing the bending force on the wider face (plank) of the board. Therefore it leaves out the applications where the bending force is applied to the narrower face (edge) of the board.

The grain angle of wood has also been used for estimating strength ([8],[9]). In a similar approach, a regression estimator transforms the average grain angle to the strength of wood. For this purpose, the capacitance based grain angle measurement and microwave based grain angle measurement tools were developed. The capacitance based grain angle measurement machine is most affected by the grain angle at the surface of the board. The microwave measurement method can produce more than one measurement. It can estimate the average grain angle, and also the moisture content and the local density of wood. The features are usually the statistical moments of the measured profile and a linear regression estimator can be used to map the features to the strength.

A finite element model of wood has been developed for estimating the strength of wood by assuming that the grain angle of wood is known at every point ([10]). Other parameters of wood such as local density, MOE, etc., are typically assumed to be constant. This model is reported to produce accurate results for single knot and double knot pieces of wood. The highest reported accuracy is  $r^2 = 0.89$  ([10]). Because of the high computation time, this method is useful for laboratory study but is not suitable for industrial speeds.

The effect of knots on the strength of wood was studied in ([11],[12],[13]). The use of knots in strength estimation is usually the limiting factor, and is applied in the visual grading of wood. There are category and boundary rules for estimating the impact of a

knot on the strength. Therefore, a board is downgraded if, based on the rules, the knot or knots are considered to be significant strength reducing factors. Although this is obviously a very qualitative assessment of their effect on strength, this method is an industry standard. This fact shows the high impact of knots on the quality of the graded wood. The geometrical features, as will be discussed later, are especially developed to include this factor for characterizing a board.

The local density of wood has also been suggested for strength estimation ([14],[15]). For this purpose X-ray machines (called X-ray lumber graders) were developed. It was proposed (arguably) as a non-contact measurement means that can replace the MOE measurement method ([15],[16]). In this case, the statistical features of an X-ray image of a board are transformed to its strength by using a linear regression model.

Visual systems (CCD cameras, infra-red spectroscopy, and laser scanning) were also used for defect detection and grading of wood ([17],[18]). These methods are based on finding the difference between the color of defects and that of clear wood. Also, three-dimensional profiling (using laser scanning) is studied for estimating the defects and shape of the board. The detected defects are related to the surface of the board and a regression estimator transforms their features to strength reducing factors of the board.

The grade of wood was estimated in ([19]) by using the slope of the grain, infrared spectroscopy, MOE, and visual scanning. A classification and regression tree ([20]) method was used to learn the best grading thresholds. The goal was to find better rules for strength classes. In the Classification And Regression Tree (CART) method the feature space is repeatedly divided into smaller sections, as long as the total estimation error is reduced.

In learning systems theory, the SPORE algorithm ([21]) was developed specifically for very high dimensional problems. This method of learning was originally developed for human-to-robot skill transfer. This method is suitable for very large dimension feature space (about 100 dimensional) and large numbers of specimens (about 10,000) and uses almost no *a priori* information about the data. As will be shown in Chapter 6, this method is one of the most successful methods for predicting the strength of wood. In this method, one generates a polynomial in the feature space ([21]) by producing two variable estimator blocks and adding the blocks to the model such that the estimation error is

minimized. The restriction of the wood strength estimation problem is that the practical number of measured specimens is limited to around 1000. This is because the measurement process is costly and time consuming. The SPORE algorithm has been tested for typical sample sizes of 10000. Therefore, this model may not be stable for a typical sample size used for wood strength estimation problem.

### **1.7. Economic importance of this knowledge**

The economic importance of this study arises from the limitations of the wood-grading standard. With the MSR lumber grading standard, three criteria must be met for every bundle that is sold. A flat MOE value ( $E$ ) and an edge MOR (the bending stress that breaks the board) value ( $F$ ) are provided for a grade. The following three criteria should be met for the assigned grade ([22]):

- a) The average flat MOE of a grade should be more than  $E$ ;
- b) The fifth percentile of MOE of the grade should be more than  $0.82 \times E$ ;
- c) The fifth percentile of MOR of the grade should be more than  $2.1 \times F$ .

In practice, a stress-grading machine measures the MOE profile of each board. The average and minimum of this profile is measured and transformed into its grade.

Recently, there have been attempts to improve the market value of the grading system. In a recent study ([23]) it was shown that in MEL grading standards the limiting factor is the ultimate tensile strength of the board. Therefore, an increased estimation accuracy of the tensile strength increases the grade yield by more than 5%. It was shown (by simulation) that an increase in MOE prediction from 0.7 to 0.9 will produce more than 17.5 million US\$/year in British Columbia ([23]).

### **1.8. Why Geometrical features are important**

As will be shown in future chapters, the problem of learning the tensile strength of lumber is an empirical learning problem. In problems with small sample numbers, the *a priori* knowledge about the input-output dependence directly improves the estimation.



The geometrical feature set is a first step in including the classical study of wood failure under stress into this learning problem. Naturally, the features that are introduced are crude and only general information is used as *a priori* knowledge. It will be shown in Chapter 6 that even this set of features produces relatively good estimation accuracy.

The potential advantage of a geometrical feature set approach is that it is based on the three-dimensional model of a board, which will be produced by using scanning profiles. Since the model of knots, the most common defect, is produced here, more refinement of features is possible by more carefully studying the effect of knots on strength. For example, through simulation enough samples can be generated so that a learning machine can find the effect of common types of defects on strength.

## **1.9. Outline of this thesis**

This thesis is presented in seven chapters. Chapter 2 describes the measurement systems used in this project. A look into the wood property that is to be measured is presented. Also, measurement restrictions that are due to the limitations of the measurement machines are presented. Chapter 3 presents the effect of common strength factors and how they are modeled. Chapter 4 presents the chosen approach for feature definition and board characterization.

Extraction of features from measured profiles is presented in Chapter 5. Details of knot detection algorithm that is the basis of the geometrical feature set is presented in this chapter. The problem of learning systems is discussed in Chapter 6. The learning problem and different learning machine structures are presented. Also, the ASEC learning machine is described in this chapter.

The Appendix presents an examination of wood failure. This study motivated the focus of this thesis and provides a basis for comparison between classical study on this subject and the approach of this thesis.

All the experimental results are given in the related chapters.

## 1.10. The Contributions of this thesis

The main contributions of this thesis are as follows:

1. For analyzing an X-ray image, image segmentation (using Bayesian classification) is developed. In this method, the statistical properties of the image is used for segmentation.
2. A geometrical model of knots is developed and the detection algorithm and transformations are presented.
3. For characterization of a board, a geometrical feature set is developed.
4. In order to estimate the accuracy of a learning system, a method for measuring the learning capacity (VC-dimension) of a regression estimator is presented.
5. The ASEC learning method (see Section 6.6) is presented. In this method, the *a priori* knowledge about the input-output relationship is used to transform the given feature set into a better set of features. This method combines feature transformation and feature selection.

## ***Chapter 2. Measurement means; practical constraints***

In this chapter the measurement systems used in the experiments are introduced and their limitations are discussed. For every wood property there are different technologies of measurement. Each technology and method of measurement is limited by its inherent restrictions that arise from the system's setup or its sensor requirements and limitations.

As shown in Section 1.2, the measurement process is the first step in characterizing a specimen. The limitations of the measurement process directly affect the accuracy of the strength estimation because it is the basis of specimen characterization and there is no other method for later compensation of the generated error. Measurement restrictions can be examined from three different perspectives.

The first limitation is that the measured property is affected by other properties of wood, that is, one property of wood cannot be singled out. For example, the measured density of wood, if measured by X-ray methods, increases with the increase of moisture content.

The second limitation in the measurement process is influenced by the measurement method and the spatial resolution of the measured profile. Obviously, a property of wood cannot be measured at a single point and is usually the average of that property over an area or volume of wood, either at the surface of the board or through the depth of the board. Any improvement of the measurement technique can result in a more sensitive machine and, therefore, improve the spatial resolution of the measured profile.

The third limitation in the measurement process is dependent on the quality of the measurement machine, the sampling rate, and the noise level of the output profile. In order to remove the measurement noise, almost all systems use averaging in a short period. For a board-moving system with a set of fixed sensors, temporal averaging leads to spatial averaging. Higher-speed sensors allow the system to generate the output faster. In the following section the physical basis of the measurement machines is discussed.

## **2.1. Measurement of the Physical Properties of Wood**

The physical properties of wood relevant to the measurement systems of this project are discussed in the following.

### **2.1.1. Modulus of elasticity of wood**

The modulus of elasticity of a specimen is measured by using bending machines. The board is bent and the force needed to bend the board is measured.

Young's modulus or modulus of elasticity (MOE), is widely recognized by the wood products industry as the best variable estimator of the bending strength of boards ([5], [6], [7]). It is defined as the ratio of stress over strain.

In the three point bending test, as shown in Figure 2.1a, force is applied at the midpoint of the board. MOE is calculated as follows.

$$MOE = P L^3 / 48ID \quad (1)$$

Where  $P$  is the load (N),  $D$  is the deflection (m),  $L$  is the span of the measurement (m), and  $I$  is the moment of inertia ( $m^3$ ).

This test is sensitive to the location of defects in the wood specimen. Therefore, the operator tries to find the weakest point on the board and places it right at the midpoint (and on the tensile stress side), where the force is being applied. To reduce the sensitivity of this, a four-point test is often used (see Figure 2.1b). In this test, the middle part of the board is in constant bending moment, therefore, the whole span between the two forcing points is uniformly tested.

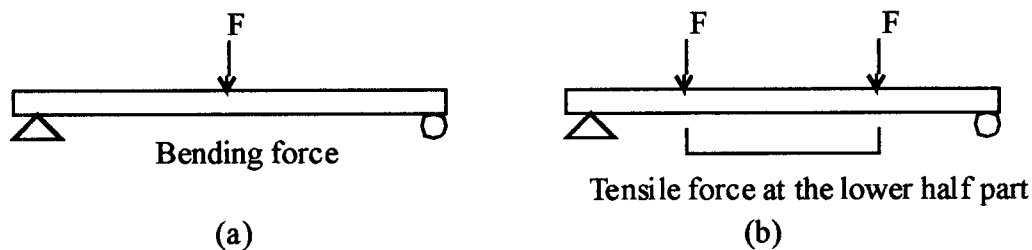


Figure 2.1 Three-point and four-point bending test.

What is shown in Figure 2.1 is the basis of static MOE measurement systems where the board is placed on holders and bending force is applied. In a mill the boards are passed sequentially through a machine that continuously measures the MOE. The CLT™ (Continuous Lumber Tester) is one of these machines. As shown in Figure 2.2 the dynamic bending machine has four rollers that are slightly displaced from a straight line. The first and last rollers are placed to grip the board and push the board through the system. The middle rollers are slightly displaced (Figure 2.2) so that they bend the board as it passes through. The rollers' locations are fixed, therefore, the bending of the passing board is fixed. The force applied by the board to the bending rollers is measured and sent to the output of the system.

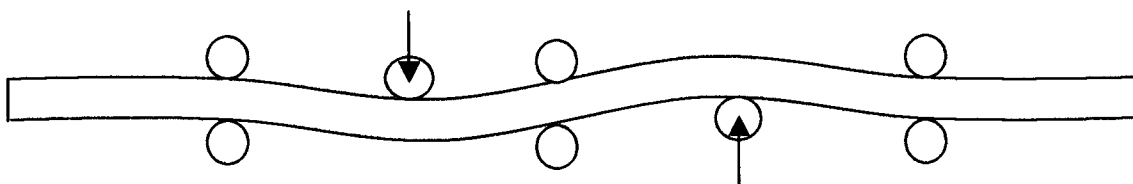


Figure 2.2 The mechanism of a Continuous Lumber Tester (CLT™, or dynamic bending machine) machine.

### 2.1.2. Moisture content

The conductivity of wood is proportional to its moisture content. Therefore, by placing two conducting nails inside the board and measuring the resistance between them the moisture content of the board can be measured. Although this factor was not used as a feature of specimens, it was regularly checked to make sure that the boards were dry (moisture content less than 12%).

### 2.1.3. Dielectric Properties

Measuring the grain angle of wood is usually based on its directional electric property. The permittivity of wood depends on the angle between an applied electric field and the longitudinal direction of the tracheid of wood. Also, the permittivity depends on the density, moisture content, and temperature of wood as well as the frequency of the electric field used to measure it. Although these factors are briefly discussed here, they are not of significant importance because the grain angle measurement is based on the difference of permittivity along the grain and across the grain.

The directional and variable permittivity of wood is based on the polarization of its material and internal water and vapor. There are five polarization types that contribute to the overall polarization of wood. These are electronic polarization, ionic polarization, dipole polarization, interfacial polarization, and electrolytic polarization ([31]). Electronic polarization is due to the polarization of the atoms of the dielectric. Ionic polarization is a result of elastic displacement of atoms in a molecule. Dipole polarization is due to alignment of the existing dipole molecules in the direction of the electric field. Interfacial polarization is due to the interaction of water and the cell wall that makes a cavity similar to a dipole. Electrolytic polarization is caused by the electrolysis of the soluble components due to the electric field. The first two polarizations are the major ones while the last three create the loss factor, making the permittivity a complex number as follows:

$$\varepsilon = \varepsilon' - i\varepsilon'' \quad (2)$$

The permittivity is a directional factor with three major directions, longitudinal, radial, and tangential. The longitudinal permittivity is the biggest factor while the radial and tangential permittivity factors are approximately the same. All three factors are increased by wood density and moisture content. The permittivity of wood decreases as the frequency of the applied electric field increases. An increase in temperature may also cause an increase or decrease in permittivity, depending on the frequency of the applied electric field ([31]). The loss factor ( $\varepsilon''$ ) is usually much smaller than the real part of the permittivity itself, and therefore can be overlooked in the qualitative analysis of the measurement systems.

#### 2.1.4. Density of wood

It was shown that an X-ray source could be used for wood densitometry ([24]). The X-ray can be generated by an X-ray tube that projects beta rays onto a metal plate. The X-ray absorption phenomenon is a combination of three absorption factors: the photoelectric absorption, the scattering absorption, and pair production absorption. The electrons of the matter rarely absorb the X-ray photons. This phenomenon is called true absorption or photoelectric absorption. Also, some of the photons are deflected by the atoms of the mass, which results in the scattering factor. In pair production absorption the ray transforms to matter (electron and positron) by the Einstein equation.

The attenuation of the ray as it passes through the board is as follows.

$$I = I_0 \exp(-\mu t) \quad (3)$$

where,  $\mu$ , is the mass absorption coefficient (1/cm), and  $t$  is the specimen thickness. This equation can be written as follows.

$$I = I_0 \exp(-(\mu / \rho) \rho t) \quad (4)$$

where  $\rho$  is density (g/cm<sup>3</sup>) and  $\mu/\rho$  is the mass-absorption coefficient in square centimeters (area) per gram ([25]). A method of calculating the mass absorption coefficient for different materials is published in [26] and tables of it are published ([25]).

As the attenuated ray comes out of the wood, it is transformed to a measurable signal by an array of sensors. Each sensor is based on ionization of its atoms (gas or crystal) in an applied electric field ([27]), generating photons in the process. Every photon absorbed by the detector generates a pulse of electric current. The count of the pulses in a time interval is proportional to the intensity of the X-ray.

#### 2.2. Measuring density with an X-ray machine

X-ray scanning is used to measure the local density of wood. As the X-ray passes through the board its intensity is attenuated. A sensor array measures the ray intensity.

The board density image is mainly used for knot detection, which leads to the board's structural model. Nevertheless, the system and method of this densitometry are commonly used and are explained in the following.

The system consists of an X-ray source tube, which projects an X-ray fan on the object, and a linear array of sensors that detects the intensity of the rays after passing through the board. A metal frame reduces the vibration caused by any moving components. Lead curtains create a shield for the operator from the X-rays. A few rollers lead the board (pushed by the conveyor) through the system.

As previously discussed, the attenuated ray coming out of the wood is transformed into a measurable signal by a linear array of sensors ([27]).

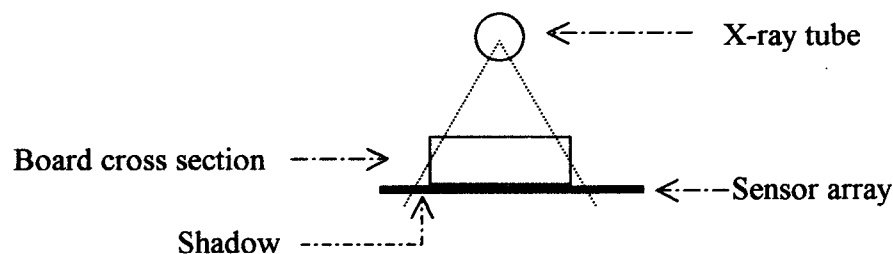


Figure 2.3 A simple diagram of the X-ray system and the shadow image results from a non-collimated X-ray. The board is extended and moves perpendicular to this page.

The measured variable is the ray density, which is calibrated to wood density. For that purpose a set of calibration aluminum plates with knot density and different thickness were used. They were placed in the machine and a polynomial was fitted to the output signal so that it transformed the measured intensity to wood density.

An array of sensors (in a single line) can capture the density profile across the board at the measurement point. Each measurement generates a density profile related to one cross-section of the board. A conveyor moves the board along, while repeated scanning takes place. In the end, an image of wood density results for the whole board. The X-ray densitometry method and the structure of the measurement system impose some constraints on the produced image that are discussed here. The X-ray generated from the



tube is not collimated; therefore, it creates a shadow of the board corner at the sides of the density profile.

If the angle of the ray ( $\theta$ ) is relatively small (for the vertical distance between the source and the sensor array) there will be little horizontal deviation ( $s$ ) in the projection of the measured signal and thus no collimation is needed, which was the case for the X-ray system used in this thesis. As will be shown, the angle of the ray can be measured from the X-ray image. The shadow of the board ( $s$ ) (as is shown in Figure 2.4) is less than three pixels in the machine used, which is considered negligible in the measured profiles and is ignored in the feature extraction procedure.

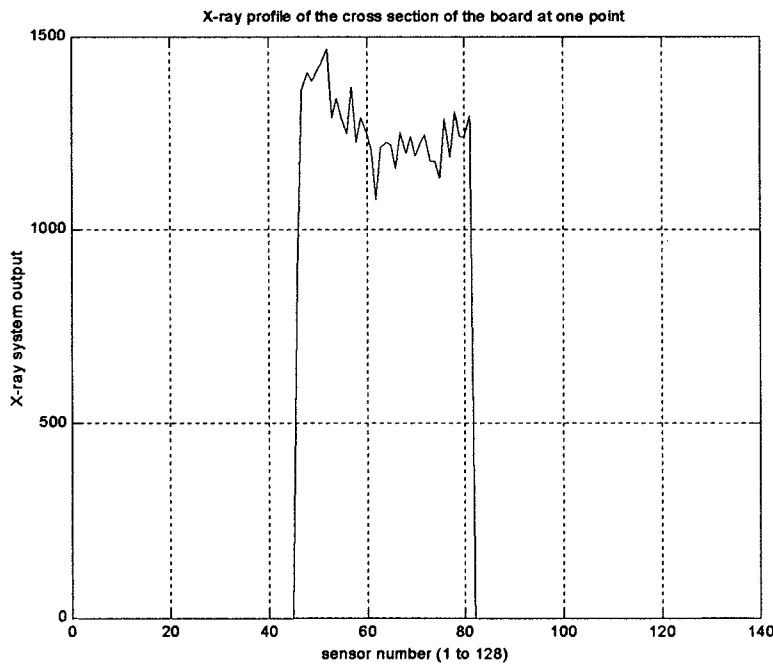


Figure 2.4 The profile of one cross-section of the board. The shadow of the board creates the side ramps.

In Figure 2.4 the shadow of the board is 2 pixels at each side and the total number of pixels for the board is 35. Using the standard board size (89mm by 35 mm), one can calculate the maximum ray angle ( $\theta$ ) with the center-line as follows (Figure 2.5):

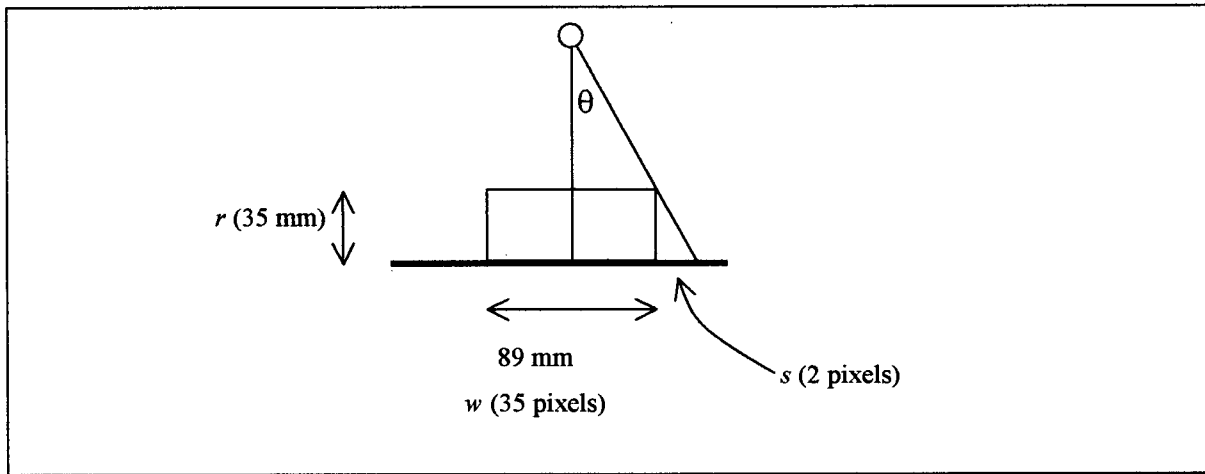


Figure 2.5 The shadow of an X-ray image.

$$\theta = \text{atan} \left( \frac{s}{r} \right) = \text{atan} \left( \frac{s}{w} * \frac{89 \text{ mm}}{r} \right) = \text{atan} \left( \frac{2}{35} * \frac{89 \text{ mm}}{35 \text{ mm}} \right) = 8.1^\circ \quad (5)$$

There is a more important shadow image that is related to the image of knots of a board. Figure 2.6 shows an imaginary knot projected on the image plane similar to the result of X-ray scanning of the board. A similar pattern is generated from the X-ray image of a knot and will be used to detect knots in a board.

The density of the image at every point is obviously the summation of the wood density throughout the thickness of the board at that point. If two perpendicular images are produced, a sense of density distribution in the body of the board can be obtained.

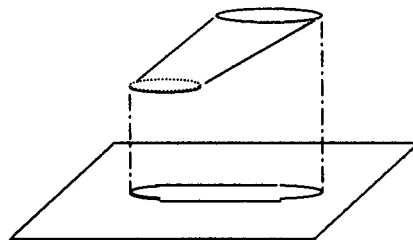


Figure 2.6 The shadow image of a knot.

The sensor array makes a fixed resolution profile across the board. The image resolution along the board depends on the speed of the board. If the conveyor can push the board

through the system with a constant speed this resolution will be constant too. Through the course of this thesis, two X-ray systems were used for X-ray scanning. One of them (made by Vision Smart) did not have any conveyor attached to it. Therefore a conveyor was placed next to the machine that pushed most of the board through the X-ray scanner and the end of the board (about 90 cm of the length of the board) would free-run through the system because of its momentum.

The deceleration of the board was compensated for as a preprocessing step when the image was being processed. The image was segmented into two parts, the first part related to the part of the board that was pushed through the X-ray system with constant speed. The second part of the image was related to the part of the board that was decelerating. A constant speed (that is the average speed during deceleration) was assigned to the second part of the image, as the momentary speed could not be estimated. The strength estimations by using each of the outputs of the two machines resulted in similar strength estimation accuracy therefore here only a typical system is described.

Another consideration regarding the X-ray image is that there should be a zero density region that surrounding the board density image. The content of the surrounding area is noise whose level is much less than the density of clear wood. A threshold level can separate these parts and extract the density image of the board.

Figure 2.7 shows the contour map of the top (for the plank of the board) and side images of the board. The lower right part of the image shows increased density that can be due to natural density variation of clear wood or related to a defect, such as a resin canal.

Vibration of the board is not considered in this case because slight vertical displacement of the board does not affect the measured image. This is because the source is far from the board surface and the slight change in the vertical location of the board does not create significant change in the sensor readings. The noise level of the system in the surrounding area is negligible, as can be seen from Figure 2.4. However, filtering by a square filter of size three was advised (by the manufacturer) for smoothing the wood density. The system stores the measured ray intensity in 16 bits of data, which covers the spectrum of black to white.

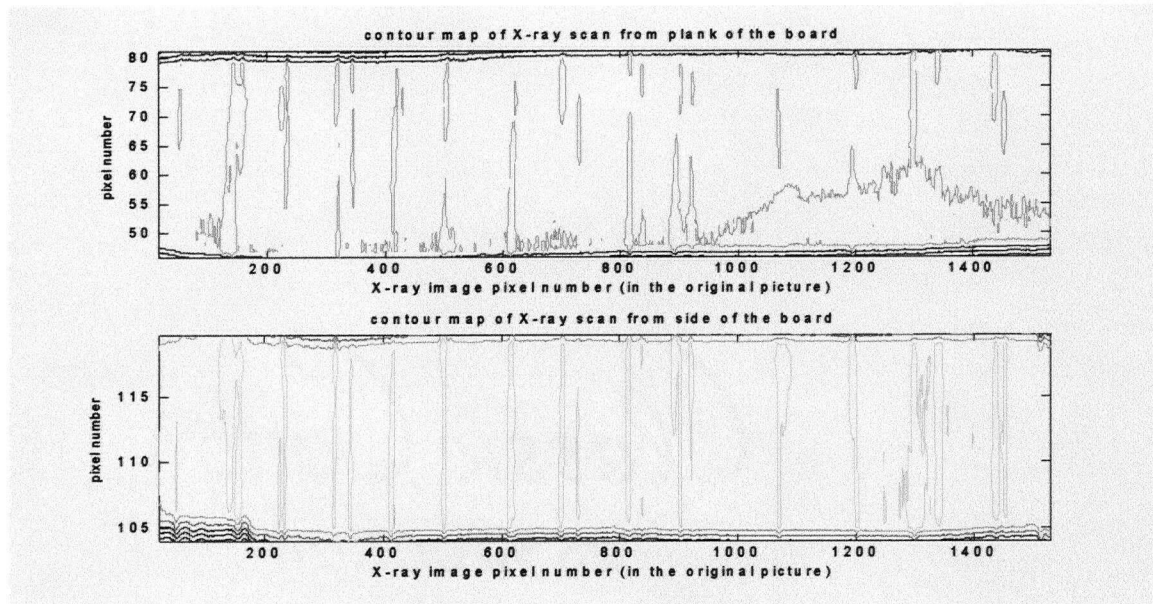


Figure 2.7 The contour map of an X-ray image.

The final digital image is displayed by using two diagrams, the contour map and the three dimensional diagram of the density. The contour map (Figure 2.7) is used for showing the general form of the image and the relative location of the defects in a board. The three dimensional density-diagram (similar to Figure 4.6) is used for showing the density variation due to the existence of a defect.

## 2.3. Measuring Grain Angle with SOG and Microwave Machines

The principle of grain angle measurement was discussed above. In this section the details of the two grain-angle measurement machines are presented.

### 2.3.1. Grain angle via an SOG machine

The capacitance based measurement system that was used in this research was based on measuring the impedance of a capacitor whose dielectric was the board at the measurement point and its operation is as follows (based on directional permittivity of wood). If two parallel strips of metal are placed on the top surface of a board and the

impedance of the created capacitor is measured, the impedance depends on the angle of the electric field of the capacitor. If the capacitor (i.e. the two metal plates) is rotated around its center the measured capacitance varies. The maximum capacitance is measured when the electric field is along the direction of the grain.

In a rotating head capacitor a cylindrical head is rotated around its axis. The bottom of the cylinder, where it contacts the board, four pieces of metal and a cross-shaped dielectric compose the circle as is shown in Figure 2.8. This arrangement creates two pairs of parallel capacitors that are perpendicular to each other (as shown by the dotted lines in Figure 2.8). Therefore if the electric field of two of the parallel capacitors is along the grain of wood then the electric field of the other two capacitors is perpendicular to the grain of wood. The angle of the head where one of the two impedances is maximized and the other one is minimized is the angle of the grain of wood.

The impedance of each of the capacitors is most affected by the grain of wood at the surface of the board. The wood that is beneath the surface is affected less because the magnitude of electric field is less inside the board. Also, the measurement is taken over the area of the head. The diameter of the measurement head was about 5cm. The underlying assumption for accurate measurement is that the grain angle, and other factors that affect the permittivity, are about the same over the area of the measurement head.

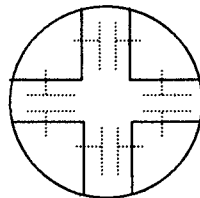


Figure 2.8 The rotating head for grain angle measurement.

The measurement head rotates at 60 cycles per second and every half cycle one measurement can be completed. Therefore the sampling rate is limited to 120 samples per second. The grain-angle sampling rate is then controlled by the speed of the conveyor. The sampling rate of SOG measurement was about 1500 samples per board, which is about one sample per 3.3mm. The system's resolution is 0.1 degree and its accuracy is ( $\pm 0.9$ ) degree and the system covers  $-89$  to  $89$  degrees ([2],[3]).

Figure 2.9 shows a profile that is measured by using this system. The large variations in Figure 2.9 are related to grain angle defects, which are usually related to knots. Small variations are related to the measurement noise. The two constant measurements at the beginning and the end of the profile is produced by the system when the board is not attached to the measurement head. The large variations, at the beginning and the end of the board's grain angle profile, are produced when the board first reaches the measurement head or leaves it.

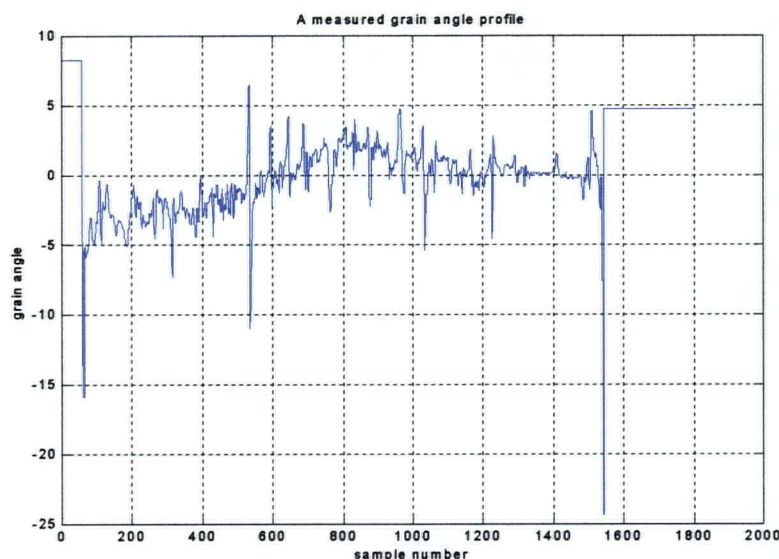


Figure 2.9 A measured grain angle profile of a board.

The grain angle measuring system consisted of a metal frame and plastic rollers to push the board along the system and before the measurement heads. A stepper motor controls the movement and speed of the board. Two measurement heads were placed on the system so that two profiles could be measured simultaneously, one from the face of the board and one from the edge of the board. The diameters of the two heads were different. One head diameter was almost equal the thickness across the face of the board and the other head diameter was almost equal the thickness across the edge of the board. Four profiles were measured by using this arrangement, two profiles from the two faces and two profiles from the two edges. By rearranging the system four extra profiles were

measured from the two faces of the board, each grain angle profile related to half of a face.

### 2.3.2. Grain angle measurement via microwaves

The microwave grain angle measurement machine is based on the directional attenuation of an electric field in wood. A simple diagram of such a machine is shown in Figure 2.10 ([28],[29]). In this machine a linearly polarized electromagnetic wave at microwave frequency ( $\approx 10$  GHz) is transmitted through the thickness of the specimen. The directional conductivity and dielectric property of wood creates a non-symmetric attenuation and delay of the wave. Therefore, the output wave is an elliptically polarized wave whose major axis is along the grain. The Microwave based grain angle measurement is based on identifying this ellipse. The standard deviation of the measured grain angle is 2 degrees.

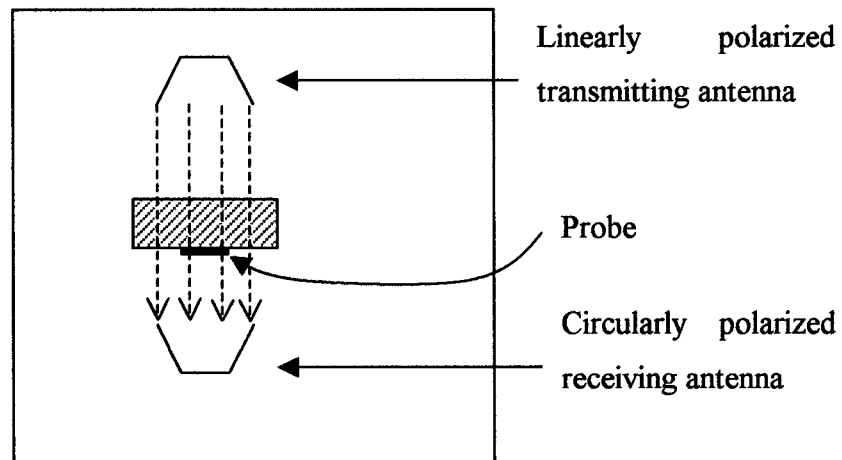


Figure 2.10 Microwave System for measuring wood grain angle

Let's say the electric field is represented by a sine wave as follows:

$$E = E_0 \sin(\omega t) \quad (6)$$

where the amplitude ( $E_0$ ) and the angular frequency ( $\omega$ ) are constant. Wood as an orthotropic material shows different loss factor and phase factor along and across its

grain. This would transform the linearly polarized electromagnetic wave to an elliptically polarized wave as shown in the following ([28],[29]).

The incident wave components on longitudinal and transversal axes are as follows:

$$E_{IP} = E_0 e^{j\alpha} \cos(\theta) \quad (7)$$

$$E_{IT} = E_0 e^{j\alpha} \sin(\theta) \quad (8)$$

where  $\theta$  is the angle of the incident field with the longitudinal axis. As the transmitted wave propagates through wood, it is attenuated and delayed by the longitudinal and transverse propagation constants.

$$\gamma_p = \alpha_p + j\beta_p \quad (9)$$

$$\gamma_t = \alpha_t + j\beta_t \quad (10)$$

Moisture content is the major factor of the wave attenuation and the wood material is the major factor of the wave delay. If the thickness of the specimen is  $d$ , the output wave is as follows:

$$E_{TP} = E_{IP} e^{-\gamma_p d} = E_0 \cos(\theta) e^{-\alpha_p d + j(\omega t - \beta_p d)} \quad (11)$$

$$E_{TT} = E_{IT} e^{-\gamma_t d} = E_0 \sin(\theta) e^{-\alpha_t d + j(\omega t - \beta_t d)} \quad (12)$$

which is an elliptically polarized wave. The assumption for the validity of equations (11) and (12) is that the two components of the electric field are uncoupled. Furthermore, based on a commonly used approximation ([30]), all internal multiple-reflections, inside the board, are neglected. The ratio of  $E_{TT}$  to  $E_{TP}$  was used as a feature of the grain angle that defines the alignment of the grain angle through the thickness of the board. In order to identify the ellipse, one needs to measure the amplitude of the electric field at least in three different angles. Therefore four probes, with a 45 degree angle between each other, were located underneath the specimen ([30]). Each probe is used to measure the amplitude of the electric field along the direction of the probe. A probe adds a 455 KHz wave to the microwave. At the other end of the machine, the modulated signal is received



and the power (rms value) of the received signal is measured. The measured power represents the power of the transmitted signal along the direction of the probe.

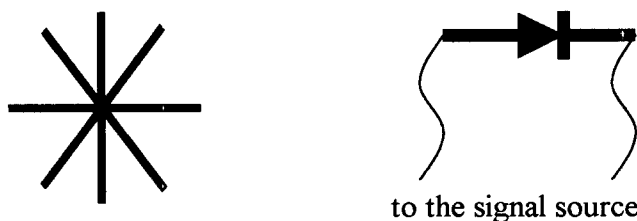


Figure 2.11 The arrangement of the probes (left) and the structure of a probe (right).

Once the amplitude (or rms value) of the signal along the direction of the probes is known, the general ellipse equation can be solved and the direction of the grain can be obtained ([29]). The measured grain angle profile is similar to that of capacitance based (SOG) machine. The difference is that the output of the SOG machine is more affected by the surface grain of the specimen, while the output of the microwave machine represents the average grain angle through the depth of the specimen.

## 2.4. Modulus of Elasticity using the Continuous Lumber Tester

As was discussed on Section 2.1, the Continuous Lumber Tester machine measures the flatwise MOE profile of the specimen by moving the board through two slightly displaced rollers and measuring the bending force. Two profiles are measured from each board (Figure 2.12) one from each side of the board. The machine is calibrated by using aluminum bars. However the average MOE of the two profiles was compared with the static MOE measurement (Section 2.1) in the lab. A linear dependence between the measured average output and the static MOE was observed. As is shown in Figure 2.12 the noise level when there is no board in the machine is negligible as compared with the signal level. Therefore, thresholding was used to extract the MOE profile from the measured profile. The two profiles that are shown in Figure 2.12 are apart by a fixed distance. This distance is equal to the distance of the two measurement cells.

The measured signal at the output of each sensor is an average over the span of the board between one gripping roller and the other sensor roller. Therefore, the signal is very

smooth and does not need a high sampling rate. As is shown in Figure 2.12 there are about 500 samples over the length of the board. About 70 centimeters from each end of the board is not measured. The distance between the two sensors was twice that length, i.e., 140 centimeters ([4]).

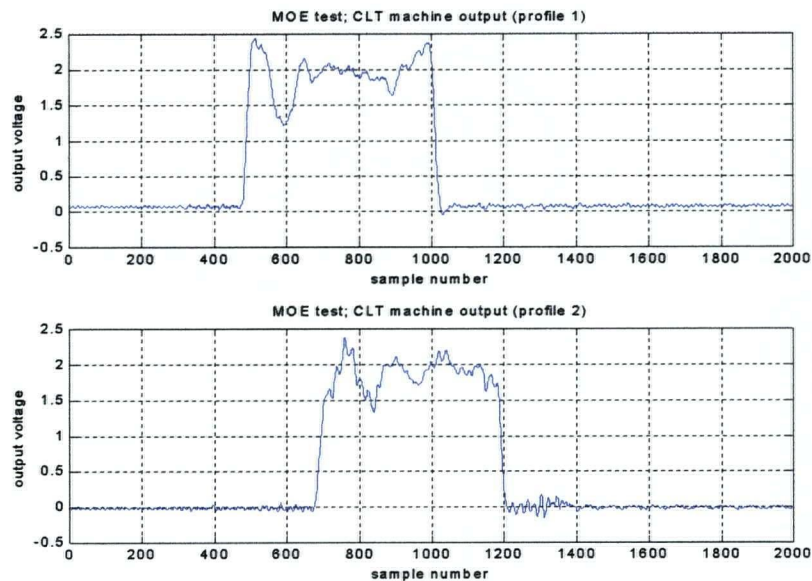


Figure 2.12 Typical output profiles of a dynamic bending machine.

#### 2.4.1. Calibration

Unlike other measurement machines the calibration information for the dynamic bending machine was available at the time of the experiments. The following test was carried through at UBC ([32]) in order to ensure that the dynamic bending machine measurement was calibrated and that it had the same reliability as the other measurement systems.

Fifteen specimens were randomly selected and were tested in a flatwise-static bending test as well as being tested by the dynamic bending machine machine. The average of the dynamic bending machine output and the average of the static MOE profile were used to calculate the calibration factors. As will be shown in the following, the result shows good linear correlation between the average dynamic bending machine and the average static MOE along the board. The repeatability of each measurement was tested as follows. For

each of five randomly selected specimens, the dynamic bending machine measurement was carried through five times. The repeated profiles were plotted in a diagram for comparison. Visual comparison shows good match of repeated profiles.

The dynamic bending machine output measurement was explained previously and a sample output was shown in Figure 2.12. Each profile was extracted from the corresponding measured (dynamic bending machine output) signal by removing the surrounding area by using a threshold level of 0.1 v. magnitude. The result is shown in Figure 2.13.

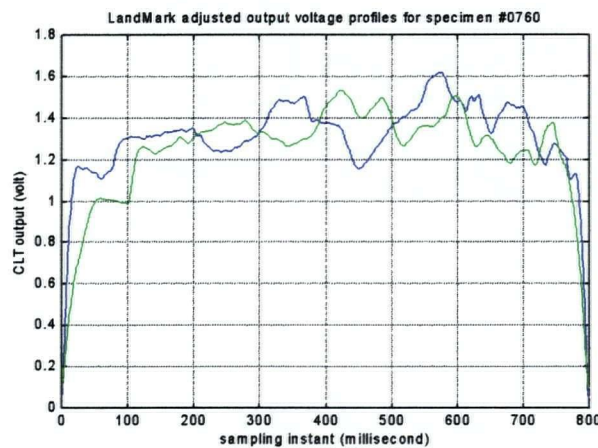


Figure 2.13 The extracted and adjusted dynamic bending machine output profile for specimen #0760

The flatwise-static MOE profiles were measured by means of two independent Static Bending machines, i.e., Metriguard Model 440 Static Bending Tester and Eldelco DART Static MOE Tester ([32]).

Static MOE measurements were taken at equal spaces fifteen centimeters along each board. Seventy centimeters from both ends of the board were not measured in order to have a proper support of the board. Two sample profiles are shown in Figure 2.14 and good agreement is obtained between the two static test machines. One of the machines measures the MOE of a specimen by three-point test and the other measured the MOE by a four point MOE test (see Section 2.1.1). The difference between the two profiles of Figure 2.14 is due to this difference in the machines.

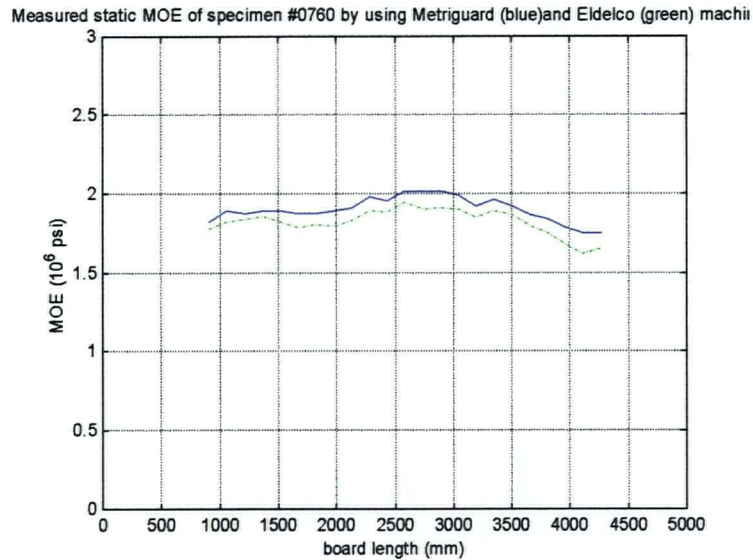


Figure 2.14 Measured static MOE of specimen #0760 by using Metriguard (blue) and Eldelco (green) machines, (a) original measurement, (b) adjusted for sampling position.

The extracted and adjusted profile shows the actual measured MOE at the represented points of the board. The average of the two profiles in Figure 2.13 (shown as V in Table 1) and the average of the two profiles of Figure 2.14 (shown as E in Table 1) were used for calibration.

The measured average dynamic bending machine and average static MOE are shown in the following Figure 2.15. The coefficient of determination of the two variables is  $r^2 = 0.92$ , which shows a linear relationship between the two variables.

**Table 1: The average CLT<sup>TM</sup> output, average MOE, and transformation coefficients for selected specimens.**

Specimen #	V (CLT <sup>TM</sup> )	E (MOE)	$C_1$	$C_1^1$
760	1.3683	1.8833	0.7265	1.3764
83	1.7110	1.8623	0.9188	1.0884
749	1.2610	1.5726	0.8019	1.2471
9	1.1853	1.3637	0.8692	1.1505
570	1.0565	1.2586	0.8394	1.1913
198	1.5896	1.8423	0.8628	1.1590
503	1.5518	1.7860	0.8689	1.1509
616	1.8549	2.0528	0.9036	1.1067
62	1.3717	1.4725	0.9315	1.0735
94	1.0792	1.4200	0.7600	1.3158
642	1.4377	1.7064	0.8425	1.1869
2	1.9447	2.1876	0.8890	1.1249
612	1.6248	1.7914	0.9070	1.1025
623	1.4489	1.7315	0.8368	1.1950
17	1.7727	2.1795	0.8134	1.2295



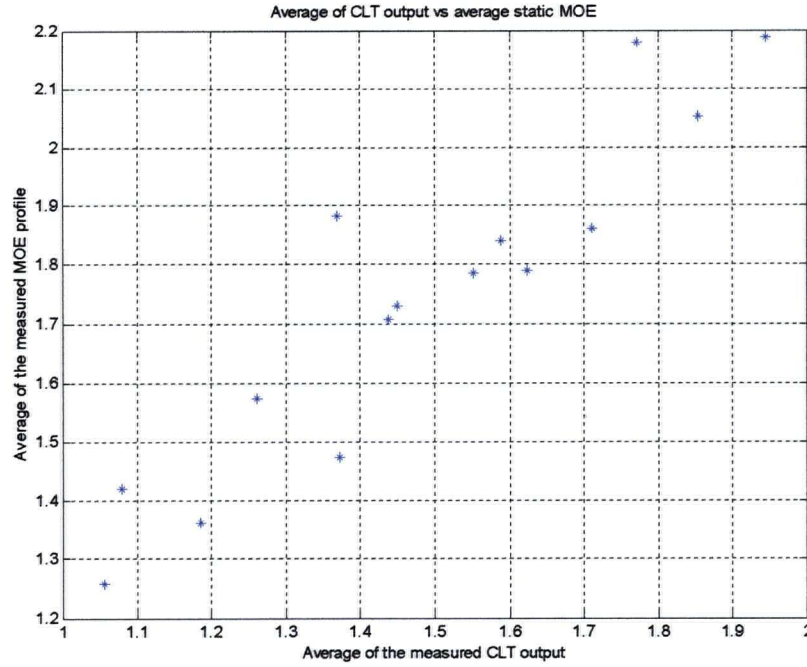


Figure 2.15 Linear relationship between the average measured dynamic bending machine output and the average static MOE.

Therefore for a linear relationship, the calibration equation of each cell (that produces each dynamic bending machine profile) is as follows:

$$V = C_1 E + C_0 \quad (13)$$

As was stated before,  $V$  is the average of the dynamic bending machine output and  $E$  is the average static MOE.  $C_0$  is the output offset of the dynamic bending machine machine and is equal to the output voltage when there is no board inside. Therefore, the part of a profile that is not related to the board (for example, see the side parts of the two profiles shown in Figure 2.12) is equal to this offset factor. Each measured profile is the output of one force-measuring cell, which measures the required force for bending the board along each of its faces. For the two cells there will be four factors ( $C_{01}$  and  $C_{02}$  for offset factors and  $C_{11}$  and  $C_{12}$  for the scale factors), which are shown in the following equations:

$$C_{01} = 0.0626 \quad (14)$$

$$C_{02} = 0.0172 \quad (15)$$

$C_{11}$  and  $C_{12}$  are calculated by comparing the measured profiles (from dynamic bending machine ) with the static MOE profiles of the selected samples and by using the following equations:

$$C_{11} = (V_1 - C_{01}) / E_1 \quad (16)$$

$$C_{12} = (V_2 - C_{02}) / E_2 \quad (17)$$

As shown in Figure 2.13 and Figure 2.16 the two profiles are almost the same therefore the scale factors are about the same ( $C_{11} \cong C_{12} \cong C_1$ ) and are calculated as follows:

$$C_1 = (E_1 + E_2) / ((V_1 - C_{01}) + (V_2 - C_{02})) \quad (18)$$

The 10% of the beginning and the end of the dynamic bending machine profile are removed in order to avoid the transient part of the measured profile. The following calibration factors were calculated by using linear regression. Equation (19) is the calibration equation for profile #1 (the output of cell #1 of the dynamic bending machine) and Equation (20) is the calibration equation for profile #2 (the output of cell #2) of the dynamic bending machine:

$$\text{Profile \#1: } E = 1.1808 (V - 0.0626) \quad (19)$$

$$\text{Profile \#2: } E = 1.1808 (V - 0.0172) \quad (20)$$

The final calibrated dynamic bending machine profiles of a few specimens are shown in Figure 2.16.

In order to check the consistency of the measured profiles, five boards were randomly selected and the dynamic bending machine profile of each were measured five times. Figure 2.17 and Figure 2.18 show that the superimposed profiles match well for both cells of the machine. The measured MOE values are very close but the location of the profiles can be compromised in some cases because of the board slippage under the conveyor rollers of the dynamic bending machine.

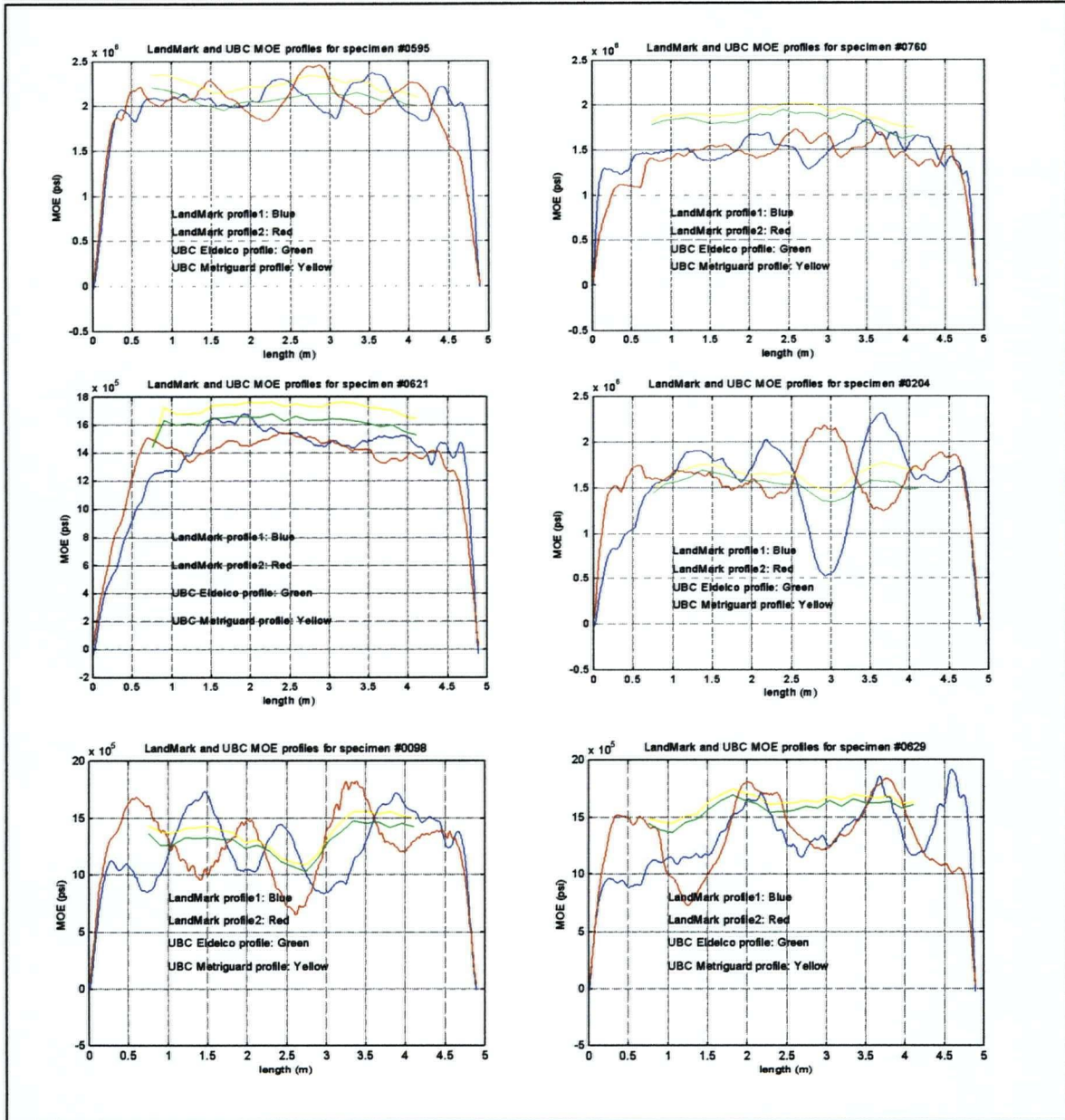


Figure 2.16 Measured MOE profiles by CLT™ and static bending methods.



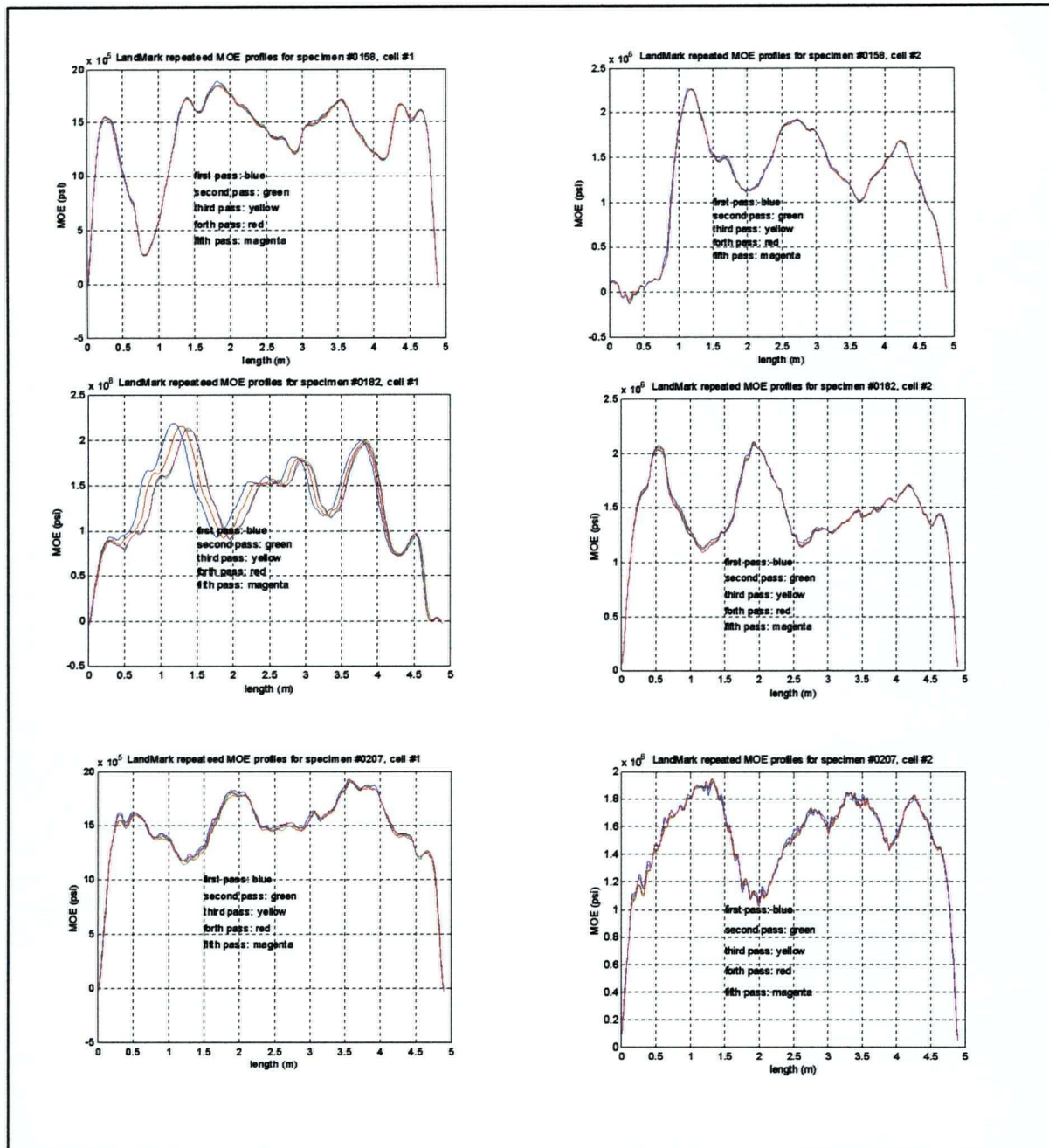


Figure 2.17 The repeatability of the measured profiles for CLT™ machine.

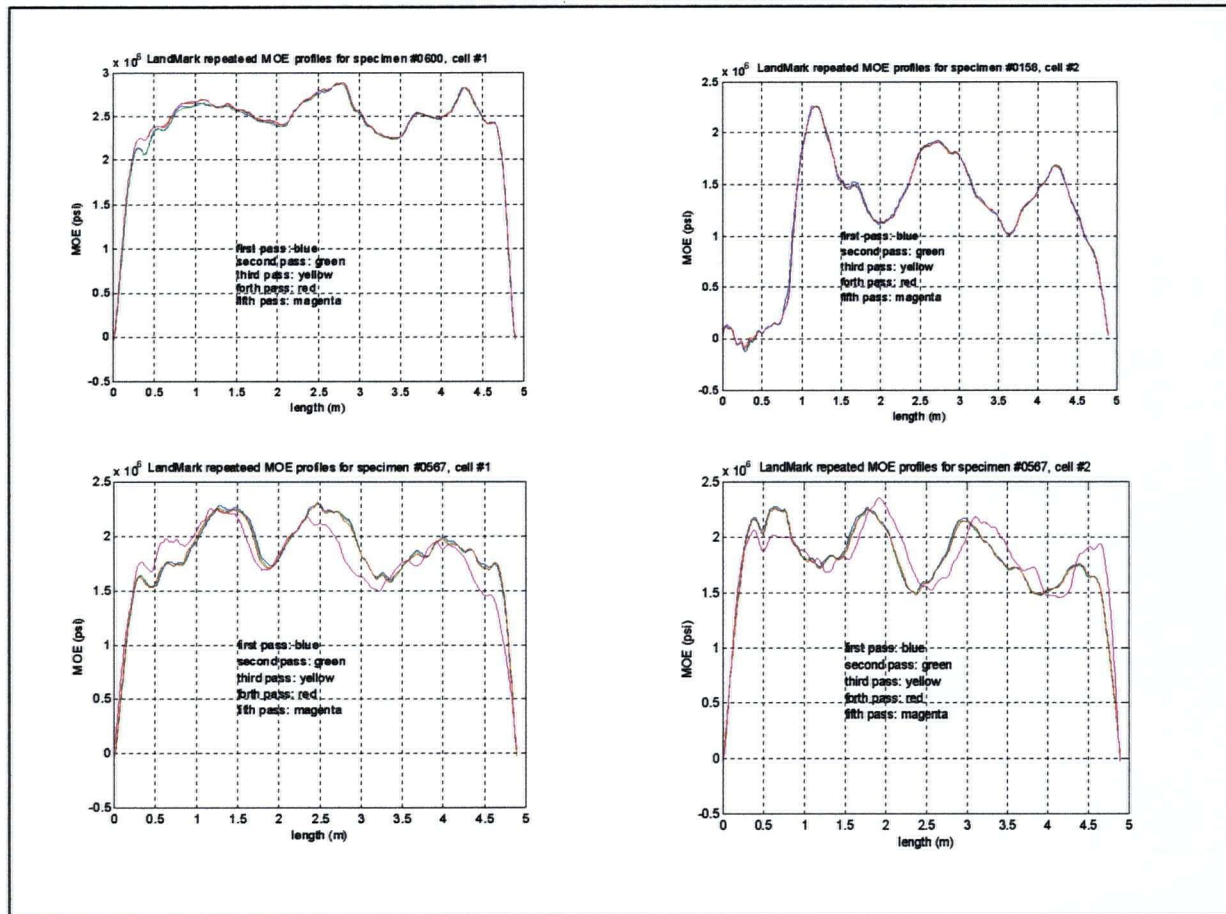


Figure 2.18 The repeatability of the measured profiles for CLT<sup>TM</sup> machine (continued).

## 2.5. Ultimate Tensile Strength Tester

This machine destructively tests the boards. The board is put under increasing tensile stress until it breaks apart. The tensile stress is recorded at every instant and is kept when the board is broken. The system consists of two grips; one moving and one fixed. A metal frame holds the grips and provides the fencing for security of the operator from wood particles that may fly around due to the release of tensile stress when the board breaks. The moving grip is driven by a pneumatic system whose measured pressure is transformed to the applied tensile stress.

### ***Chapter 3. Modeling wood failure***

If a board is tested in tension, a three-dimensional stress distribution is generated inside the board. This stress distribution is not usually uniform due to the variation of wood properties inside the board. A crack usually is initiated in one part (or more) where the stress level is high. If the applied tensile stress (to the board) is increased, the chance of wood failure under stress is increased. Once the wood fails at a point of the board, it generates a higher stress zone around it. Since the property of wood usually changes gradually, the excessive stress can also cause the failure at a neighboring point. This phenomenon can generate a chain of failures that finally ends when the board breaks apart ([33]).

In the following sections the factors influencing the strength of wood are presented. Also, the measurable features of wood that have important effects on strength are explained. A board is composed of clear wood and various defects, which interact when tensile stress is applied. For example, assuming that wood property is known in a neighborhood, by knowing the geometrical features of the defects in the board and wood properties, the stress distribution may be estimated. Each of the features (defects and clear wood) and their effect on total strength are discussed in the related sections.

The discussions of each section are mostly limited to qualitative descriptions of the related physical factors. A feature or a set of features will represent each physical factor. Therefore, the qualitative discussion provides an insight into the problem and helps in defining the features. An exact relationship between each physical phenomenon and the tensile strength is not needed (and in most cases it is not known). The collective relationship will be learned by the estimation (or learning) algorithm, which will be discussed in subsequent chapters.

Two very important aspects of wood are the complexity of its structure (as a fibrous anisotropic material) and variation in its local features (due to various defects). Therefore, even if the features of wood were accurately measured at every point of the board and if

the strength model was fully developed, the computation of a board's tensile strength doesn't seem feasible for on-line use. Therefore, simplification seems an integral part of this problem if a practical strength estimator is desired.

### 3.1. Tensile strength of clear wood as a random variable

In strength-of-material testing in the early twentieth century it was observed that the tensile strength of similar specimens are dispersed and cannot be modeled by a single value. The tensile strength of matter, therefore, was modeled by a probability density function. Weibull weakest link theory ([34],[35]) is a theoretical explanation of the probabilistic behavior of material strength. This theory was based on two major assumptions. The first assumption was that the cause of fracture is the flaw (microscopic cracks) in the material and that these flaws are independent and are randomly scattered in the volume of the material. The second assumption was that the entire material holds up against the applied stress until one of the flaws fails. This assumption is actually the weakest link theory, which is applicable to brittle material like glass or china. It was shown that this assumption is valid for wood if the mode of fracture is brittle.

In non-brittle material the strength of a specimen is not equal to the strength of its weakest element and the failure of the specimen is a random process rather than a sudden event. When the fracture starts, stresses are redistributed and it doesn't necessarily lead to immediate collapse or failure. Also, neighboring points can affect the stress tolerance of an element of the specimen.

For the brittle case, the probability of fracture at a point,  $S$ , is related to the stress level,  $\sigma$ , over the volume of the material,  $V$ , by the following equation (Weibull distribution, [34],[35]):

$$\log(1 - S) = - \int_V n(\sigma) dv \quad (21)$$

where  $n(\cdot)$  is a heuristic function and is defined as follows:

$$n(\sigma) = \left( \frac{\sigma - \sigma_0}{\sigma_u} \right)^m \quad (22)$$

where  $\sigma_0$ ,  $\sigma_u$ , and  $m$  are properties of the material. The probability of failure,  $S$ , is equal to zero for  $\sigma < \sigma_0$ . The location parameter,  $\sigma_0$ , is not usually measured due to the practical difficulty in finding a minimum value for ultimate strength of material. Therefore  $n(\sigma)$  is commonly defined as follows.

$$n(\sigma) = \left( \frac{\sigma}{\sigma_u} \right)^m \quad (23)$$

There are different reported parameters in the literature. For SPF (Spruce-Pine-Fir) softwood lumber  $\sigma_u \approx 30$  and  $m \approx 3$  is used for simulation (covariance  $\approx 40\%$ ) ([23]). For Douglas fir, an average tensile strength of 14200 p.s.i. (102 MPa) and standard deviation of 3500 p.s.i. (26 MPa) is reported (covariance  $\approx 25\%$ ) in [36].

It was observed ([37]) that the size (the bigger dimension of the cross-section) of the specimen affects the strength and caused modification of this model ([38],[39],[40],[41]). However, it was shown that the size effect vanishes if the thickness of material is more than 30 mm ([42]). The modifications of this model that account for redistribution of stress due to the existence of crack as well as size effect were developed ([43],[44]).

### 3.2. The effect of grain angle on strength

Grain angle of wood is the angle between local longitudinal axis of wood grain and the longitudinal axis of the board. The average grain of a tree doesn't necessarily grow along the longitudinal axis of the tree. In some cases the grain may grow in a spiral form around and along the trunk of the tree ([45]). The average grain angle affects the strength of a small specimen and is described by the Hankinson equation ([46]). The impact of grain angle on strength is due to the fact that the strength of wood is directional. Tensile strength of wood along the grain is the highest while the tensile strength is lowest across the grain. Although related to the species of the tree, the longitudinal strength is usually many times larger in magnitude than the strength perpendicular to the grain. The

following equation (Hankinson equation) is an accepted relation between wood property (e.g., tensile strength or MOE) and wood grain angle ([47]):

$$N = \frac{PQ}{P \sin^n(\theta) + Q \cos^n(\theta)} \quad (24)$$

where,  $N$  is the desired property of the test specimen,  $P$  is the property along the grain,  $Q$  is the property across the grain, and  $\theta$  is the angle of the grain with respect to the stress direction.  $n$  was empirically derived, and is usually between 1.5 and 2 for tensile strength. In [48],  $P$  for tensile strength of Douglas fir is reported to be equal to 13800 p.s.i. (99 Mpa) and  $Q$  for the same tensile strength is reported to be 340 p.s.i. (2.4 Mpa). Although improvements to Hankinson equation were suggested ([49],[50]), this equation is still widely used.

In a board, not only the average grain angle affects the strength of the specimen, the local grain angle can affect the strength even more severely. Local grain angle is created by deviation of grain of the trunk of the tree into its branches. Also, the deviation of grain angle around a branch that is a consequence of natural growth of the tree can have considerable effect on the strength of a piece of lumber.

Another effect that is not as important, for strength estimation, is that the type of failure also changes depending on the angle between tensile stress and the grain. If tensile stress is perpendicular to the grain, failure in clear wood is like that of brittle material ([38]), while if the tensile stress is along the grain failure in clear wood is like that of tough material ([51]).

### 3.3. The effect of density variation on strength

One of the features of wood that was known for many centuries to be related to the strength was its density. Although not highly correlated with strength, the specific gravity of wood relates to the density of fibers of wood, i.e., the thickness of the cell walls. The correlation between density and strength is valid within a species, and even valid within the body of a tree. Each ring of a tree represents one year of its age. The light color part is early wood that grows in spring (when water is abundant). The cells that grow in this

season have bigger void at the center of the cell. The dark color part is related to late wood (summer and fall). The grain that grows in this season has smaller voids at the center of the cell and is denser and stronger than that of early wood.

One of the reasons why wood shows orthotropic behavior is this difference between the strength of the two types of wood in a ring ([38]). The body of a tree shows a cylindrical symmetry that arises from the relative axial symmetry of growth rings. Therefore, at every point the properties of clear wood are relatively constant in three major directions; along the tree, along the radius of the trunk, and tangent to the trunk (perpendicular to the other two axes).

Also, the difference between heartwood (close to the center of the tree) and sapwood (away from the center and carrying water) as well as juvenile wood (the wood developed when the tree was young) and the rest of the tree can affect the strength of a board.

### **3.4. The effect of holes on strength**

Although holes are not likely to be seen in a board, natural holes can be created in wood due to decay or missing knots. There are two effects that can be associated with holes; smaller effective cross-section of board and increased local stress arise due to the shape, size, and location of the hole.

Usually, the increased stress that is caused by the shape of the hole initiates the fracture, which propagates and results in the failure of the board. The holes with sharp ends that are pointing across the board cause largest local stress. Therefore even a small saw cut, because of its shape, could be more strength reducing than a larger circular hole. However, the shape of natural defects is not usually sharp.

### **3.5. The effect of knots on strength**

Knots may be the most common defects that exist in lumber. As the trunk of a tree grows radially, the base of each branch is gradually taken into the body. Different types of knot can be generated based on how the branch is incorporated in the trunk wood. A knot can

appear as a distinguishably defined cylindrical shape in a board, or it can appear as an acute change of direction in local grain.

Knots have two features that affect the tensile strength of a board. The first feature is that knot wood almost does not contribute to the strength of the board's cross-section at its location. Therefore, a knot is sometimes modeled by a hole in the board. However, the shape of the knot cannot always be simply modeled as a hole because there is also density variation from clear wood to knot wood that makes the effect of a knot more complicated. The second effect is that due to the existence of the knot, the grain of wood bears local variations. This variation of the grain reduces the local strength of wood around the knot. In practice, it is obvious that knots are the cause of failure in most cases ([11],[12]).

Visual grading that usually deals with the largest knot of a board is an example of empirical information that transforms knot features to board grades. The size of a knot and its closeness to the edge of the board are two important factors that reduce the strength. Knot area ratio represents what percentage of the cross-section of the board is actually taken by the knot. This, in fact, is the increase in the stress on clear wood section of the board cross-section.

### **3.6. Other factors**

Other factors such as wane, resin canal, and decay affect the strength but they are hard to model or learn by a learning algorithm. The reason is that these factors don't have any specific shape and density and that they don't happen frequent enough for a learning algorithm to be trained for their condition.

Any error in measuring the features of wood will have an impact on the estimated strength. This factor can be modeled as a shift in the feature space.

Some of the features of wood are not measurable. For example grain structure, which has a defining effect on strength, cannot be measured directly. Therefore, a heuristic approach to strength estimation will be limited to a specific species or a group of species.



### 3.7. A 2x4 board as a system

It is almost impossible to see a board without any defects. The impact of defects on strength is so severe that the failure of a board is almost always assigned to a dominant defect or defects. A comparison of board tensile strength and the tensile strength of small samples of clear wood from the same boards ([11]) shows low correlation. This comparison substantiates the observation of defect dominance on tensile strength.

The actual failure of a board under tensile stress usually takes place due to the interaction of more than one feature. For example an edge knot initiates the fracture, but the propagation of the fracture is dependent on the general strength of the wood. In practice, there can be more than one fracture that is initiated but the first one that results in the total failure ends the experiment. Therefore, a board should be considered as a system of wood features such that their interaction determines the tensile strength of the board.

Since a board usually has many defects in it, different strength zones can be defined (not always easily/possible). A study ([52],[53]) showed that if defects are closer than a certain distance to each other, they could create local weak zones. In practice, it can be seen that close knots, for example, create low strength zones. This is due to the fact that high density variation and high grain variation as well as high stress zones exist in a small region, which lets the fracture be initiated and propagated in a relatively low stress level as compared with the average strength of the board. This fact is not only related to knots but also related to local high grain angle (like spikes), resin canals, or some times local decayed zones. However, knots are the most common factors that cause the failure of a board ([11]).

One more factor that must be considered in board tensile strength estimation is that all the elements of a board are random variables. This means that all the elements of the system are random variables. Therefore the problem of board tensile strength estimation, in general, can be considered as a nonlinear mapping of input random variables (features) to the strength.

### 3.8. Proposed approach to modeling strength as a continuous function of wood properties

It is the goal of this thesis to find a mapping that transforms the measured properties of a board to its tensile strength. The place where a failure initiates and the path on which the crack will grow are not known. Although it is desired to get a detailed and accurate measurement of all the physical attributes of a board, the complexity of analysis will make the study practically impossible. Therefore, the problem of transforming a board's physical properties to its tensile strength leads to transforming selected features of measured profiles to board tensile strength.

Let's say  $X$  is the set of strength factors that should be measured in order to estimate the strength of the board. The set  $X$  is the set of all the necessary profiles. If  $X$  were known, the estimation error would be due to model development; i.e., board characterization (Chapter 4) and/or the learning method selection (Chapter 6). It is obvious that  $X$  cannot be fully produced by practical measurement. The measured profiles cannot be detailed enough to represent all the local variations of wood properties, as was obvious from the limitations of measuring machines of Chapter 2. For example the measured grain angle profile of the board can only be considered as a coarse presentation of three-dimensional grain-direction of all the points of the board. Also assuming that we know all the strength affecting factors, not all the factors are being measured due to unavailability of a suitable measuring machine and/or due to lack of reliability of an existing machine. The available set of measurable strength factors,  $\hat{X}$ , is a subset of  $X$ , and its elements of  $\hat{X}$  are usually the smoothed form of the elements of  $X$ . Any improvement of the existing measurement machines or any invention of new machines that measure a new property of wood would improve the estimation by making  $\hat{X}$  more similar to  $X$ .

One characteristic of the measured profiles, that is, the elements of  $\hat{X}$ , is that each one can be a single number, a vector, or an image (matrix). Any change in the measurement machines affects  $\hat{X}$ . Figure 3.1 shows the problem of strength estimation as a learning problem ([21],[54]). In this figure,  $x$  is a vector that represents  $X$ ,  $u$  is the vector that is not related to the output but can be a part of the input feature set as a part of exploratory

data analysis,  $v$  is the vector of features that are included in the generated feature set and represents  $\hat{X}$ .  $g(x)$  is the actual process and  $\hat{g}(\hat{x})$  is the model that is developed by the learning machine.  $y$  is the actual output and  $\hat{y}$  is the estimated output.

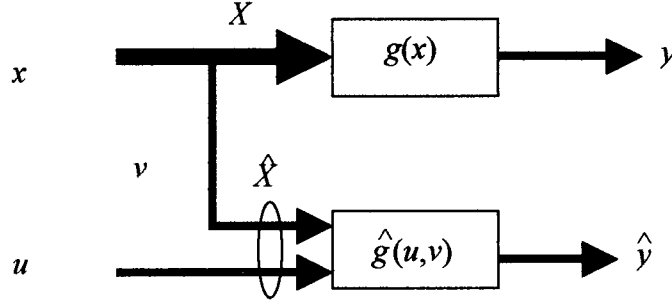


Figure 3.1. Strength estimation as a learning problem.

To reduce the dimensionality of the problem the measured profiles,  $\hat{X}$ , can be transformed into a feature space,  $F$ , that represents  $\hat{X}$ . For example, a board can be identified by its defects and a general description of its average wood properties. The feature space,  $F$ , could also be very high dimensional and is not necessarily unique. Since there aren't any rules in defining the features, there can be many redundancies in the feature set or one may define a feature that is contributing to strength estimation.

Some *a priori* knowledge about any dependence between the strength and the feature set helps in developing a more accurate model. As will be discussed in the Appendix, the problem of wood failure under tensile stress was studied in the past. Therefore, it is desired to choose a set of features that are physically meaningful and best represent  $F$  so that the existing knowledge of strength of material can be used at model development stage that will be discussed in Chapter 6

The selected feature space can be denoted by  $\hat{F}$ . Since the tensile strength,  $S$ , is a random number it can be assumed that it is a hyper-surface in the measurable physical features space, with added uncertainty. The estimated tensile strength of the board,  $\hat{S}$ , is the hyper-surface in  $\hat{F}$  that estimates  $S$ .

A board can be related to a point in  $\hat{F}$ . The estimation process requires the transformation of this point to  $\hat{S}$ . As the point is transformed from one space to the next some uncertainty, or variance, is added to it. Once the methods of measurement, feature extraction, and model development are chosen the estimated strength for a given point in physical space is a random variable.

Let's say the estimated strength of a board with measurable features  $\hat{F}$  is denoted by  $s(\mu, \sigma, \lambda)$  where  $\mu$ ,  $\sigma$ , and  $\lambda$  are location, scale, and shape factors, and  $s$  is a random variable with known distribution ([55]). The three factors of the distribution are functions of the feature ( $\hat{F}$ ), assuming that boards are uniquely identified in the feature space. Therefore, the tensile strength of a board can be studied as a random function (for the random nature of material failure) of the features as follows.

$$s(\mu, \sigma, \lambda) = s(\mu(\hat{F}), \sigma(\hat{F}), \lambda(\hat{F})) = g(\hat{F}) \quad (25)$$

## ***Chapter 4. Characterizing a board; features***

There are two classical approaches to solving the strength estimation problem. One approach is based on the assumption of having relatively accurate knowledge about wood properties at every point of the board. This approach leads to an analytical approach that can be used in a method like linear elastic finite element modeling ([38],[64]). On the other hand, there have been estimation methods that were based on general properties of wood ([65]). In this case general features of the board are mapped to the strength of the board. The existing standard of wood grading is based on such an approach.

The strength estimation methodology of this thesis is something between the two approaches but closely related to the second approach. A model of board structure (fundamentally similar to what is done in the first approach) was developed. Then the characterizing features of the board were extracted from this model. Finally, the features was mapped to the strength of the board.

In this method, the failure of wood is considered to be a function of general properties of wood at a section of the board. The basic assumption here is that wood properties don't change much in a small region. Therefore, a few (although it is not known how many) features of that section can characterize it. Since the exact location of fracture initiation and the path of fracture propagation are not needed, the features should be mapped to a single value that is the final stress level that leads to the failure of the board. It should be noted, as was explained in Chapter 3, that the strength of a board section is generally a random variable. However, the variance of strength at a point in the feature space is not studied in this thesis and the average of the strength distribution was estimated. Therefore, the strength of a board can be modeled as  $g(\hat{F})$  where  $g(\cdot)$  is a real function of the real valued feature vector  $\hat{F}$ .

The problem of characterizing the board for strength estimation is basically an attempt to define the weakest part of the board and the general properties of the board. Therefore

two types of features were produced; one type for the weakest section and one type for general properties of wood.

It may seem relatively straightforward to detect the weakest section or sections of the board and develop the strength estimation model for that section. The underlying assumption here is that the weakest section can be indicated and its features can be used for strength estimation. This deduction arises from the fact that the causes of many failure cases are major defects that can be easily seen.

If no distinct defect exists, interaction of defects can go beyond the neighboring defects. Therefore, it is not always clear as how to distinguish different sections of a board. Therefore, characterizing a board cannot be considered a trivial task. Also, it can be expected that the feature set to be almost definitely noisy.

As was explained above, a single major defect or a combination of defects can cause the failure. In many cases of failure a single defect or a group of defects can cause the failure of the board. The section of the board that has the dominant defect(s) is the weakest section of the board. In such cases a feature that represents the worst defect is needed in the representative feature set. The minimum of the MOE profile is related to the weakest section of the board or equivalently to the biggest defect. Also, in many cases the board contains defects that are similar. In such cases there isn't one distinct defect that causes the failure. These defects increase the probability of the failure at lower stress levels. Therefore, a set of features that represents the overall existence of defects should be included.

From Chapter 3 it seems obvious that there are two basic factors that govern wood strength estimation; wood strength at a point or a section of the board and the structure of the board. The former one defines how strong the material is and the latter one defines the stress distribution and stress concentration that has a definite reducing impact on strength. The stress distribution in a board is not going to be studied in this thesis but a set of features that seem to be related to the stress concentration should be included in the representative feature set.

The definition of features depends on the measurement constraints. The goal is to produce a set of features that are most relevant to strength or failure and can be accurately

measured or calculated by using the measured features. The understanding of measurement constraints is important at this stage because some of the features that may seem most relevant may not be measured properly. The measurement systems and their restrictions will be explained in Chapter 5. Local density, grain angle, and bending modulus of elasticity are the measured profiles. The representative features are therefore related to one of these measured properties. Obviously, the unmeasured properties of wood will add to the uncertainty of the estimation, as was explained in Chapter 3.

#### **4.1. A general view on defining the features**

The need for extracting a feature set from the measured profiles of a board arises from the fact that analyzing the measured profiles of the board is practically impossible because they contain numerous redundant or highly correlated pieces of information. For reducing the unnecessary pieces of data, features of the measured profiles are used to develop a model of the board. These features are basically the characterizing features of the board. The features that will be introduced serve a double purpose, they represent a board (distinguishing different boards) and they represent the strength-related physical features of the board. Therefore the representative feature set needs to be a minimum necessary size. In order not to compromise prediction accuracy for feature meaningfulness, the method of Figure 4.1 can be followed. The meaningfulness of the features makes them different than the features that are usually introduced in dimensionality reduction techniques as these techniques map (linearly or nonlinearly) the given set of features to a fewer number of features. For example, in a three layer neural network with few hidden layer neurons, the output of each neuron is dimensionally reduced feature. Obviously in a problem like wood strength estimation such a technique cannot be much of help in using any existing information from other studies. However, if the X-ray image is transformed to the board model, it can be transformed to Knot Area Ratio (KAR), a feature that has been studied, by other researchers ([13]), as an effective strength-affecting feature.

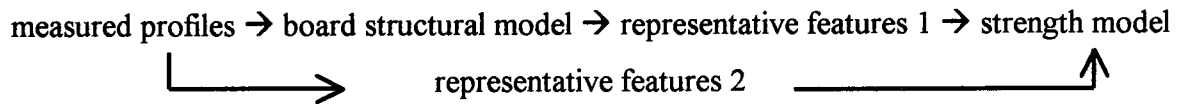


Figure 4.1. Steps of strength estimation method.

The board structural model is the generalization of the measured profiles. This model is developed based on the existing knowledge of the constraints that should exist on the measured profiles. However, the model tends to be simple and tends to remove the anomalies that are not considered in the model. For example, a small-size density variation is usually considered a measurement error. Therefore the measured profiles can generate features that can add to the generated features through the board model, shown as *representative features 1* in Figure 4.1.

A learning system, shown as the “strength model” in Figure 4.1, uses a fixed length of features. Therefore, the feature set should have fixed length. However, the fixed length feature set imposes restrictions on the estimation that can best be studied by the complexity of signals. In the following, the effect of a fixed number of features on the estimation accuracy is discussed.

It is obvious that the complexity of a board structure is variable. A board with many small defects is structurally more complex than a board with few major defects, although the strength of both boards may be the same. For example if a board with twenty knots is to be characterized, twenty sets of features for the knots and a set of features for the clear wood is needed. But if a board with three large knots is to be characterized, only three sets of features for the knots and a set of features for the clear wood are needed. Since transformation of input to output needs a fixed length of features, the feature sets of clear wood and defects will be transformed into a fixed length vector of features that is the board feature set.

It may seem that the board model is not needed because one could extract a set of representative features directly from the measured profiles and feed those to the strength estimator (Figure 4.1). The board model allows one to develop a set of features that are physically meaningful, for example the features that are related to structure of the board.



By using this model the existing knowledge from fracture mechanics and wood grading can be used in order to develop a meaningful representative feature set. This fact will become obvious when knot feature extraction is involved.

The characterization of a board starts by categorizing its strength affecting factors. The factors that affect the failure (or strength) of the board are distinguished and are extracted from the measured profiles. A set of features that best characterize each factor will be produced. Based on the feature sets of all the factors, a set of features that best characterizes the board will be generated.

The strength factors are categorized into clear wood and defects. The average properties of clear wood contribute to the strength of the board. The defects, on the other hand, are the strength reducing factors. The measured profiles are the density, grain angle, and modulus of elasticity. The strength model is a learning system (parametric and non-parametric) with fixed length of input vector.

In order to produce a fixed length vector of features as the representative feature set of the board, the statistical representation of the features is used, i.e., the similar features of defects in a board are represented by their statistical features. For example, a feature of a knot is its volume, but a representing feature for the related board is the average of all knot volumes. The clear wood of a board will be represented in the same fashion. The clear wood can be represented more accurately because it is one entity whose properties change gradually.

This method is based on the regularity of the defects. If the defect sizes are about the same and their strength reducing effect is about the same, then this method can represent the probability of failure at lower strength levels.

## **4.2. Properties of clear wood**

Characterizing the clear wood of the board is basically defining what is regular in the board. For that purpose, the regular part of a measured profile will be defined and will be extracted from the measured profiles. Here, a definition of clear wood and its attributes is provided and will be used to set the methodology for clear wood feature extraction and

clear wood modeling. Naturally, every measurement technique imposes some constraints on the measured property and therefore the methodology of feature extraction will be developed accordingly.

The clear wood of a board is the non-defect wood that is related to the regularly grown portion of the trunk of the tree with two distinct attributes. The first attribute is that this portion of a board contributes most of the body of the board. The second attribute is that the properties of clear wood are either constant or vary gradually. However, randomness of properties can be expected because clear wood is a natural product. The first attribute, statistically speaking, states that the chance of a point being clear wood is more than that being a defect. The second attribute leads to using averaging or using regression for extracting the clear wood properties from the measured profiles. These attributes are the basis for extracting clear wood properties from the measured profiles.

A density profile is a two dimensional profile in the form of an image. Every pixel of the image is proportional to the density of that point of the board. If the pixels of the image are considered as random numbers and presented as a histogram diagram, clear wood contributes the major shape of the distribution. The statistical mode of the distribution is the most common density of the board and defines the average density of clear wood. Also, the natural variation of density in clear wood is obviously a second feature. This feature is a strength-reducing factor (Figure 4.2). Therefore, the density of clear wood at every point of the board is modeled by its average density and by its natural variation that is modeled by the standard deviation of the histogram diagram.

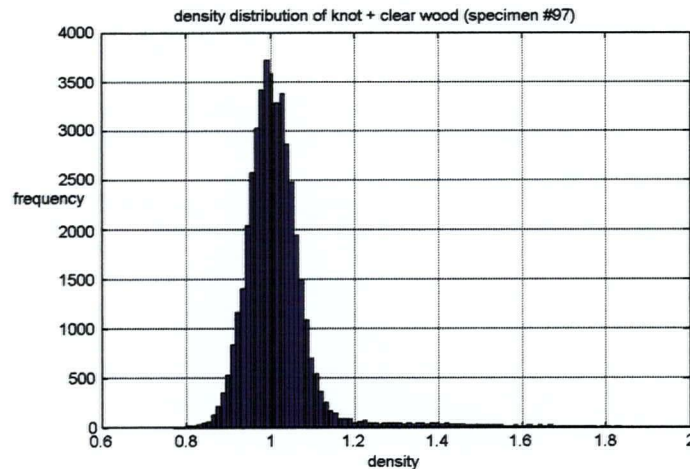


Figure 4.2. The histogram representation of density distribution of a board

If the clear wood density is assumed to be constant at all clear wood points of the board, the mode of this distribution is the natural choice (Figure 4.2). The mode of distribution is the simplest form that can be used to model the clear wood density. Although, the average of the distribution is usually close to the mode, the mode value has the slight advantage that it is not affected by large density affecting defects like decays or resin canals. The standard deviation should be independently stored as a feature. Also if the density is assumed to change gradually along and across the board, a plane that is a linear regression of the image can model the clear wood density at every point. In geometrical defect detection it will be shown that two images are needed to develop the model of the defect. One image is related to the wider face (plank) of the board, that is, the top image, and one is related to the narrower face (edge) of the board, that is, the side image. The two images can produce the density of all the points of the board.

The grain angle profiles of the board cannot be analyzed in the same way as done above. The sampling rate in this measurement is low therefore the number of measurements in this profile is considerably less than that of density profile. The sharp local grain angle variations that can be seen in the grain angle profile are related to the grain variation around knots. The shape of grain angle variation in the measured profile depends on the relative location of the measurement head with respect to the center of the knot. The histogram representation does not generate a clear shape of distribution (Figure 4.3)

because the defects usually take a large portion of the profile. The grain angle features of clear wood should be extracted from the grain angle profile. The average and standard deviation of the measured grain angle profile are considered as the grain angle features of the clear wood.

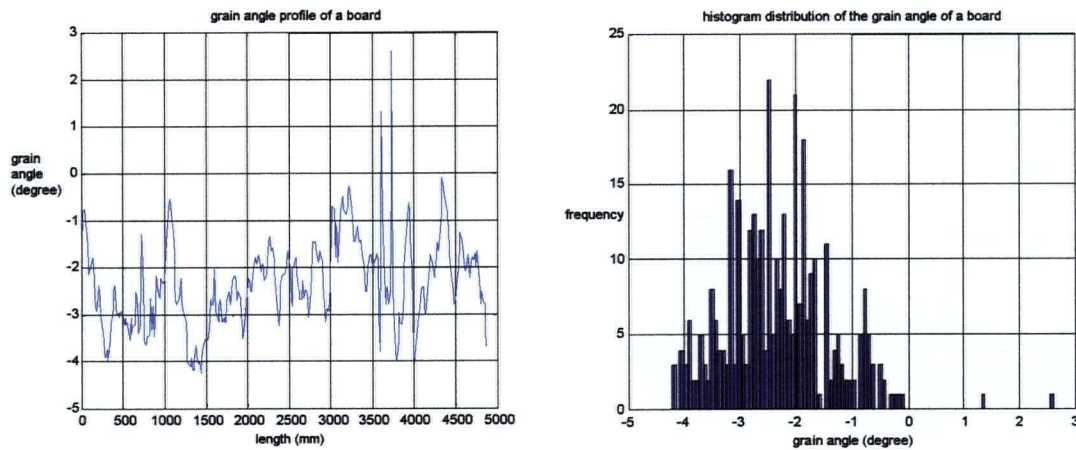


Figure 4.3. Grain angle profile of a board and its distribution.

The bending modulus of elasticity measurement of a board is an average over a span of the board. This profile does not vary much and the sampling rate is low. Therefore, similar to grain angle profile, the profile is better for feature extraction than its histogram diagram (Figure 4.4). However, unlike the grain angle profile the defects of the board do not create any sharp changes in the profile. Therefore, the local measurement of the MOE of a point can be assigned to the cross-section of the board at that point. Although the variation in the MOE profile is gradual, the minimum of this profile usually indicates a weak zone in the board that is either related to the major defects of the board or a weak clear wood zone. This feature is not related to clear wood general properties and will be considered as a feature of the entire board.



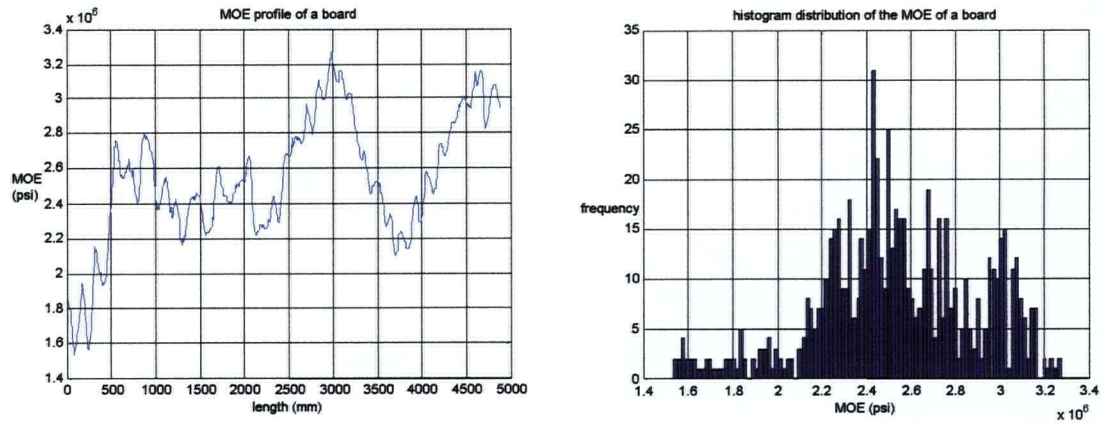


Figure 4.4. The MOE profile of a board and its distribution.

### 4.3. Characterizing the defects

Characterizing the defects starts with detecting the defects from the measured profiles. Defect detection can be categorized into three problems: detecting general variations, sharp local variations, and extreme values. The general variations were explained in Section 4.2, along with clear wood properties, in the form of standard deviation of the measured local density. The local variation detection is based on the existing difference between the density or grain angle of the defect and the clear wood. The extreme values are the minimum values of wood density and modulus of elasticity, which are the weak points of the board.

Knots are the distinct defects that most commonly exist in wood. They were studied and modeled in detail by using their geometrical features. This is possible because of a knot's defined shape as well as its distinct density difference with respect to clear wood density. Each representing feature will be an indication of the defect's size, location (across the board), or shape. A grain deviation appears as sharp changes in the profile (Figure 4.3). This can be detected by using thresholding techniques. No other specification is assigned to it because the shape and size of the variation depends on the relative location of the measurement head with respect to knot location. It will also be shown that the grain variation can be modeled by having the knot model.

These attributes of a feature are basically the properties that similarly affect the strength. The size of a defect reduces the clear wood material. The shape of the defect affects the high stress zones that are generated by the defect when the tensile stress is applied. Also, the location of the defect across the board influences the strength reducing effect of the defect. If a similar defect is closer to the edge of the board, it more likely initiates a fracture.

Also general variations of clear wood or general minima of strength factors are the features that reduce the strength. Therefore, the standard deviation of the measured properties (from clear wood values) in Figure 4.2, Figure 4.3, and Figure 4.4 are the features that can represent some of the defects of wood. Also, the minimum of the measured profile for MOE (Figure 4.4) and the minimum of density across the board are the features that are related to weak points in the board. The minimum of density across the board is the average density in a small section of the board (Figure 4.5). Such general minima or maxima are not seen in grain angle profile of the measured samples.

When the specific features cannot be defined, general features that may be related to the strength can be defined. For example a grain angle defect may not have a specific shape or size or location but its maximum grain deviation defines the weakest part of that defect and is related to the likelihood of failure at that point. Similarly a decayed zone can be identified because of its regional reduced density level but a specific generic shape cannot be assigned to it. Also, the location across the board cannot be assigned to the decay zone because its large size makes this feature meaningless.

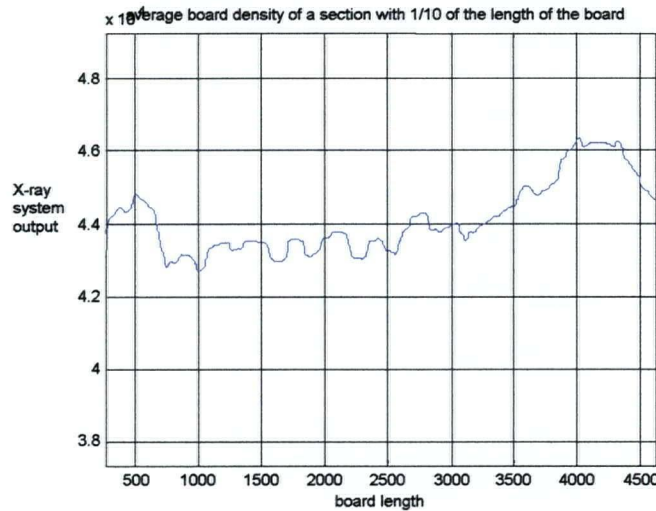


Figure 4.5. The local density profile of the board generated by a moving window. Window span is 1/10 of the length of the board.

#### 4.4. Geometrical features of a knot

In this section the geometrical features of a knot will be explained. The geometrical features of a knot are the features that are obtained by modeling the shape of the knot. The features characterize a knot's size, its shape, and its relative location with respect to other knots. The creation of a knot is due to the existence of a tree branch, which has a defined circular cross-section. The remainder of the branch keeps its defined cross-section shape and leads to the geometrical model that will be produced in Chapter 5.

The knot can be detected in the density image of the board because its density is usually distinctly different than the density of clear wood. Figure 4.6 shows the X-ray density profile of a knot in a piece of wood. The three-dimensional image of such a knot can be imagined as what is shown in Figure 4.7. The flat part of the density profile in Figure 4.6 is related to clear wood that surrounds the knot and the protrusion is related to knot wood. It will be shown in Chapter 5 that two density images of knot that are taken from mutually perpendicular directions are sufficient to identify the model of the knot (Figure 4.7).

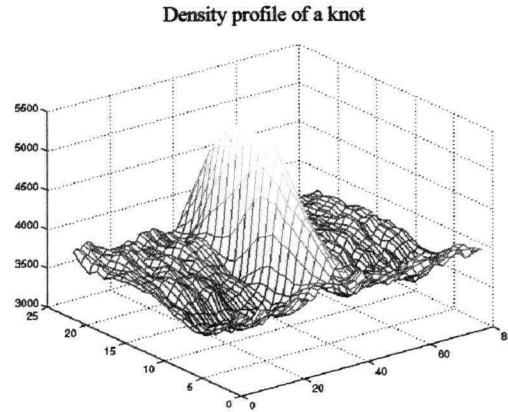


Figure 4.6. The density profile of a knot as compared with the density of clear wood of a board.

This model is based on the generalization of the cylindrical shape of a branch. Therefore a knot is modeled by a cone, which is cut by two parallel planes. The cone represents the knot and the parallel planes are the surface of the board. The features of the geometrical shape are the features of the related knot. More detailed features can be extracted from this shape because of its increased accuracy as compared with other defects.

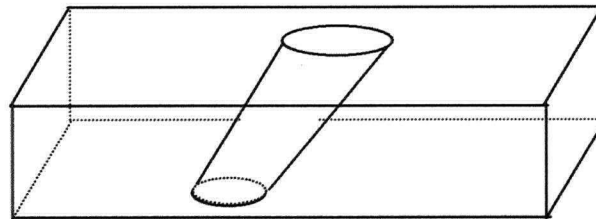


Figure 4.7. Conic model of a knot.

Knot volume, its axial deviation, the taper, the average radius, the distance to the closest knot, and knot-area-ratio are the features of a knot. The axial deviation of the knot is the standard deviation of all knot points from the central line of the board. Knot taper is the ratio of the smaller diameter to the larger diameter. Knot average radius is the average radius of the knot cone in the board. Knot area ratio is the maximum of the ratio of knot cross-section to the board cross-section. The Knot Area Ratio (KAR) is a visual quality level ([13]) measurement that is used in industry for down grading a piece of lumber if a large knot exists in it.



Knot volume and average knot radius represent the reduction in the material of the surrounding clear wood area. Similarly, knot-area-ratio represents the minimum clear wood material at the cross-section of the board where the knot is. The axial deviation of the knot from the center of the board shows how close the knot is to a corner of the board. Knot taper represents the possible existence of the tip of the knot in the board that can create a high stress zone. The distance of the knot to its closest neighbor can be used to represent the expected increase in stresses that can be created from knots that are very close to each other.

#### **4.5. Characterizing feature sets are not unique**

In defining a set of features, it should be noted that different sets of features can be defined that represent similar properties of wood. It is obvious that any transformation of a feature or a group of features will generate the same estimated strength as long as the learning system can learn the inverse of the transformation. Since there aren't any rules for defining features, the definition of features that are representing the same factor is avoided. For example, the average and the mode of the density histogram of a board (Figure 4.2) are two features that are conceptually representing the same factor. The mode of the board's density profile is the regular density of the clear wood of the board, which is related to the strength of the board. The average of the X-ray image can be considered the same as long as the defects are neglected.

In defining the representative features of a board their physical meaning will be considered. Therefore, the emphasis is on the features that are related to board model as opposed to those that are related to the measured profiles (Figure 4.1). The advantage of this methodology is that the existing knowledge from wood grading and wood mechanics can be used for feature definition.

## 4.6. Board model

The board model is the basis of generating a set of features that physically make sense. For that matter, the clear wood properties that were explained in Section 4.2 are extracted from the profiles and inserted at all the points of the board model. By using the geometrical features of knots a basic three-dimensional model of the board can be generated (Figure 4.8). This model is a structural model of the board. The knots are detected by using the geometrical method and are modeled by holes of the same shape.

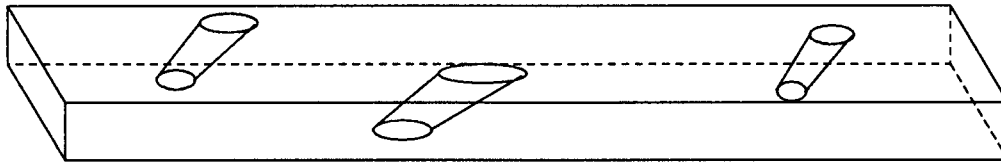


Figure 4.8. The model of a board that consists of the clear wood and defects.

The clear wood part of the model at every point indicates the density, grain angle, and modulus of elasticity of wood at that point. The density of wood at every point is the average of estimated density from the two density images after removing the knots. There may be more than one grain-angle profile of the board. The grain angle of wood at a point is resulted from the interpolation of the measured profiles after removing the anomalies (the trace of the defects like knots) in the profile. The modulus of elasticity profile is usually measured as a pair, one from each side of the board. The average profile shows the average MOE of the measured points. The MOE of all the points in a cross-section of the board are assigned the value of MOE at the related point in the measured profile.

It is obvious that the complexity of the characterizing feature set increases as the number of the defects changes. This can be seen from the fact that if each defect is characterized by a set of features the total number of features increases if the number of defects increases ([66],[67]). The complexity of the characterizing feature set can be related to the accuracy of the strength estimation of the board.

As was mentioned before, the characterizing feature set of the board is a fixed length vector of features. In order to transform the variable number of features to a fixed number

of features the statistical moments were used. The clear wood features and general variations of clear wood were defined before. The features that are related to the board model are the features that are related to geometrical features. Knot volume, its axial deviation, the shortest distance to a neighboring knot, taper, and knot area ratio are the features that were added.

## ***Chapter 5. Extracting features from measured profiles***

In this chapter the extraction of features from the measured profiles is presented. The features, that will represent a board for strength estimation, were discussed in Chapter 4 and the measured profiles were presented in Chapter 2. Statistical features are used for representing the grain angle measurement and the modulus of elasticity measurement. The X-ray images, however, are used for deriving the location and shape of knots. Most of this chapter is related to detecting knots in an X-ray image and modeling the shape of the pattern.

As was explained in Chapter 4, a model of board density is produced by using the X-ray scanning of the board. This model consists of the knots of the board and clear wood density distribution. A knot produces a local protrusion (a bump) in the regular surface that is the X-ray image of the clear wood. Since the X-rays are almost parallel, the edge of this bump is a pattern that can represent the shape of the knot and will be used to extract the parameters of the knot model. It will be shown that two patterns are sufficient to produce the model of the knot. The three dimensional shape of a knot and its location in the board is the basis of geometrical features of the board.

Shown in Figure 5.2 is an X-ray image of the wood density of a board and the surrounding area. The surrounding area is related to the portion of the measured profile where there is no board in the system and the system output is the noise level. Since the noise level in all the measured profiles is considerably lower than the signal level, the portion of the measured profile that is related to the board can easily be separated from the surrounding part of the profile by thresholding.

### **5.1. The image of a knot in an X-ray image; the shadow image**

As was explained in Chapter 4, the density image of a knot is generated by X-ray scanning of the board. Figure 5.1 shows the density image of a knot, which is a part of an

X-ray image. As is shown in this Figure a knot can be seen as a local peak in a relatively flat surface that is related to clear wood of the board. Because of the large density difference between clear wood density and knot wood density (as can be seen in Figure 5.1) a threshold level can separate knot pixels in this image. This is the basis of knot detection that will be derived in the following sections.

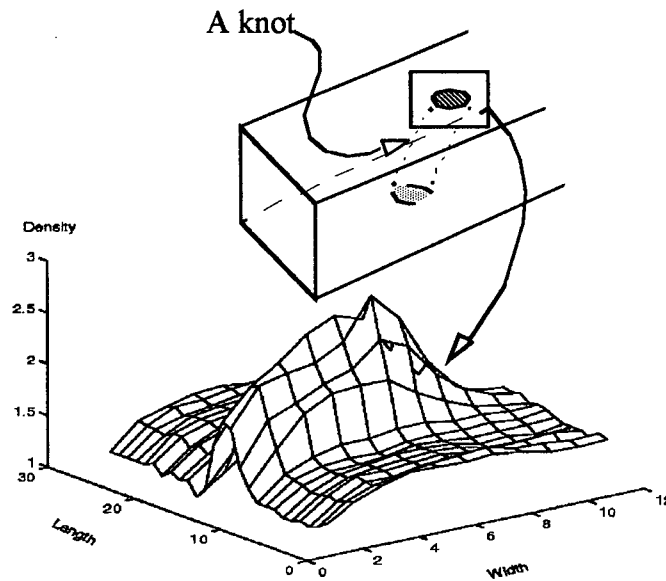


Figure 5.1. The X-ray image of a knot

A contour map can show the existence and location of knots in an X-ray image. Figure 5.2 shows the contour map of a measured profile. The two relatively vertical lines are the edges of the board (the surrounding area were removed) and the closed elliptic contours are related to the knots. Each closed elliptic contour is related to a density variation similar to Figure 5.1.

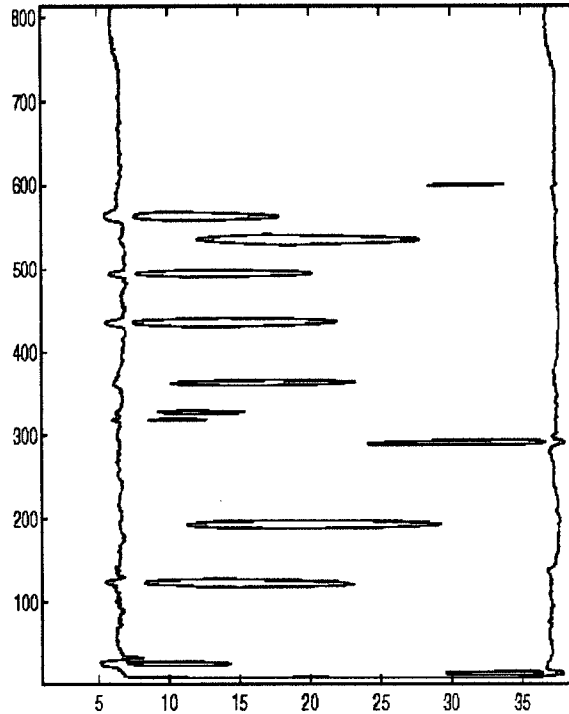


Figure 5.2. Contour map of an X-ray image.

## 5.2. Thresholding; classifying a pixel into knot or clear wood classes

The existence of knots and the related wood density variation makes the density distribution of the board a bimodal distribution ([68]). In order to classify the image pixels into two classes of clear wood and knot wood, Bayesian based thresholding was used. The density of wood is assumed to be normally distributed, therefore, knot wood and clear wood can be modeled by two normal distributions whose averages and standard deviations were measured from the existing sample boards. For example, the distribution of density in knot and clear wood of a board (as will be discussed in detail later in this chapter) from a randomly selected board shows that the mean and standard deviation of clear wood distribution (assuming a normal distribution) are  $\bar{x}_c = 1$  and  $\sigma_c = 0.05$ ,

relatively. The mean and standard deviation of knot wood distribution are  $\bar{x}_k = 1.39$  and  $\sigma_k = 0.44$ , relatively. The distributions are shown in Figure 5.4.

The problem of separating the image pixels into knot pixels and clear wood pixels was cast as a Bayesian decision-making problem. Every pixel of the image should be classified into one of two populations with known probability distributions (Figure 5.3). In deriving the threshold level two points were kept in mind, detecting knots and preserving the shape of the knot pattern. Detecting small variations in clear wood density of the board was not of interest here. The knot patterns, on the other hand, were specifically important because they were transformed to geometrical features of the board. Therefore, the threshold value was chosen so that it generates the least chance of error in the output knot pattern.

Although the density distribution of clear wood pixels of the X-ray image is related to the natural variation in wood density, assuming that knots are the only defects of the board, the density distribution of knot pixels is affected by different factors as is explained in the following. The density distribution of a knot depends on the type of the knot. For example, if the density distribution of knot wood is modeled by a Normal distribution the mean of the distribution changes as the type of the knot changes. There are few types of knots therefore one can expect few possible density distributions. The threshold level is specific to the two density distributions. A single threshold level imposes error in classification procedure. However, this error cannot be avoided because the types of knots cannot be distinguished by using their X-ray images.

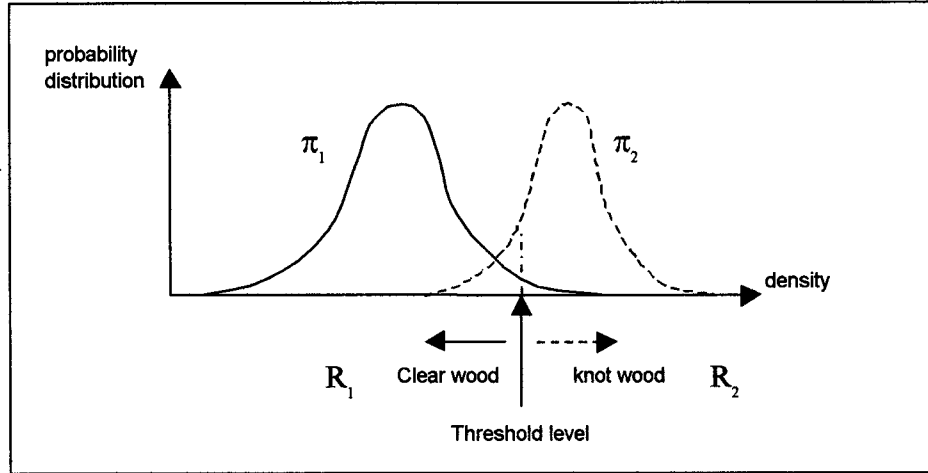


Figure 5.3. Probability distribution of clear wood density and knot wood density.

The actual statistical distribution of a knot-related X-ray image pixel is more complex because it comprises two components: the statistical distribution of knot wood density, and the statistical distribution due to the projection of a knot cone onto the image plane. For a tilted knot, for example, the density value of a pixel changes gradually from the knot wood to a combination of knot wood and clear wood depending on how far the pixel is from the center of the knot.

### 5.2.1. Bayesian classification

Let's say  $\pi_1$  is the population of all the clear wood pixels in an X-ray image and  $\pi_2$  is the population of all the knot wood pixels in the same image.  $\mathcal{A}$  is the population of all the image pixels ( $\pi_1 \cup \pi_2$ ).  $X$  is a random variable that is defined as the measured density at a pixel in the X-ray image.  $x$  is an observation, i.e., it is the value of a pixel in an X-ray image.  $R_1$  and  $R_2$  are the two regions of  $X$  that define clear wood and knot wood respectively.  $q_1$  is the *a priori* probability of a pixel belonging in  $\pi_1$  and  $q_2$  is the *a priori* probability of the same pixel belonging in  $\pi_2$  ( $q_1 + q_2 = 1$ ).  $p_1$  is the conditional probability distribution function of  $\pi_1$  and  $p_2$  is the conditional probability distribution of  $\pi_2$ .  $C(2|1)$  is the cost of misclassifying a pixel into  $R_2$  (i.e., classifying it as a knot pixel) while it belongs in  $R_1$  (it is a clear wood pixel). Also,  $C(1|2)$  is the cost of misclassifying a pixel into  $R_1$  (i.e., classifying it as clear wood) while it belongs in  $R_2$  (it is a knot pixel).



The conditional probability of the pixel belonging in  $\pi_1$  (clear wood) given that its value (the observation) is  $x$  can be expressed as follows:

$$p_1 = \frac{q_1 p_1(x)}{q_1 p_1(x) + q_2 p_2(x)} \quad (26)$$

The cost of misclassification is

$$C = C(2|1) q_1 \int_{R_2} p_1(x) dx + C(1|2) q_2 \int_{R_1} p_2(x) dx \quad (27)$$

$R_1$  and  $R_2$  are chosen so that:

$$R_1: \frac{p_1(x)}{p_2(x)} \geq \frac{C(1|2) q_2}{C(2|1) q_1} \quad (28)$$

$$R_2: \frac{p_1(x)}{p_2(x)} < \frac{C(1|2) q_2}{C(2|1) q_1} \quad (29)$$

The probability that a pixel is clear wood is much higher than that it relates to a knot, therefore

$$q_1 \gg q_2 \quad (30)$$

Since  $q_1$  and  $q_2$  are the probability of a pixel belonging in a knot and clear wood correspondingly, they can be estimated by using the total clear wood area and total knot area in a sample X-ray image.

The location, size, and shape of a knot will be detected by using the edge of the particular knot in the X-ray image. Extracting these parameters is the whole purpose of edge detection. If too many pixels are misclassified from knot to clear wood it disfigures the shape of the knot edge pattern, therefore affects the rest of the process. If the threshold level is set to a higher level the resulted parameters of a knot will be biased to a lower value (because the size of the pattern will be smaller), which can be tolerated. Therefore, the cost of misclassifying a clear wood pixel to a knot pixel ( $C(2|1)$ ), is much higher than the cost of misclassifying a knot pixel to a clear wood pixel ( $C(1|2)$ ).

$$C(2|1) \gg C(1|2) \quad (31)$$

Therefore  $R_1$  is for all  $x$ 's that are non-zero and  $R_2$  starts where  $p_1(x)$  diminishes to almost zero at a high percentile point. If  $x_T$  is the threshold level,

$$R_2 = \{ \forall x > x_T \mid p_1(x_T) = \frac{C(1|2) q_2}{C(2|1) q_1} \cdot p_2(x_T) \approx 0 \} \quad (32)$$

If  $x_T$  is a high percentile (for example 99 percentile) of  $p_1(x)$  it satisfies equation (32).

### 5.2.2. Edge detection

As can be seen from previous section, knot edge detection in an X-ray image of a board relies on finding  $p_1(x)$ ; clear wood density distribution, by using the X-ray image. The frequency distribution (a histogram diagram) related to the X-ray image pixels is the summation of two normal distributions with different average and standard deviations. Knowing the fact that the density of wood changes in a very limited range as well as equation (30), one can conclude that  $p_1(x)$  has a dominating effect regarding the overall statistical measures of the combined frequency distribution. Therefore, the mean and the standard deviation of the overall distribution can be measured and used directly in  $p_1(x)$ . In the following it is shown that the statistical mean of  $p_1(x)$  is almost equal to the statistical mean of the overall distribution:

$$m_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(1)} \quad (33)$$

$$m_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_i^{(2)} \quad (34)$$

where  $x_i^{(1)}$  is a pixel that belongs in  $p_1(x)$  and  $x_i^{(2)}$  is a pixel that belongs in  $p_2(x)$ .

$$\begin{aligned} m &= \frac{1}{n} \sum_{i=1}^n x_i \\ &= \frac{1}{n} \left[ \left( \sum_{i=1}^{n_1} x_i^{(1)} \right) + \left( \sum_{i=1}^{n_2} x_i^{(2)} \right) \right] \end{aligned} \quad (35)$$

$$= \frac{1}{n} [n_1 m_1 + n_2 m_2]$$

$$m = q_1 m_1 + q_2 m_2 \quad (36)$$

Equation (30) and  $q_1 + q_2 = 1$  result in the following:

$$m \approx m_1 \quad (37)$$

The standard deviation of the overall distribution does not have such a simple relationship with the distribution of the overall distribution as shown in the following. The standard deviation of the overall distribution  $s$  is as follows.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \quad (38)$$

$$\approx \frac{1}{n_1-1} \left[ \left( \sum_{i=1}^{n_1} (x_i^{(1)} - m_1)^2 \right) + \left( \sum_{i=1}^{n_2} (x_i^{(2)} - m_1)^2 \right) \right]$$

The first term of equation (38) is the standard deviation of the clear wood distribution. Usually  $m_2$  is a multiple of  $m_1$  therefore the second term of equation (38) is not negligible.

The standard deviation of clear wood can be estimated by using the portion of the overall distribution that is mostly contributed by clear wood. Assuming that the standard deviations of the two distributions are in the same order, one can conclude, from equation (30), that the maximum frequency of knot wood distribution is much less than the maximum frequency of clear wood distribution. For example for  $q_1 = 10 q_2$ , the maximum frequency of clear wood distribution is about ten times as much as the maximum frequency of knot wood distribution. This fact can be seen in Figure 5.4-a where the clear wood distribution obviously dominates the overall distribution (Figure 5.4-a, d, and e).

The samples within a neighborhood of the mode value of the overall distribution is a censored representation of clear wood density distribution and can be used for estimating the standard deviation of clear wood distribution. A censored sample of a distribution is the sample of the distribution when a part of the distribution cannot be sampled. The limits of this neighborhood is defined by a threshold level, which should be more than

$q_2 / q_1$  in order to avoid the maximum frequencies of knot wood distribution. If  $q_1$  is assumed to be almost equal to  $10 q_2$ , the threshold level should not be less than 10%. The limit of the neighborhood can be set to be 20% of the mode frequency of  $p_1(x)$ . The standard deviation of the clear wood density distribution is equal to the estimated standard deviation of the pixels within this neighborhood.

Once the parameters of  $p_1(x)$  are estimated the threshold level for knot detection will be one of its high percentiles (like 99 percentile).

### 5.2.3. Experimental data

The threshold level detection is studied on a typical sample. The overall statistical distribution of the density of wood is derived from the X-ray image of the sample. The actual density distributions of clear wood and knot wood were extracted subjectively for comparison. The statistical mean and standard deviation of the overall distribution will be compared with the statistical mean and standard deviation of clear wood distribution. The mean of the overall distribution will be used as the estimation of the mean of clear wood distribution. The 20% maximum frequency neighborhood will be used to estimate the standard deviation of clear wood density distribution. The threshold level for knot edge detection will be derived based on the actual and estimated clear wood density distributions. The cost of misclassification will be derived for each case.

The distribution of wood density (clear wood and knot wood) is shown in Figure 5.4-a. This diagram is the density distribution of all the pixels of the X-ray image of specimen #97. One can easily see the dominating effect of clear wood density distribution by comparing the overall density distribution (Figure 5.4-a) with clear wood density distribution (Figure 5.4-e) and with knot wood density distribution (Figure 5.4-d).

The actual knot wood density distribution and clear wood density distribution were extracted from the X-ray image for comparison. The pixels of the X-ray image were divided into knot and clear wood subjectively and with a confidence gap; i.e., the sampled knot pixel values are high enough to be obviously recognized from clear wood pixels and the sampled clear wood pixels were far from the knots (Figure 5.5). The gap affects the distribution of the knot density, as the lower values are more likely to be cut

out. Figure 5.4-d and Figure 5.4-e show the two distributions. It is obvious that the average density of knot wood is considerably higher than the average density of clear wood.

The mean and standard deviation of the overall distribution are shown in equations (39) and (40):

$$\bar{x}_t = 1 \quad (39)$$

$$\sigma_t = 0.97 \quad (40)$$

The mean and standard deviation of actual knot wood distribution and clear wood distribution are as follows:

$$\bar{x}_c = 1 \quad (41)$$

$$\sigma_c = 0.05 \quad (42)$$

$$\bar{x}_k = 1.39 \quad (43)$$

The mean and standard deviation of the estimated clear wood distribution based on 20% of maximum frequency is as follows:

$$\bar{\hat{x}}_c = 1 \quad (44)$$

$$\hat{\sigma}_c = 0.04 \quad (45)$$

and the neighborhood of the mode is as follows:

$$[0.91 \ 1.09] \quad (46)$$

Equations (41), (42), (44), and (45) show that the mean and the standard deviation of the clear wood distribution are closely estimated.

One can conclude from Figure 5.4-b and -e that the threshold level should be 1.2. In this case the probability of misclassifying a clear wood pixel to a knot pixel is almost zero. Therefore, the cost of misclassification is as follows:

$$C_1 = C(1|2) q_2 \int_{R_1} p_2(x) dx \quad (47)$$

$$C_1 = C(1|2) q_2 \times 0.14 \quad (48)$$

The 99 percentile of estimated clear wood density distribution is 1.09. Therefore, the threshold level by using this criterion will be 1.1. The cost of misclassification is as follows:

$$C_2 = C(2|1) q_1 \int_{R_2} p_1(x) dx + C(1|2) q_2 \int_{R_1} p_2(x) dx \quad (49)$$

$$C_2 = C(2|1) q_1 0.032 + C(1|2) q_2 0.064 \quad (50)$$

If it is assumed that

$$C(1|2) = 10 C(2|1) \quad (51)$$

and

$$q_1 = 10 q_2, \quad (52)$$

the cost of misclassification is as follows:

$$C_2 = 3.264 C(1|2) q_2 \quad (53)$$

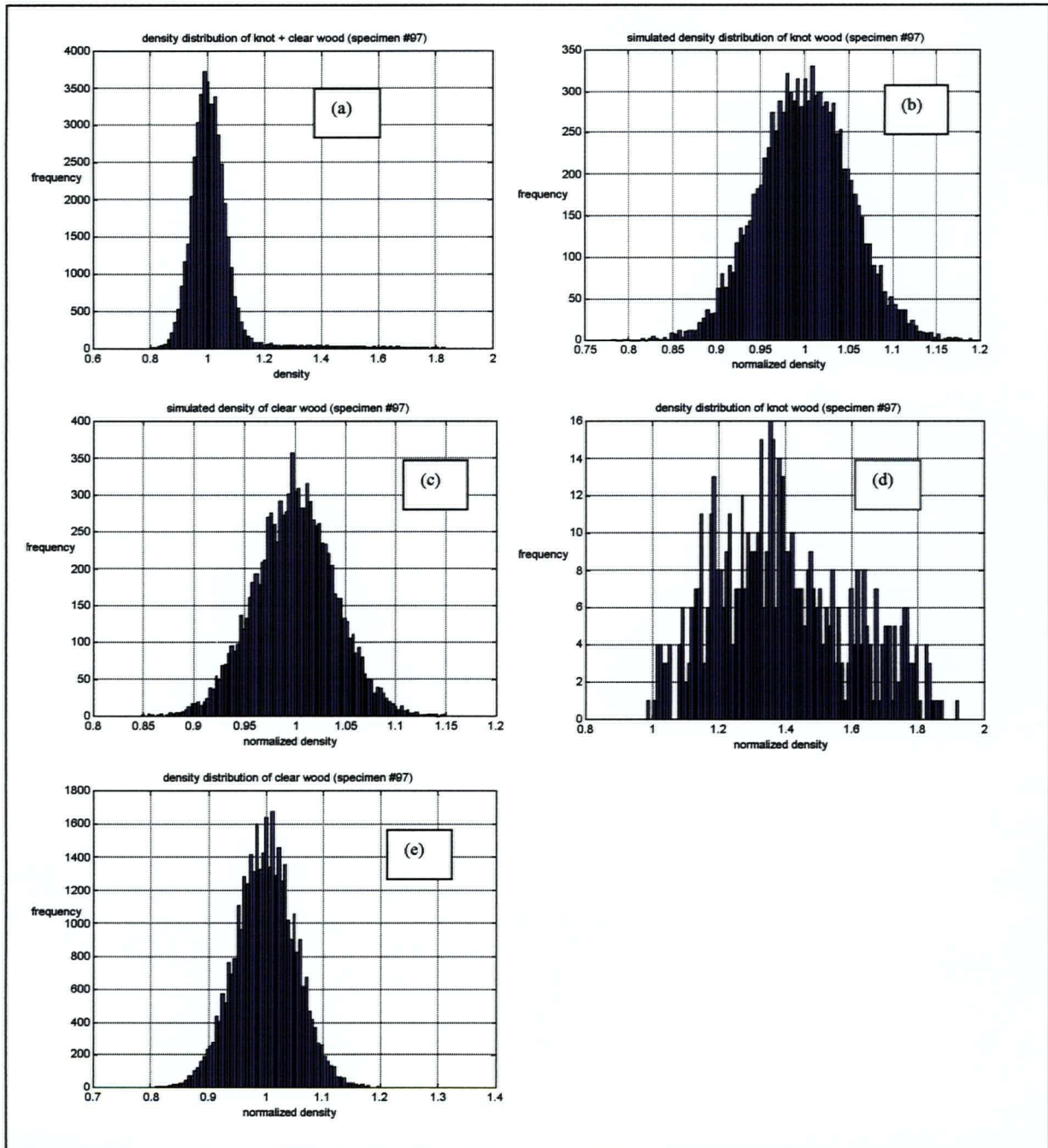


Figure 5.4. Density distribution of X-ray image pixels of a specimen (a) overall-distribution; the distribution of all X-ray image pixel values (b) simulated density distribution of clear wood (Normal distribution with similar mean and variance), (c) simulated density distribution (Normal distribution) of knot wood (d) density distribution of knot wood (normalized to average density of clear wood) (e) density distribution of clear wood (normalized to the average value).

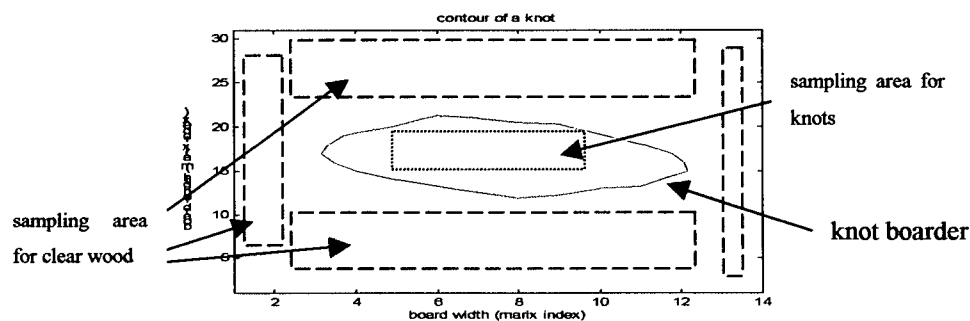


Figure 5.5. Sampling pixels of a knot from an X-ray image.

Figure 5.5 shows the samples pixels for the distributions of Figure 5.4. The area around each knot of the board is separated and the clear wood pixels and knot pixels are separated as shown in Figure 5.5. The threshold level that shows the border line between knot pixels and clear wood pixels is equal to the average of minimum and maximum wood density in the image of Figure 5.5.

### 5.3. Indexing the pixels of a knot

Once the threshold level is found and the knot pixels are detected, the neighboring pixels that are related to a knot are grouped. All the pixels that belong to the same knot are adjacent in the X-ray image. Once one pixel that is related to a knot is found all the neighboring pixels can be found and marked. Therefore, indexing the pixels of knots in the X-ray image is a double search. The first search is through all the pixels of the image looking for a knot pixel that was not indexed before (looking for a new knot). Once a pixel is detected it is indexed and a second search starts. This search indexes all the pixels that are neighbor to this pixel or are the neighbors of the neighbors of this pixel. Therefore, all the knot pixels that are related to the same knot are found and indexed in this search. Once all the pixels of one knot are indexed the first search resumes to find another knot.



A simple routine that can be used for indexing all the pixels of the same knot is as follows. The first found pixel (the result of first search) is placed in a stack. Then all the neighboring pixels are checked to see if they are knot pixels. All the neighboring knot pixels are placed in the same stack as well. Then all the pixels in the stack can be retrieved and their neighboring pixels are checked to see if they are knot pixels. A neighboring knot pixel is added to the stack if it already doesn't exist in the stack. The search terminates when all the pixels in the stack are checked and no new pixel is added to the stack.

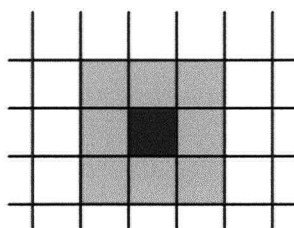


Figure 5.6. A search for neighboring knot pixels. Once the black pixel is known to be a knot pixel, all the gray pixels are checked too.

#### 5.4. Knot pattern

The knot pattern is the edge of the image of a knot. The edge pixels are the outermost pixels of a group of pixels that are indexed to be related to a knot. The inside pixels can be easily detected because all their neighboring pixels are knot pixels. The edge pixels are simply non-inside pixels. Figure 5.7 shows an edge pattern that is obtained by using the method presented above. This pattern is simply the edge of the density profile shown in Figure 5.1.

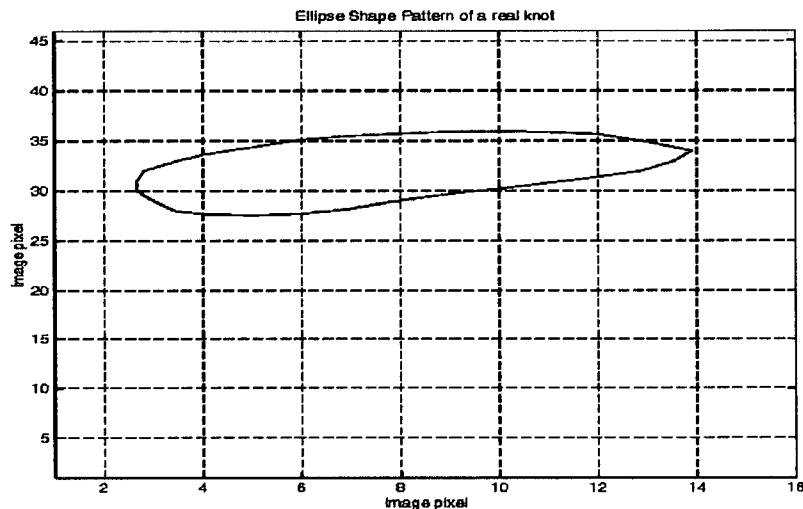


Figure 5.7. A knot pattern (from its image)

Solving this pattern for knot parameters is a basic problem in our knot detection procedure. The knot pattern can be imagined to be the orthographic projection of a cone onto an image plane ([69]). The cone is cut by two parallel-planes. Figure 5.8 shows the cone and its images on two image planes. Deriving the parameters of the cone from the features of the pattern in x-y plane will result in solving the knot detection problem. It will be shown that the pattern in x-z plane can easily be solved.

The patterns that are shown in Figure 5.8 are equivalent to knot patterns in an ideally simple case. In this case, the clear wood does not have any variations (standard deviation equal to zero) therefore the threshold level can be set very close to the average clear wood density. Also, the knot density is distinctly different than the clear wood density at all points of the knot.

Because the knot pattern results from the projection of all the points of the cut cone onto the image plane, different sections of it can be distinguished. This pattern consists of two half ellipses whose major axes are along the same line; therefore, their minor axes are parallel. The two ellipses are connected by two straight-lines as is shown in Figure 5.8. The center part of the pattern is a trapezoid whose two bases are equal to the diameters of the cone at the two ends.

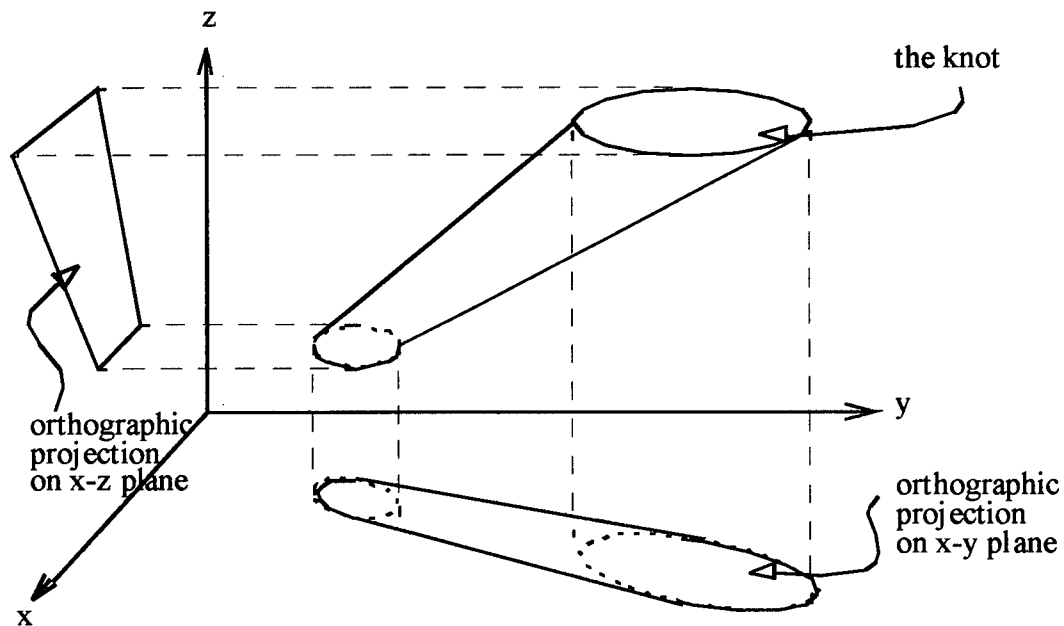


Figure 5.8. Two orthographic projections of a knot. Projection line is perpendicular to the image plane (x-y and x-z planes).

The special shape of this image can be proven to be correct by circulating the generator line around the axis of the cone and following its projection on the image plane. It can be proven that the symmetry axis of the pattern is the projection of the axis of the cone. Also, the centroid of the image lies on symmetry axis of the pattern. It will be shown that in most cases one of the extremities of the symmetry axis of the pattern is the farthest point from the centroid of the contour. Therefore, the symmetry line can be drawn from the centroid and the farthest point; that is, the point with maximum distance from the centroid, that is on the perimeter of the pattern.

The axis of a knot cone is assumed to be closely radial; therefore, it wouldn't have a large component along the tree's length. It means that the edges of the board will not cut the pattern parallel to the symmetry axis (i.e., its major direction). This axis will be parallel to the large axis of the board.

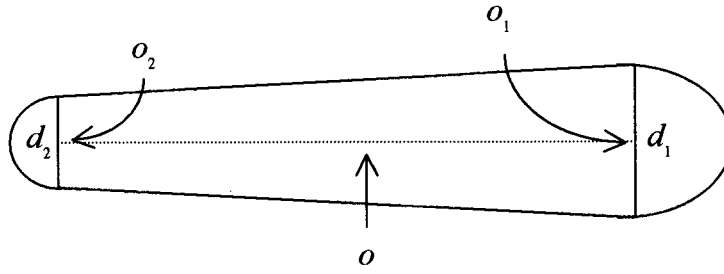


Figure 5.9. The pattern of a knot.  $o$  is the centroid of the pattern. The dashed line indicates the symmetry line of the pattern.

#### 5.4.1. Knot model parameters

A knot is modeled by a set of parameters that are easy to derive and useful in practice. The two centers of the two ellipses of the knot and the radii of the cone at the two ends can define the knot cone. The two centers define the axis of the cone and the two radii define the acute angle of the cone and size of the knot. Therefore, eight parameters can define the knot.

The centers of the two knot ellipses are projected to the centers of the two half ellipses of the knot pattern ( $o_1$  and  $o_2$  in Figure 5.9). The two diameters of the cone are projected onto the two minor axes of the corresponding half ellipses ( $d_1$  and  $d_2$ ). Once the two patterns in two perpendicular image planes are obtained the knot parameters can easily be solved.

#### 5.4.2. Measuring and calculating the features of the pattern

A simple solution to finding  $d_1$  and  $d_2$  assumes that the knot pattern is a trapezoid. What is needed to find  $d_1$  and  $d_2$  are the two areas of half trapezoid and the height of the trapezoid from the center to the two ends. If the area of the smaller end is  $a_1$  and the area of the bigger end is  $a_2$ ,  $d_1$  and  $d_2$  (see Figure 5.10) are found as follows:

$$d_1 = \frac{2a_1}{\lambda_1} - d \quad (54)$$

$$d_2 = \frac{2a_2}{\lambda_2} - d \quad (55)$$

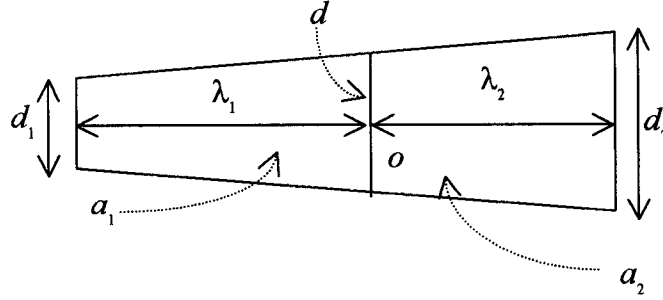


Figure 5.10. The approximate image of the knot.  $o$  is the centroid of the pattern.

In order to measure  $d_1$ ,  $d_2$ ,  $\lambda_1$ , and  $\lambda_2$ , the centroid of the pattern,  $o$ , is found first (see Figure 5.11). The farthest point of the edge of the pattern from  $o$  is the tip of the smaller ellipse.  $p_2$  is the point that the angle of  $(p_1, o, p_2)$  is closest to 180 degrees.  $p_3$  is the point with  $\angle(p_1, o, p_3) = 90^\circ$  and  $p_4$  is the point that  $\angle(p_2, o, p_3) = 180^\circ$ . The pattern is then divided into two halves on the two sides of  $(p_3, p_4)$  line. Let's say  $p$  is a random point in the pattern, if  $\vec{a} = (o, p) \perp (p_3, p_4)$  is the perpendicular projection of  $\vec{a}$  on line  $(p_3, p_4)$ , it defines which side of the pattern the point  $p$  lies. The angle of  $\vec{a}$  and  $(o, p_1)$  ( $\angle((o, p_1), \vec{a})$ ) indicates the side. If this angle is zero  $p$  belongs in one side ( $p$  is closer to  $p_1$ ) and if it is  $180^\circ$ ,  $p$  belongs to the other side ( $p$  is closer to  $p_2$ ). Figure 5.11 shows the four points and the center of the pattern.

Adding the area of all the pixels of one side gives the area of one side of the pattern.  $\lambda_1$  is the distance between  $o$  and  $p_1$ ,  $\lambda_2$  is the distance between  $o$  and  $p_2$ , and  $d$  is the distance between  $p_3$  and  $p_4$ . From equations (54) and (55) the two diameters of the two ellipses are obtained. The projections of the two centers of the cone on an image are estimated by shifting  $p_1$  and  $p_2$  into the connecting line by  $d_1 / 2$  and  $d_2 / 2$ .

When the pattern ends to the edge of the board it becomes a trapezoid (Figure 5.8, the pattern in x-z plane). In such a case the detection problem becomes a straightforward

problem. In such a case the sections of the pattern that touch the border of the board are measured. The length of each section is  $d_1$  and  $d_2$ , respectively.

The diameter of the knot cone at the two ends equals  $d_1$  and  $d_2$  that are readily measured from either x-y or x-z pattern. The centers of the two ellipses of the knot are obtained from the two projections of each center that are obtained from each pattern. Also the centroid of the knot is estimated by using the two centroids of the two patterns.

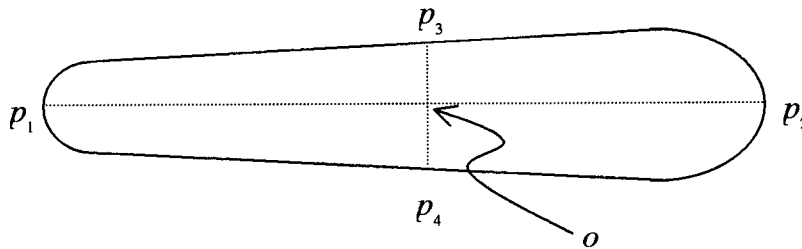


Figure 5.11. The four points of the pattern

In the following the block diagram of knot pattern analysis is presented. In this algorithm, it is assumed that all the elements of the pattern are previously detected and saved in an  $n \times 2$  matrix where  $n$  (a variable) is the number of pixels in the knot pattern.

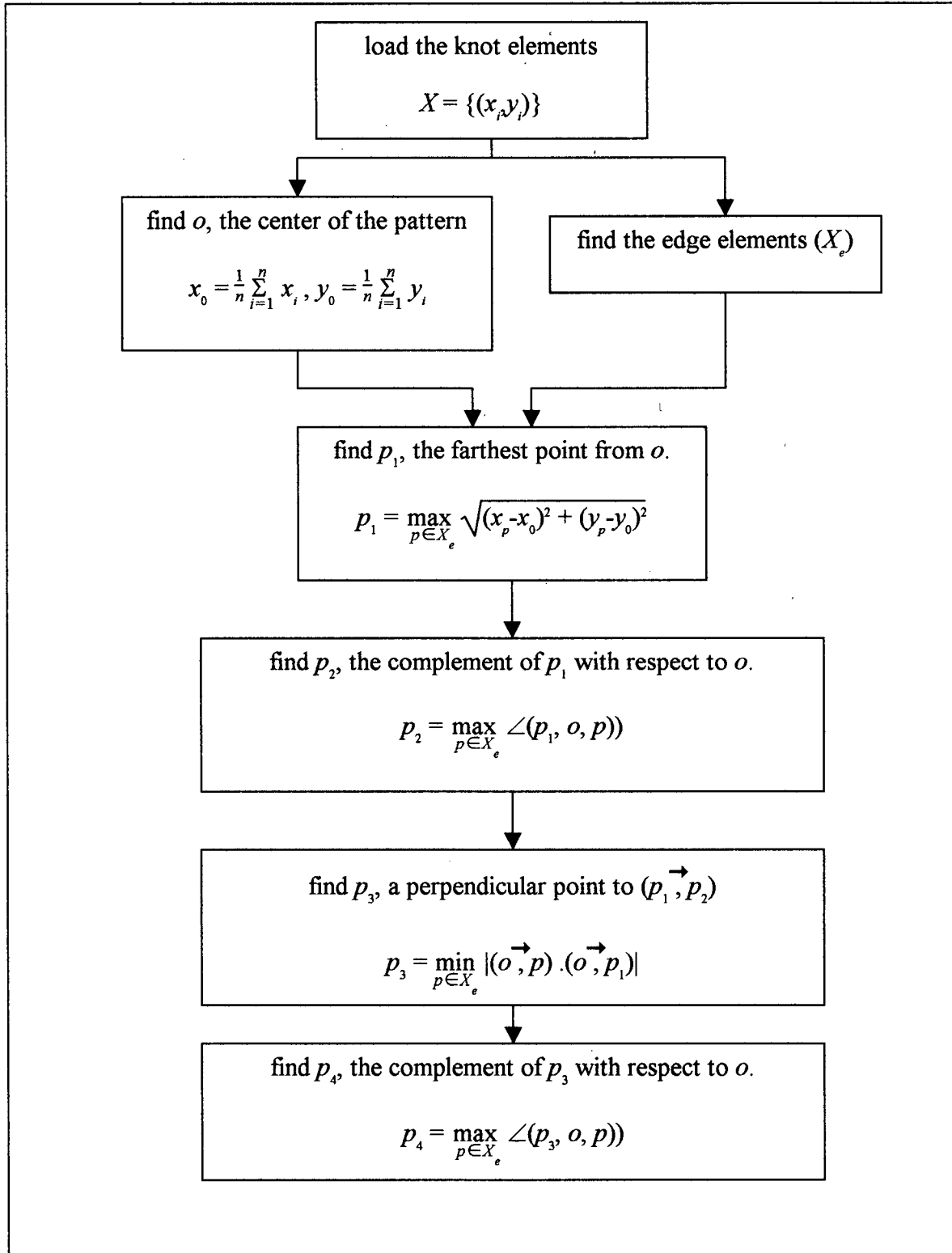


Figure 5.12. The flow chart for knot pattern analysis.

### 5.4.3. Matching patterns

In order to match the two patterns in either of the images the longitudinal location of the centroids and the span of the patterns (the length of the pattern along the board) are used, as is explained in the following. The two patterns are matching if the two spans are overlapping. If more than one pattern overlaps with a pattern, the ones with closest centroids are chosen. The two sides of the two patterns are also matched by finding the smaller end and the bigger end. The smaller diameters are averaged to result in the smaller diameter of the knot. The projected centers of the ellipse on these images result in the center of the smaller ellipse of the knot. Similarly the parameters of the larger ellipse of the knot can be obtained.

### 5.5. Geometrical features of a knot

The geometrical features of the knot are indicators of its size, taper, and deviation from center. The volume of a knot,  $d_1 / d_2$ , axial deviation, and knot area ratio are the geometrical features of a knot.

In order to measure the volume and the axial deviation of a knot, a model of the knot is generated (by using the measured parameters) in a cubic matrix. By using the center of the two ellipses the axis line of the cone can be generated. The cone is generated as a parametric equation. For every point of the matrix cube its projection on the axis of the knot is obtained. The radius of the cone at this point is obtained by linearly interpolating the two radii of the cone. If the distance of the cube pixel from the axis is more than the related radius of the cone the pixel is marked as a knot pixel.

Adding the volume of all the knot pixels of the cube results in the volume of the knot. Knot area ratio is equal the ratio of knot area to board area at every cross-section of the board.

The axial deviation is the standard deviation of all the pixels of the knot when compared with the central line of the board. The axial deviation of the knot is shown in the following:



$$\text{Axial deviation} = \frac{1}{N} \sqrt{\sum_{n=1}^N (d(p_n, p_c))^2} \quad (56)$$

where  $N$  is the number of knot pixels,  $d(p_n, p_c)$  is the distance between two points,  $p_n$  is the  $n$ th knot pixel and  $p_c$  is the projection of  $p_n$  on the central line of the cube.

## 5.6. X-ray image analysis; extraction of board image and defect detection

As was explained in previous chapters, detecting knots from the X-ray image is the cornerstone of geometrical features for characterizing a specimen. In this section the extraction of a board X-ray image from the measured X-ray image and the extraction of knots from the board X-ray image will be presented. In Section 5.2 the image segmentation problem was generally related to the density distribution of clear wood and the required equations for extracting useful information from an X-ray image were developed. In this section those equations are applied to a typical specimen.

The measured X-ray image can be divided into three parts: the surrounding area, the clear-wood area, and the knot area. Each area of the image is detected and separated from the rest by thresholding the density of the image pixel. Separating the board X-ray image from the rest of the image is simply done by a fixed thresholding because the noise level is not comparable with the signal level.

The overall statistical distribution of the density of wood is derived from the X-ray image of the sample. The actual density distributions of clear wood and knot wood were extracted subjectively for comparison. The statistical mean and standard deviation of the overall distribution will be compared with the statistical mean and standard deviation of clear wood distribution. The mean of the overall distribution will be used as the estimation of the mean of clear wood distribution. The 20% maximum frequency neighborhood will be used to estimate the standard deviation of clear wood density distribution. The threshold level for knot edge detection will be derived based on the actual and estimated clear wood density distributions. The cost of misclassification will be derived for each case.

### 5.6.1. Edges of the board

Extracting the image of the board from the X-ray image was overlooked to this point. A threshold level can extract this segment of the image as well. First the noise level in the sample image is measured. Figure 5.13 shows the comparison of the output levels. The noise level is related to the same run that the image of the board is extracted. For that purpose the average of the sensor output before the board reaches the location of the sensor is used for representing the noise level. This comparison shows that the signal to noise ratio is at least 100. Also the corners of the board cross-section profile shows that the image does not have large shadow area. The shadow image appears as a slow slope of the line descending from the clear wood density to the surrounding area.

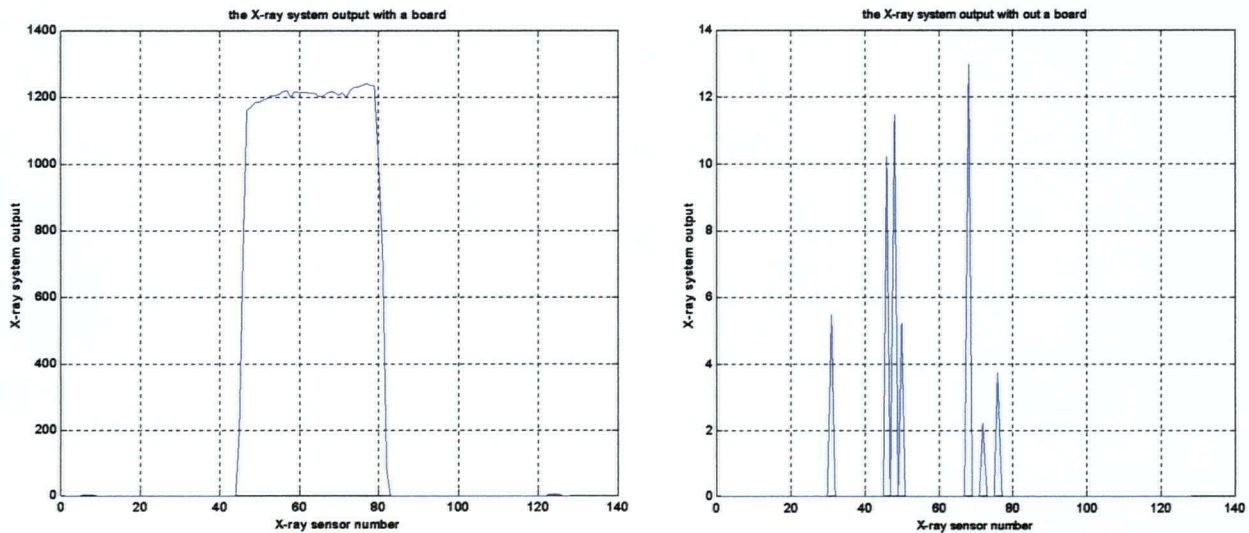


Figure 5.13. The output of the X-ray system with (top) and without (bottom) the sample board. The maximum sensor noise level is about 0.01 of the average signal level.

Therefore, a threshold scale of 0.6 was used to separate the board image from the surrounding area. This level is well below the wood density level and well beyond the noise level (see Figure 5.4-a).

For segmentation of the X-ray image it is assumed that the board is located in a straight extension along the image. Therefore, the average values of pixels along the image and

across the image are found. The search starts from both the beginning and the end of the image. When the average pixel value is more than the threshold level, the location is marked as the edge of the board.

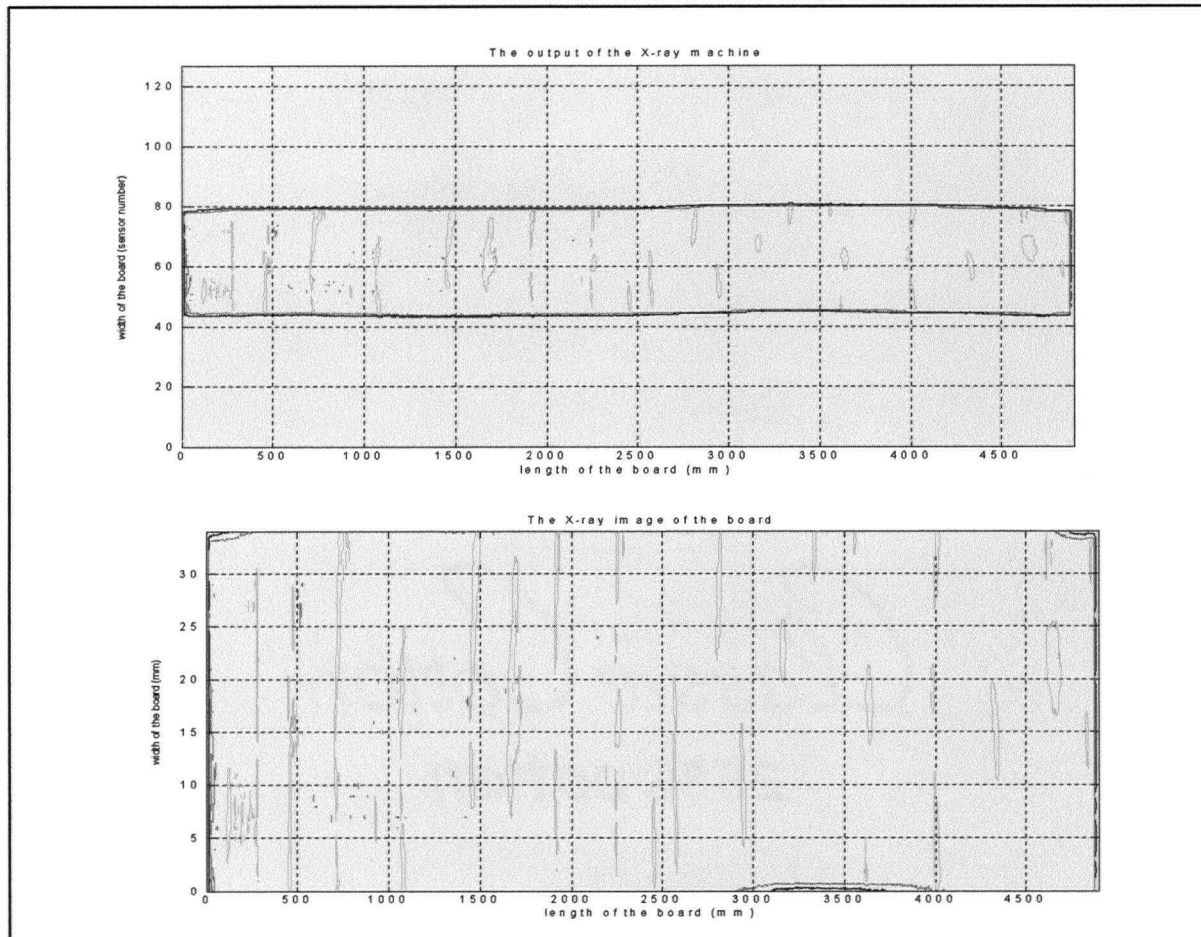


Figure 5.14. Output of the X-ray imaging machine (top) and the X-ray image of the board (bottom).

The coordinate of the image of the board in Figure 5.14 is used for knot detection. The same method is used for processing the edgewise image of the board.

### 5.6.2. Knot detection

One can conclude from Equation (32) and Figure 5.4-b that the threshold level should be 1.2. In this case the probability of misclassifying a clear wood pixel to a knot pixel is almost zero. If this level of thresholding is used the edges of knots of the board would be as shown in Figure 5.15.

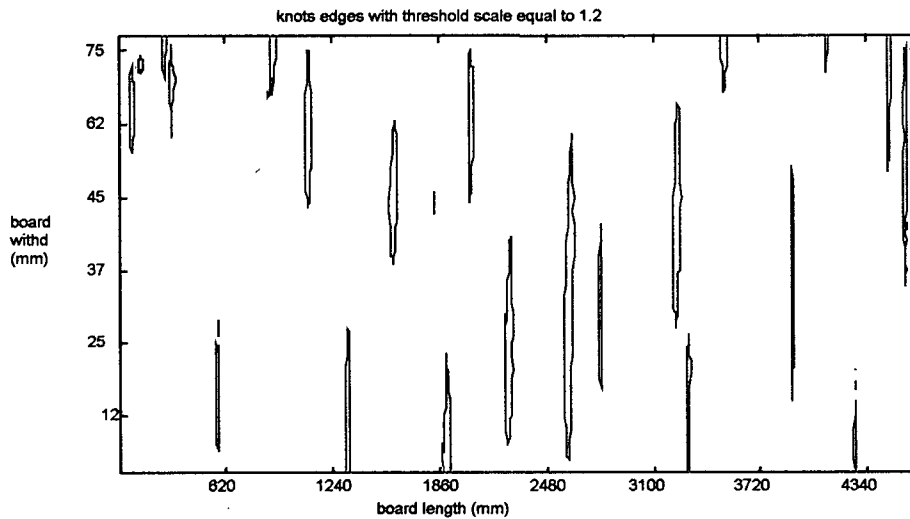


Figure 5.15. The edge of all knots of a board detected by fixed Bayesian based thresholding.  
The threshold level is 1.2 times the average clear wood density.

The edges of all knots of a board are detected by fixed level Bayesian based thresholding. The threshold level is 1.1 times the average clear wood density. The edges of knots of the board that are detected by using this thresholding are shown in Figure 5.16.

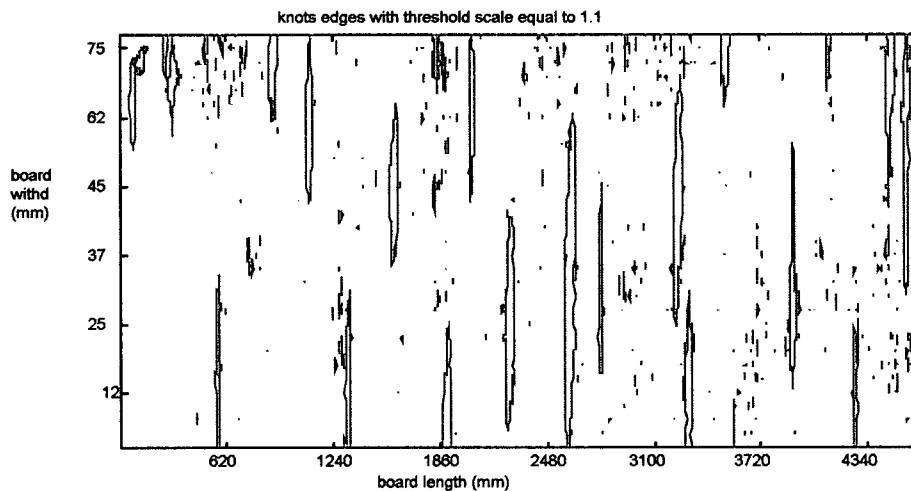


Figure 5.16. The edge of all knots of a board detected by fixed Bayesian based thresholding.  
The threshold level is 1.1 times the average clear wood density.

Other randomly selected specimens are shown in Figure 5.17 and Figure 5.18.

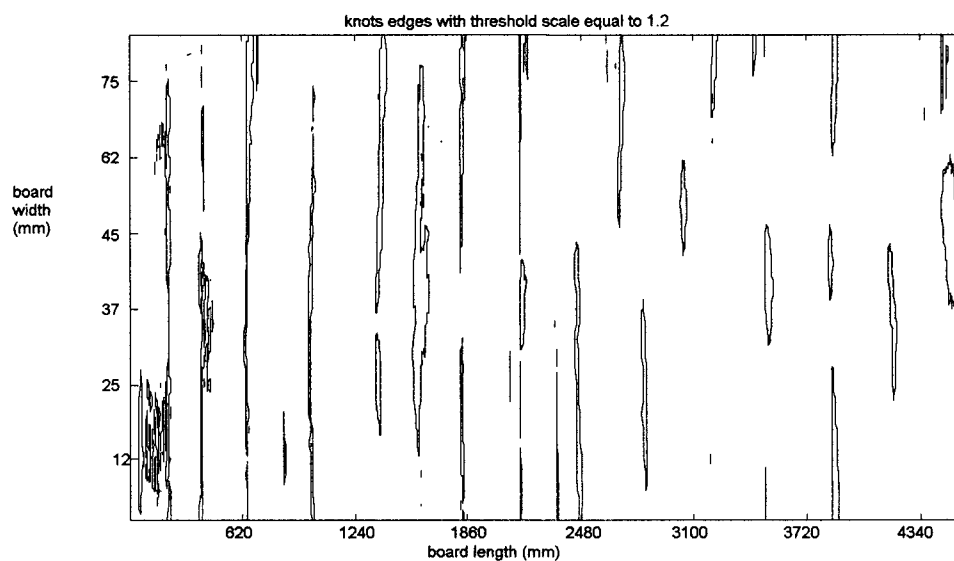


Figure 5.17. Knot detection by using Bayesian thresholding. Specimen 130.

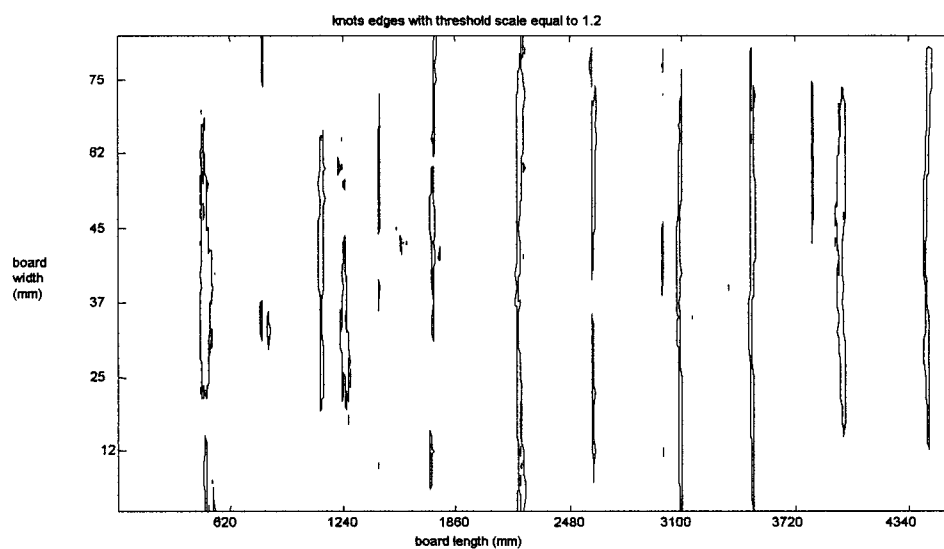


Figure 5.18. Knot detection by using Bayesian thresholding. Specimen 70.

The following figures show the effect of thresholding to the image of a knot.

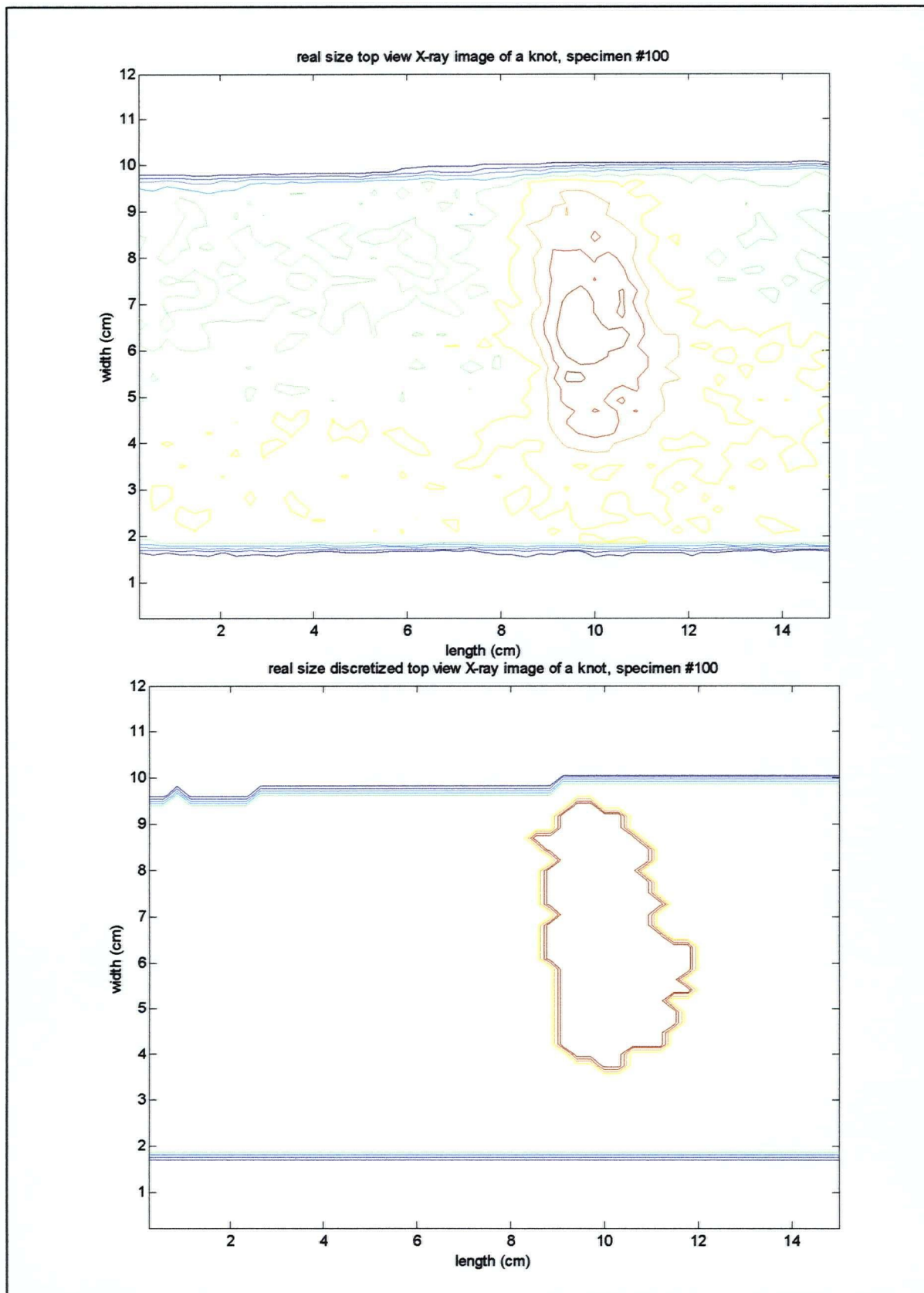


Figure 5.19. The top view X-ray image and discretized X-ray image of a knot.

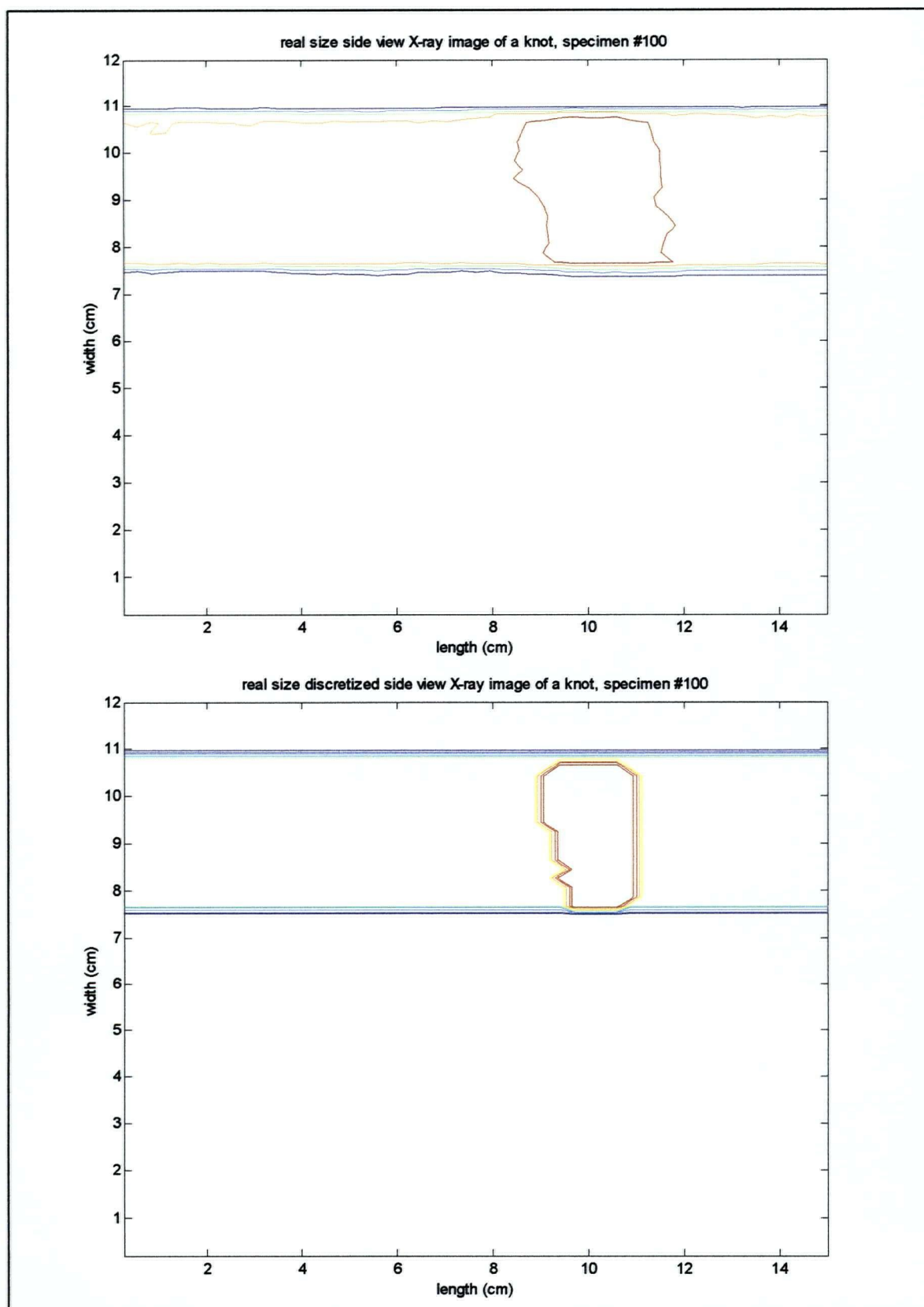


Figure 5.20. The side view X-ray image and discretized X-ray image of a knot.



### **5.6.3. Three dimensional knot shape detection (simulation)**

In this section, the projection pattern is solved and all its parameters are detected. A projection pattern is a pattern that is generated in an x-ray image of the knot. A knot in a perfect board and its x-ray image(s) are artificially generated. Simulation was used for comparing the generated parameters with detected parameters and for analyzing the accuracy of the detection method. The error of the knot detection algorithm will be reflected on strength estimation and, therefore, it is reasonable to expect that any improvement in the knot detection algorithm will improve the strength estimation accuracy.

A knot can be modeled by using a higher density cone (than clear wood) inside a cube. The cone of the knot is defined by seven parameters; the center of the knot (three parameters), the direction vector (three parameters), and the acute angle of the knot (one parameter).

The X-ray image of the board is basically the orthographic projection of the object (i.e., the board) on the image plane. In the following it is shown that one can start from the X-ray image and obtain all the parameters of the knot. The board itself is a 3-D matrix. Each element of the matrix that is inside the knot (can be defined by using vector algebra) has the value of one while each element that is outside the knot has the value of zero. The image (a 2-D matrix) is obtained by adding the pixel values along a specific direction.

Since the image is simulated, its edges represent the most accurate pattern that can be obtained by measurement. In fact the actual measured images are not this accurate because they suffer from general blurring and lack of uniform material density distribution in the knot and in the clear wood area. The fact of having perfect image in this case allows the threshold level to be chosen close to the clear wood density and obtain a pattern that is almost as accurate as a theoretical pattern.

The detection procedure starts by detecting the existence of a knot due to the existence of some points in the image with higher density than the median value. Then the clear wood density is used for finding one point in the knot area. The entire pixels that are in the knot are searched for finding the edge points and the centroid of the image. The image is



presented by its edge so that the inner density of the knot does not affect the procedure (this makes the procedure more general).

The farthest point from the image centroid and the centroid itself are located on the symmetry line of the pattern. There is a complementary point to the farthest point that lies on the symmetry line and the edge of the pattern. The distance of these two points is the length of the pattern. The width of the pattern is the longest line segment that is drawn between two points of the edge and is perpendicular to the symmetry line. In this simulation the length of this line is assumed to be close to the length of the line segment that passes the centroid and is perpendicular to the symmetry line. These two line segments are called the symmetry line and the perpendicular line in this report. Having the centroid and the farthest point from it, one can find the three other points by using vector algebra.

The areas of the pattern that are at either side of the perpendicular line are given different names. The area of inner side of the pattern from the perpendicular line towards the farthest point is called the vertex-side area and the other area is called the non-vertex-side area. Each of them is found by scanning the entire pixels of the pattern and finding what side of the line the pixel is located. Then the area of the pixel is added to the related area. The vertex-side area is used for finding the smaller diameter of the knot cone.

The simulation parameters are as follows.

$(x_0, y_0, z_0)$  = vertex location

$(v_1, v_2, v_3)$  = the direction vector of the knot

$\varphi$  = cone acute angle

$\psi = \arctan\left(\frac{\sqrt{v_1^2 + v_2^2}}{v_3}\right)$  is the angle of the knot.

$L = \text{Max}\{(38-z_0)[\tan(\psi+\varphi) - \tan(\psi-\varphi)], (38-z_0)\tan(\psi+\varphi) + z_0\tan(\psi-\varphi)\}$  is the length of the pattern. In this equation 38 is the thickness of the board.

$l = 2 \times ((38-z_0) / \cos(\psi)) \cdot \tan(\varphi)$  is the larger diameter of the knot

$l' = 2 \times (z_0 / \cos(\psi)) \cdot \tan(\phi)$  is the smaller diameter of the knot.

The detected parameters are as follows.

$L$  is the length of the pattern.

$l$  is the width of the pattern, which is equal to the larger diameter of the pattern.

$l'$  is the smaller diameter of the pattern.

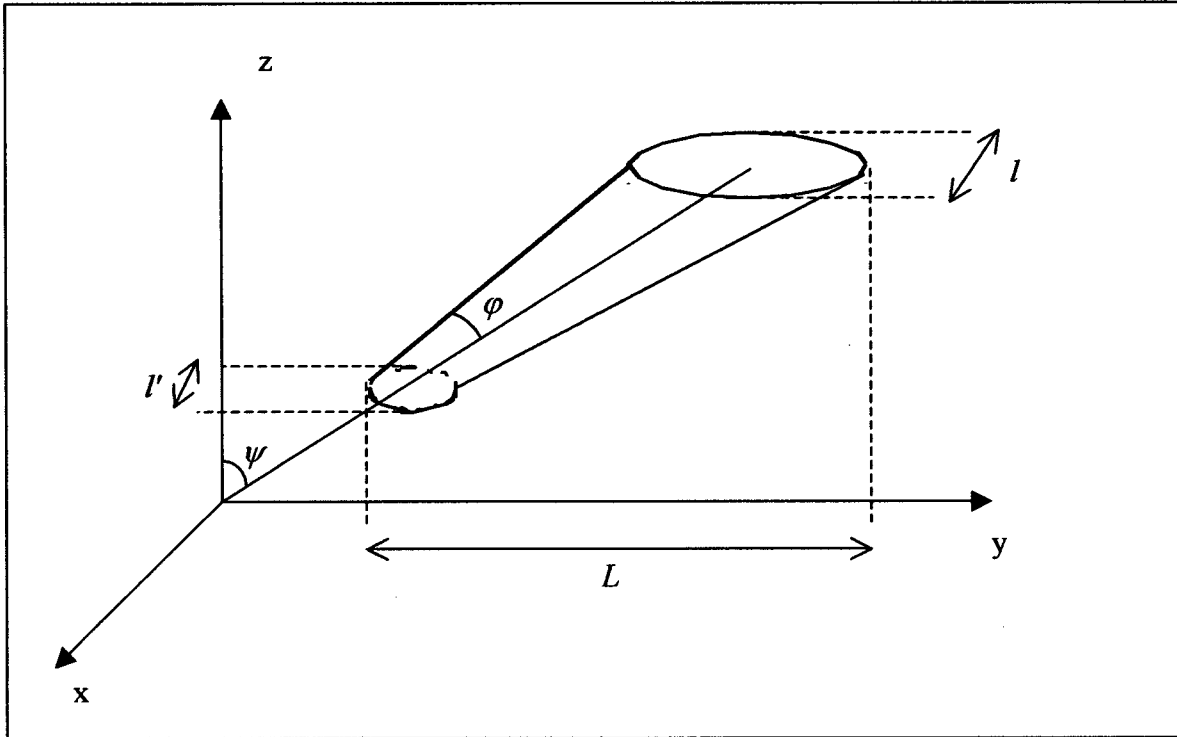


Figure 5.21. The diagram of a knot cone and its parameters.

$\psi$  is the solution of  $L \cdot \cos(\psi) - 38 \sin(\psi) = \frac{1}{2} \cdot (l + l')$ . Since  $\psi$  is in  $[0, \frac{\pi}{2}]$ , and since  $\cos(\psi)$  and  $\sin(\psi)$  are always decreasing in  $[0, \frac{\pi}{2}]$ , the above equation has a unique solution. The location of the knot vertex is kept fixed at  $(x_0 = 0, y_0 = 0)$  and  $-20 \leq z_0 \leq -35$  in increments of 2. Two relatively small cone acute angle values are chosen to be  $\phi = 3, 5, 7$  degrees. Three values for the knot angle are  $\psi = 25, 28, 31$  degrees.  $v_1$  is equal to  $v_2$  so that by choosing  $\psi$  the three direction vectors can be determined. For the given knot angles ( $\psi$ ) and knot vertex location  $(0, 0, z_0)$ , the smaller

diameter of the knot would be less than 2.5 cm and the larger diameter would be less than 5 cm. Therefore, the size of the knot is relatively small and the accuracy of the model can be assessed by this simulation.

A three dimensional matrix simulated a small board segment ( $38 \times 89 \times 100$  mm). Every element of the matrix represents a point in the artificial board. Each dimension of the matrix is one dimension of the board ( $x$ ,  $y$ , or  $z$ ). The matrix elements that are inside the knot are equal to one and the rest of the elements are equal to zero. These element values are selected for simplicity, such that the edge of the knot can be easily detected in a simulated X-ray image as follows. The simulated X-ray image is a two dimensional matrix whose element values are obtained by adding the elements of the simulated board in one direction. The  $x$ - $y$  simulated image (equivalent to the top X-ray image) is obtained by adding up the elements of the simulated board along the  $z$  axis. Similarly for obtaining an element value of  $x$ - $z$  simulated image, all the elements of the simulated board along the  $y$  axis are added. All the nonzero elements belong in the knot image.

Two X-ray images of a simulated knot are shown in Figure 5.22. The simulation parameters are  $\psi = 35$ ,  $\phi = 5$ , and  $z_0 = -30$ . The resolution of the image is 0.5 mm.

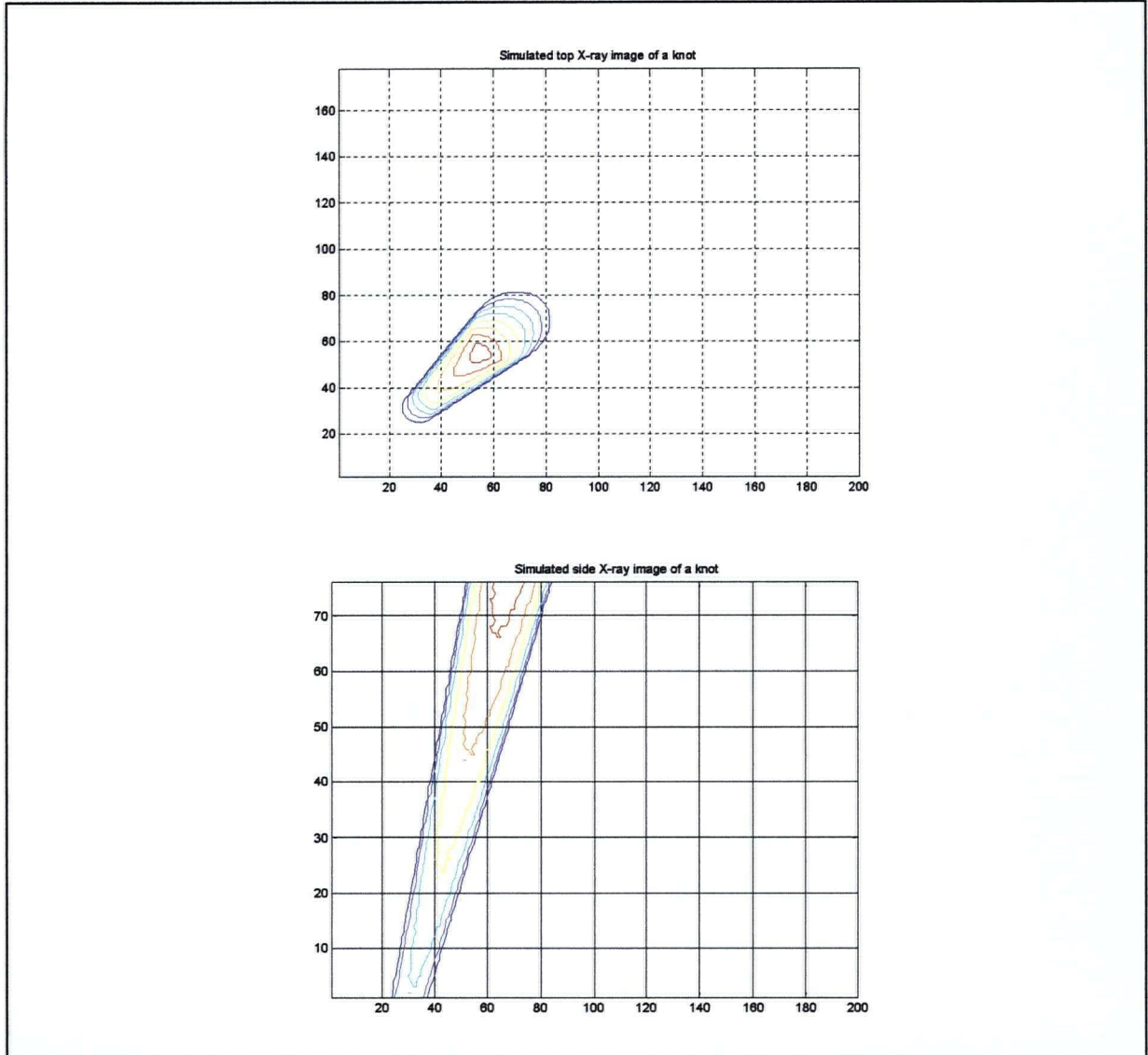


Figure 5.22. Contour diagrams showing the simulated top X-ray image (top) and side X-ray image (bottom) of a knot. The simulation parameters are  $\psi = 35$ ,  $\varphi = 5$ , and  $z_0 = -30$ .

Image resolution is 0.5 mm.

Figure 5.23 shows the error of detection (for the image resolution) for larger and smaller diameters of the simulated knot. The coefficient of determination ( $r^2$ ) between the detected smaller diameter of the knot and the generated smaller diameter is 0.84, and the  $r^2$  between the detected larger diameter of the knot and the generated larger diameter is 0.98. The error of the larger diameter in some cases is relatively big.

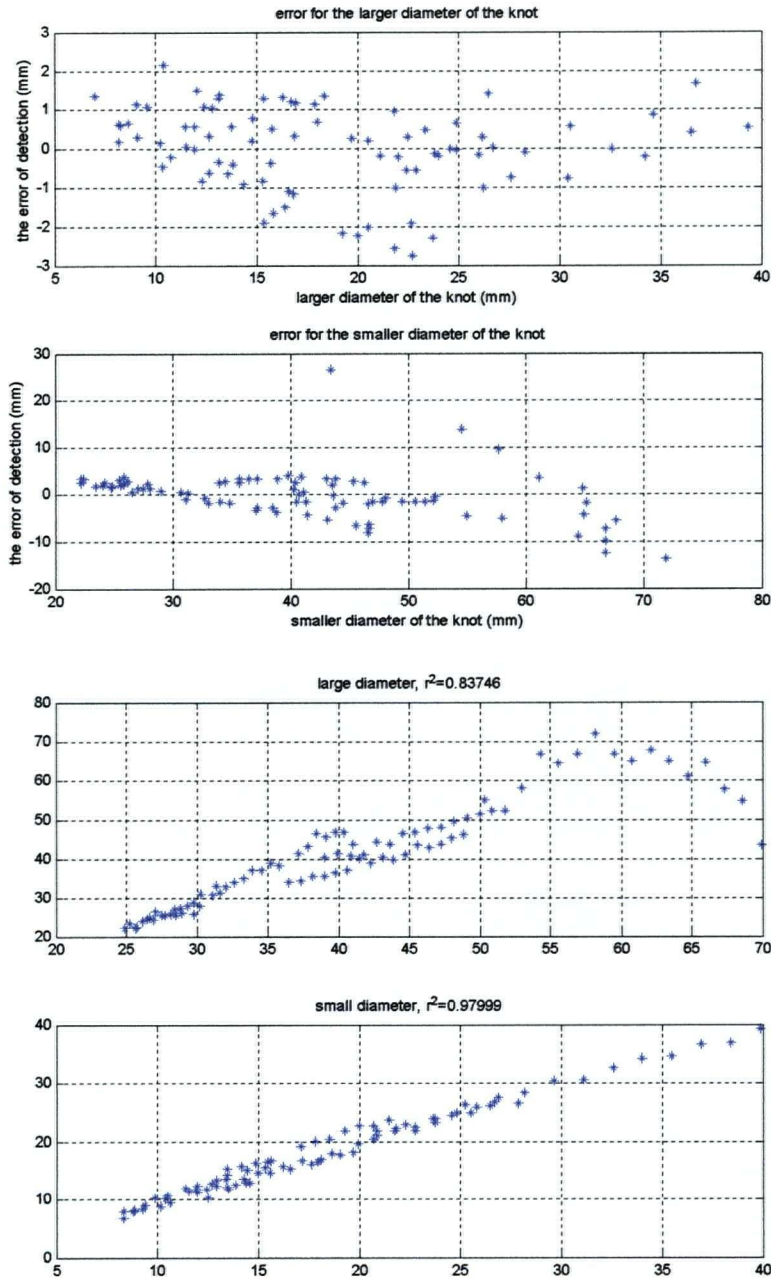


Figure 5.23. Estimation accuracy for the knot diameters. The horizontal axis represents the generated diameter and the vertical axis shows the detected diameter.

The set of knots that are simulated contains the knot size/locations that can stretch outside the limit of the board and cause excessive error in the detection. The following Figure shows a simulated knot that stretches outside the limit of the board and produces error in detection.

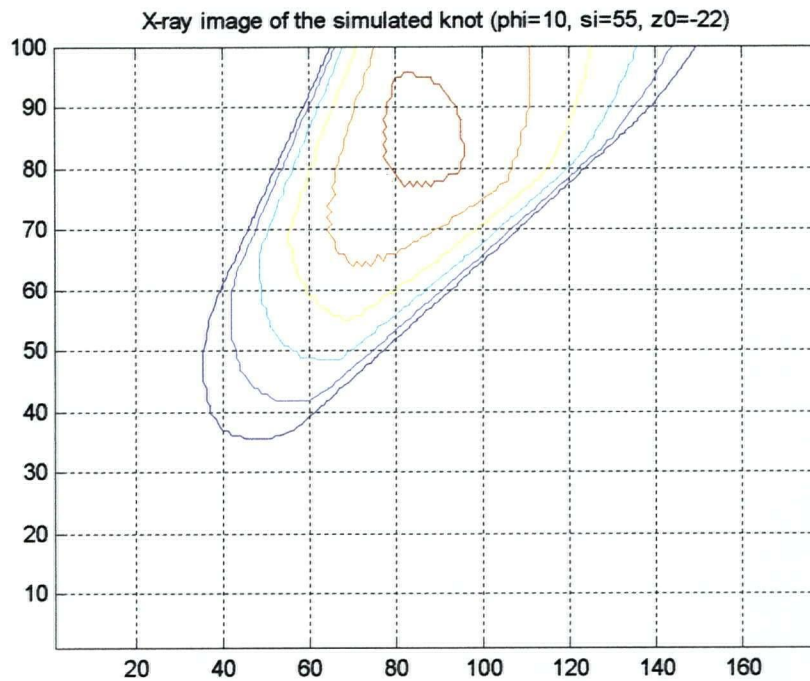


Figure 5.24. The simulated knot pattern that is partially stretching outside the board.

### 5.7. Knot detection for real data

The knot detection algorithm was applied to a few samples for comparison of the detection of knot size and location with image data. The results are shown in the following figure:



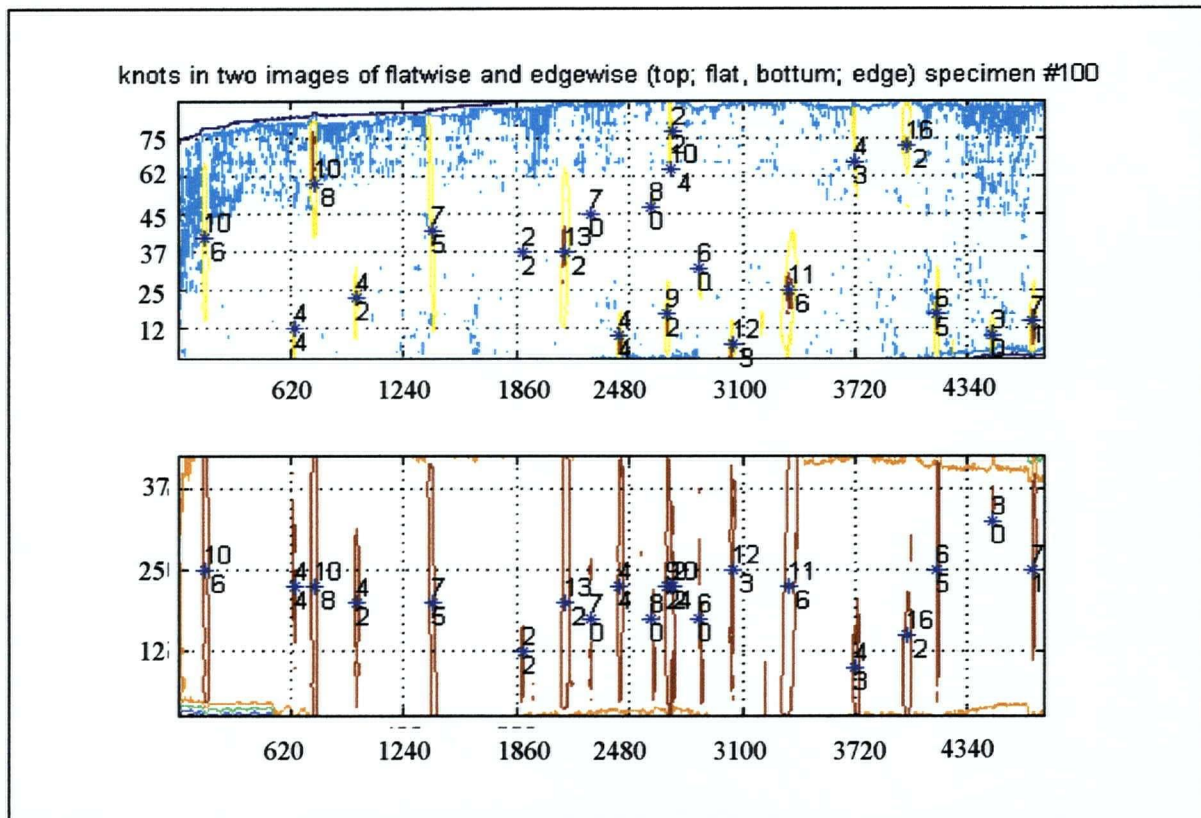


Figure 5.25. The result of applying the knot detection algorithm to the X-ray image of specimen #100.

The location of each knot is detected and marked by an asterisk. The two knot-cone diameters are also printed at the location of the knot. For example, the two numbers (10,6) show that the diameter of the knot at one of the faces is equal to 10 and it is equal to 6 at the other face of the board.

The board can be regenerated from the knot parameters and an artificial X-ray image can be produced. The result is shown in Figure 5.24.

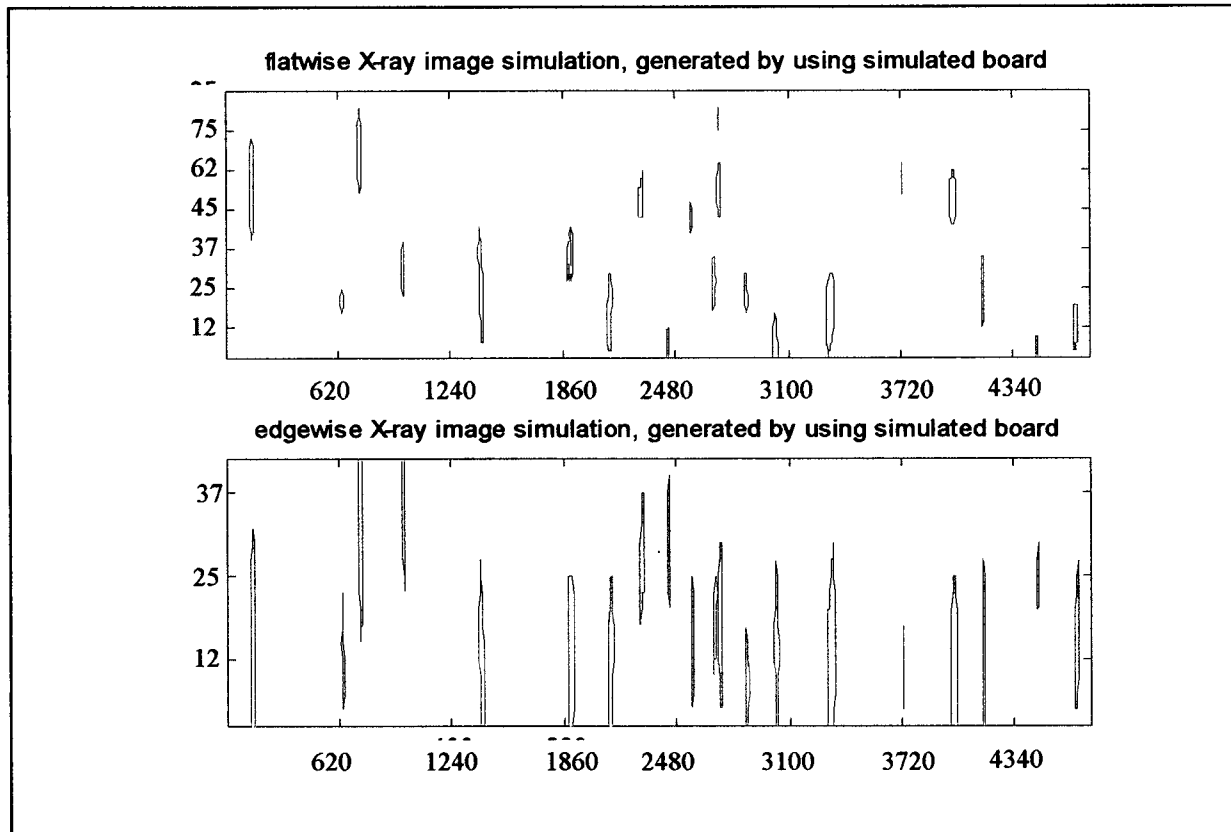


Figure 5.26. The simulated X-ray images. These images are generated from the geometrical features of specimen #100.

Also location of knots can be detected in CCD camera images of a board. About 80% of the knot patterns in each simulated image (Figure 5.26) overlap with those in the discretized form of the X-ray image (Figure 5.25). In order to calculate the overlap of the two images, the X-ray image is discretized the same way as is described in knot detection algorithm and the output image is a binary with pixel values of zero or one. The pixels that represent a high-density defect are marked as one and the rest of the pixels are marked as zero. The simulated X-ray images are similarly processed. Therefore, the overlap of the two images is equal to  $\sum_{i=1}^M \sum_{j=1}^N x(i,j)x_s(i,j)$ .  $M$  and  $N$  are the number of rows and columns of each matrix. Dividing the overlap of the two images by the sum of defect area in the simulated image ( $\sum_{i=1}^M \sum_{j=1}^N x_s(i,j)$ ) gives the ratio of the overlap.

As a byproduct of knot detection algorithm, the location and size of knots can be indicated in other measured profiles such as CCD camera image (Figure 5.25).



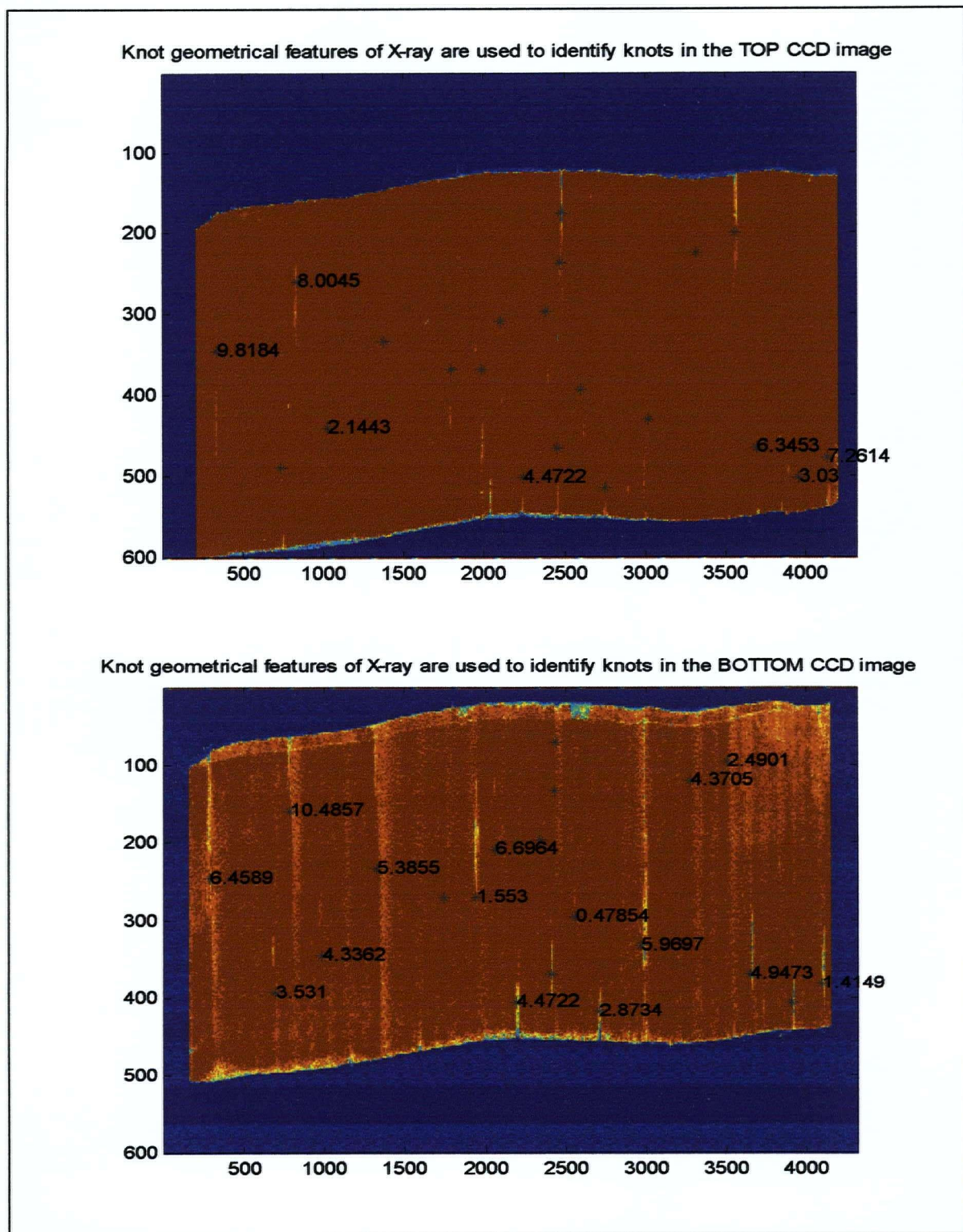


Figure 5.27. The location and size of knots in CCD camera images are detected from geometrical features.

## 5.8. X-ray and SOG interrelation

Similar to density variations in a board the grain angle variation can be divided into two factors, the variations due to the existence of knots and other variations in the regular grain angle of wood.

Figure 5.28 shows the variations in the grain angle profile that is caused by the existence of knots. The type of variation in the grain profile (as can be seen in Figure 5.28) depends on the relative path of the (grain angle) measurement head with respect to the knot. Therefore, the shape of the grain angle variations in Figure 5.28 varies. In Figure 5.29 all the profiles of a sample board as well as the board image are shown.

It is interesting to note that the variations due to the existence of knots (Figure 5.28) can be modeled using the geometry of knots. An approximate model proposed in [70] is based on fluid flow around a solid obstacle in its path. The average grain angle in this model replaces the direction of the fluid stream. The direction of the fluid at every point around the obstacle defines the direction of the grain angle around the knot ([70]). Assuming that the general trend of clear-wood grain angle is the direction of a line with a fixed angle with the longitudinal direction of the board, one can generate the local variations of grain angle around a knot by generating the fluid direction around the knot ([71]).

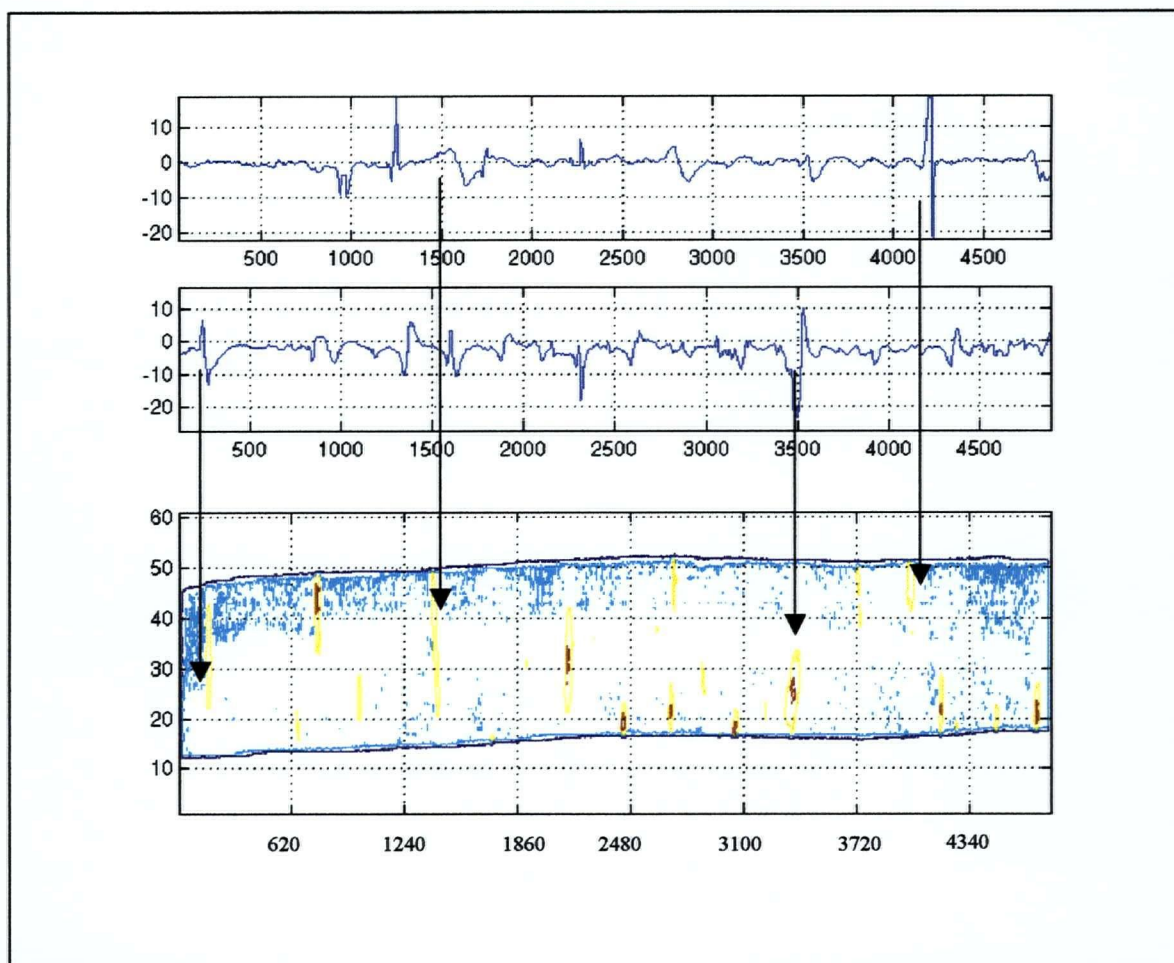


Figure 5.28. Most of the knots show a grain variation in the grain angle profile. The top two profiles are grain angle profiles of the top half and bottom half of the specimen #100. The bottom image is the X-ray image of the same specimen.



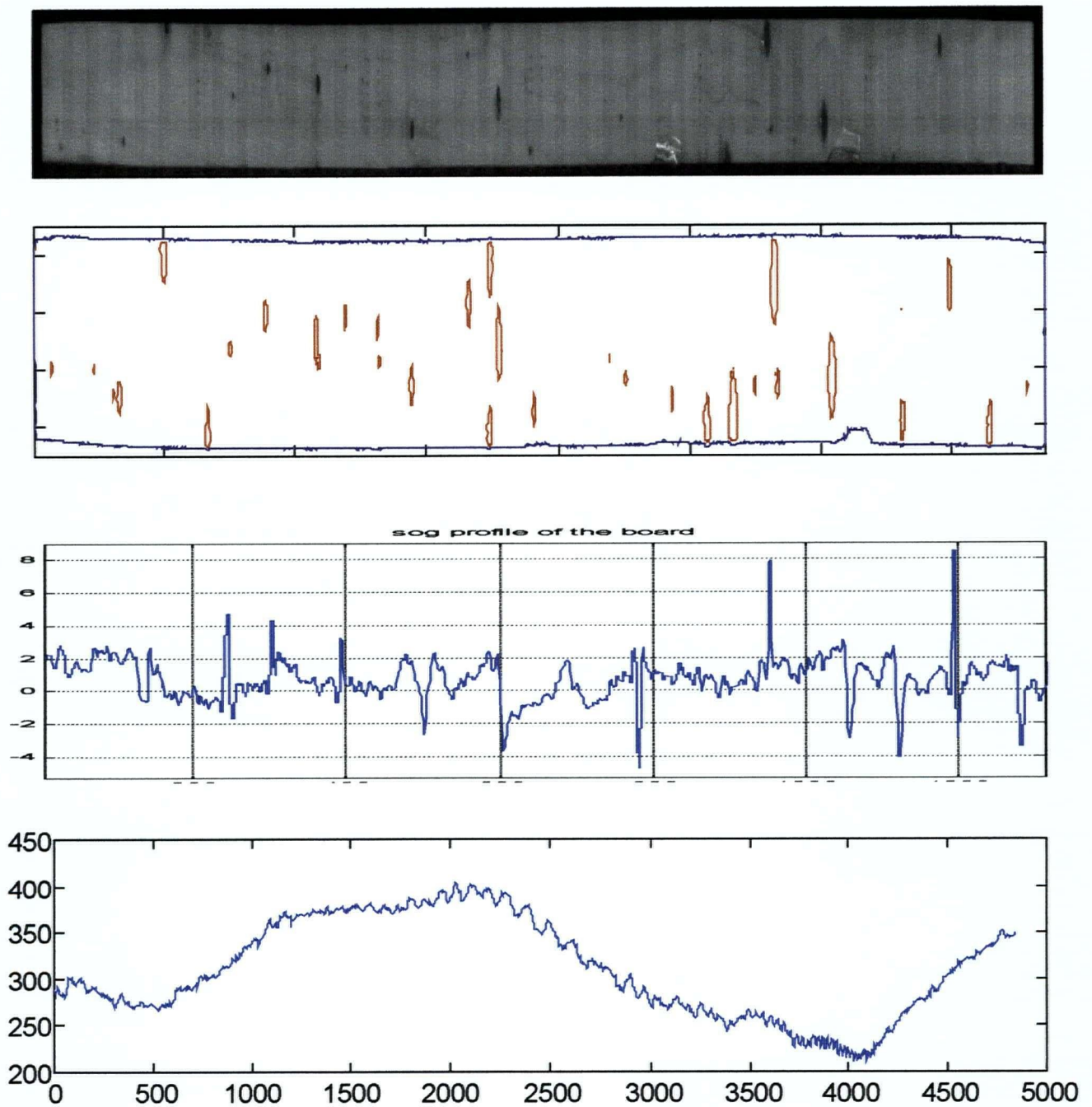


Figure 5.29. Visual image (top), X-ray image (second from top), SOG profile (second from bottom), and MOE profile (bottom) along the board (mm).

## 5.9. Sufficiency of sampling rate; the problem of resolution

The sampling rate of the natural properties of wood that is regarded as the resolution of the measurement system can have direct impact on the accuracy of the estimation because it can change the information content of the input signal to the estimation process. The resolution of the image compared to the size of a knot is the controlling factor in geometrical feature sets. The longitudinal resolution of the current X-ray image is close to 3mm ( $\sim 1/10$  inch), which is very much smaller than the size of strength determining knots.

In the following, the effect of the sampling rate of the signals on the estimation accuracy will be studied by using Fourier analysis of MOE and SOG. By comparing the spatial distance of measurement points (along the board) with the sampling instances of an analog signal, one can analyze the sufficiency of the sampling rate for measured signal. As will be shown in the following, the sampling rate of the slope of the grain signal and the modulus of elasticity signal are much more than the signal bandwidth, therefore no aliasing takes place and no part of the information is lost. Figure 5.30 shows the Fourier transform of an MOE signal. In order to capture the effect of the sampling rate, the average of the signal is removed. Figure 5.30, shows that the bandwidth of the signal is much less than the sampling rate.

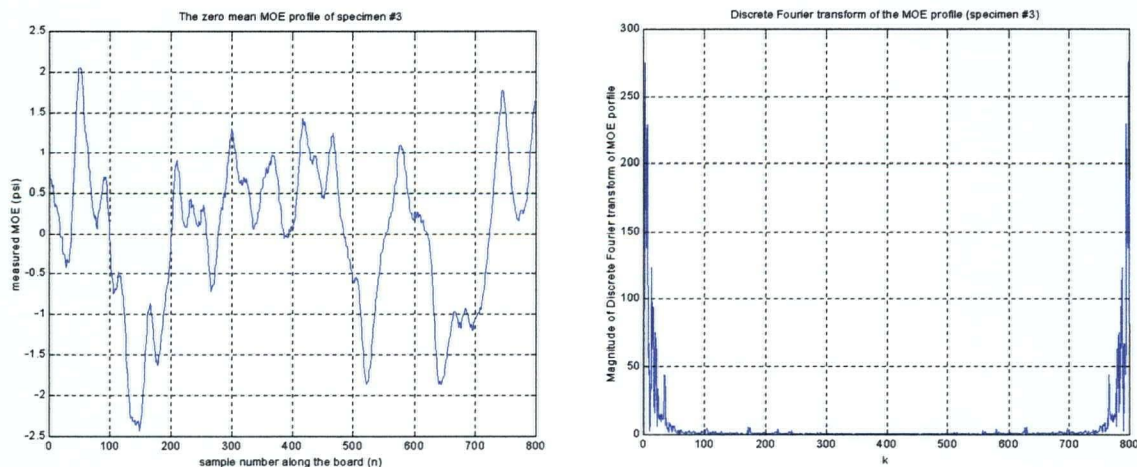


Figure 5.30. The measured signal and its Fourier-transform of a typical MOE profile with zero mean (specimen #3).

$m$  = The measured MOE profile of the board

$m' = m - \bar{m}$  = The zero mean profile (shown in Figure 5.30)

$$M(k) = \frac{1}{N} \sum_{n=1}^N m'(n) e^{2j\pi(k-1)(n-1)/N} \quad (57)$$

As is shown in Figure 5.30, the maximum of the Fourier transform magnitude ( $|M(k)|$ ) at low frequency is more than 275 ( $\max(|M(k)|) > 275$ ), and the magnitude of the Fourier transform is less than 0.8 ( $|M(k)| < 0.8$ ) for  $k > 300$ . Therefore, the energy of the signal at the tail of the Fourier transform (near  $k=400$ ) is much less than the energy at low frequency levels ( $k < 50$ ) therefore the measured signal will not be affected by aliasing.

Similarly, the grain angle profile and its discrete Fourier transform are shown in Figure 5.31. The average of the grain angle profile is not removed because the average grain angle is usually small.

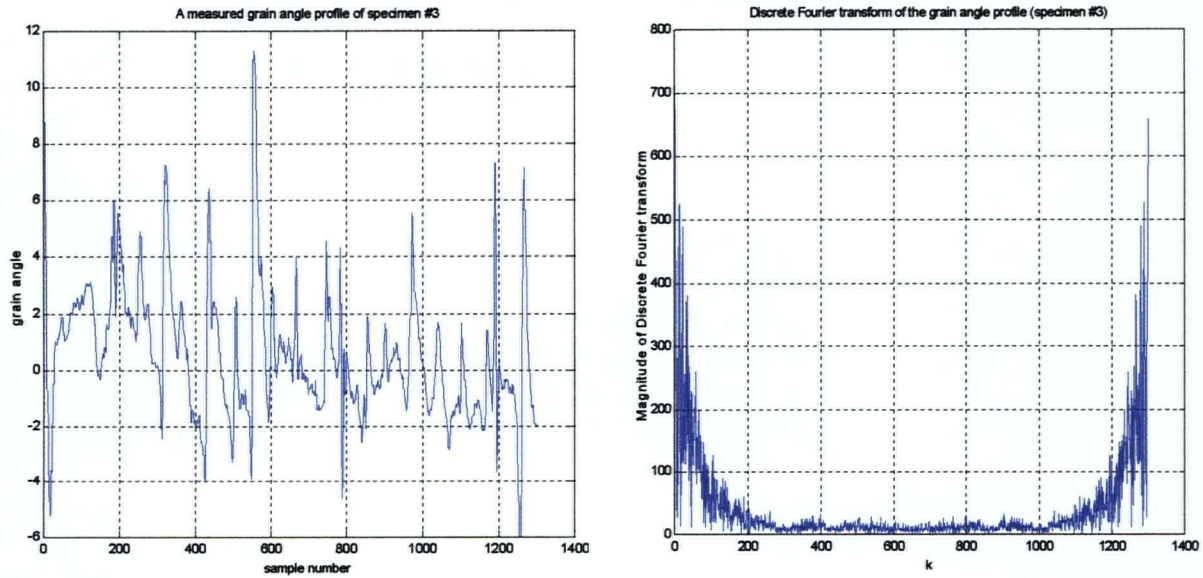


Figure 5.31. The measured signal and its Fourier-transform of a typical SOG profile (specimen #3).

Similar to equation Figure 5.31, the discrete Fourier transform of the grain angle profile can be written as follows.

$s(n)$  = The sampled grain angle profile of the specimen.



$n$  = The sample number.

$N$  = Total number of samples in a profile. The number of samples is 1300.

$S(k)$  = The discrete Fourier transform of  $s(n)$ , defined as follows:

$$S(k) = \frac{1}{N} \sum_{n=1}^N s(n) e^{2j\pi(k-1)(n-1)/N} \quad k = 1, \dots, N \quad (58)$$

As is shown in Figure 5.31, the maximum of the Fourier transform magnitude ( $|S(k)|$ ) at low frequency is more than 300 ( $\max(|S(k)|) > 300$ ), and the magnitude of the Fourier transform is less than 20 ( $|S(k)| < 20$  for  $k > 600$ ). The energy of the signal at the tail of the Fourier transform (near  $k = 650$ ) is much less than the energy at low frequency levels ( $k < 200$ ) therefore the measured signal will not be affected by aliasing.

## 5.10. Preparation of specimens and measurement process

During the course of this thesis two complete sets of data were produced and analyzed. The first set of boards was used for developing the required skills and producing the programs. The second set of boards was used for actual analysis and comparison of models and methods.

The boards were selected from typical production from mills in BC. A few of the boards had suffered damage during shipment and were removed from the database. Once the set of boards for the experiments was selected each board was assigned one number (starting from 1). The code of each board was written on one cross-section of the board and was used in all the measurements. By using the board code the direction and side were assigned to a board by having the code in front and upward (not upside down). This direction was needed to be consistent through X-ray measurement for geometrical feature extraction.

The boards were dry (moisture content less than 12%). The moisture content of the boards was measured before all the measurements to make sure that the boards remain dry. The boards were kept under roof and away from rain and snow. For transportation,

the boards were first wrapped in plastic covers to make sure that they are kept from the elements.

The CLT™ machine has two outputs (BNC connections) that port the (amplified) output voltage of the two MOE sensors. An Atlantis data acquisition board (in a laptop) with five channels of analog inputs (only two channels were used) was connected to the CLT™ output port. The driver program of the data acquisition board had the option of transforming the existing data files from binary to text. This option was later (in the laboratory) used to transform the data file format from binary to text format. When the size of the data files was not too big, (i.e., it could be saved on internal hard-drive), the text format was more desirable because the file could be checked by a variety of different applications and proved useful through the experiments.

The Newnes Advantage2 X-ray machine has a set of measurements and wood cutting tools that are controlled by a group of computers. This machine also had a conveyor system that consistently moved the board through the machine. The output of the X-ray sensors was saved in the machine's memory. The data files were then downloaded from the computer and transferred to the laboratory at UBC. The Newnes Advantage2 output files were named in sequential order. As each board was going through the machine an assistant registered the code of the board. The table was later used to change the file name so that it showed the board code. This will be explained in detail later.

### **Removing the surrounding area**

In the following, three terms will be used frequently that should be defined first. The measured signal is related to what is saved in a data file. There can be more than one measured signal in a data file. For example a MOE data file contains two measured signals. The measured signal of a data file (MOE, SOG, or X-ray) either contains a measured one-dimensional profile (or simply a profile) of the board or it contains an image (X-ray image) of the board. We use the term *profile* for one-dimensional profiles as well as the images.

All the measured signals contain a surrounding area to each profile. For one-dimensional profiles the surrounding area is the output of the measurement system before and after the



actual profile. For X-ray images the surrounding area is the part of the image that is not covered by the board. The surrounding area contains only measurement noise and is easily separated from the measured profile by thresholding as follows. Since the density of clear wood in X-ray image is detected by statistical methods, the threshold level was set to be almost half of the density of clear wood. A scan algorithm searches for the beginning and the end of the board in two dimensions of the image. Starting from the start or the end of each axis of the image, the scan algorithm searched for the point where the measured density level was more than the threshold level and marked them as the beginning or the end of the X-ray image.

The surrounding area in MOE signals was only the noise level (at a level of 0.001 volt) that was much below the MOE profile level (at least 2 volts). In order to extract the measured MOE profiles from the data files a fixed threshold level (0.1 v) was used. Starting from the start or the end of a profile, a scan algorithm searched for the point where the measured signal level was more than the threshold level and marked them as the beginning and the end of the MOE profile. As was discussed in Chapter 2 the dynamic bending machine (CLT™) produces two MOE profiles. If the two profiles are significantly different it shows that the board is bent. In order to reduce the effect of the bend in a board the average of the two profiles was saved as a third output signal of the dynamic bending machine. Therefore, the number of MOE profiles is three.

In the capacitance-based grain-angle-measuring system the output is kept constant when the measurement head is not in contact with wood. For finding the beginning and the end of the measured grain angle profile, the scan algorithms search for any sudden change in the measured signal. Once the two points (start and end) are marked a small portion of the profile that corresponds to a few centimeters of the length of the board were removed in order to exclude the transient response of the measurement machines. The output was considered as the measured profile of the board and was used for feature extraction. There were six measured profiles per specimen. The structure of the capacitance based grain angle measurement was discussed in Chapter 2. The size of the head was about 5 cm. In order to cover the wider face of a board two scans were carried out. Also, one scan for each narrower face of a board was carried out. Therefore, the output of the capacitance based grain-angle-measuring machine is six signals.

Similar to MOE signals the surrounding area in Microwave signals was noise. In order to extract the measured grain angle profiles from Microwave signals the average of the signal was used as a threshold level. Starting from the start or the end of a profile, a scan algorithm searched for the point where the measured signal level was more than the threshold level (half the average of the measured signal). The rest of the algorithm is the same as that of the capacitance based system. The output signals of the Microwave system were; grain angle, grain alignment, longitudinal signal amplitude, transverse signal amplitude, longitudinal signal phase, and transverse signal phase.

**Table 2: The number of output signals of measurement machines**

	Board density (X-ray)	Grain angle (capacitance based)	Grain angle (Microwave)	Modulus of Elasticity (CLT™)
The number of output profiles	<b>2</b>	<b>6</b>	<b>6</b>	<b>3</b>
The type of output profile	Two dimensional density image	One dimensional profile	One dimensional profile	One dimensional profile

### **Practical considerations and the database**

Boards were first coded in the laboratory and few broken specimens were removed. The broken specimens have been broken in transportation or have a relatively large fracture along and partly across the board. The board code was written at the cross section of each specimen.

There was little or no control on keeping the order of specimens in a set when the specimens were to be taken from a bundle.

All the data files that were produced by the measurement machines were in sequential order with regard to the measured specimen. For example the name of the stored data files from the dynamic bending machine were CLT\_0001.data, for the first specimen that was measured and CLT\_0002.data for the second specimen.

In every measurement process the code of every passing board was registered by an assistant as follows:

1 0014

2 0528

Therefore, it was known that the file CLT\_0001.dat was related to specimen 0014 and that the file CLT\_0002.dat was related to specimen 0528.

It is obvious that any mistake in registering the specimen code would directly cause a mix up in the database. The experiments had to be finished within a limited time slot due to the availability of the Newnes machines. Therefore, there was little chance for repeating a part of an experiment.

In order to avoid specimen mix up there was a repeated measurement of the same specimen as is explained in the following. After a specified number (say 20) of the specimens were measured the last specimen was measured again. This repeated measurement made two relatively identical data files in the measured data files.

**Table 3: Specimen file names**

Sequence of the passing board	The board code	Final file name
1	0014	CLT_0001.dat
2	0528	CLT_0002.dat
⋮	⋮	⋮
19	0751	CLT_0019.dat

20	1049	CLT_0020.dat
21	1049	CLT_0020.dat
⋮	⋮	⋮
40	0081	CLT_0040.dat
41	0081	CLT_0041.dat
⋮	⋮	⋮

Any error in registering the board codes would appear as a shift in this table. There was rarely any mistake in registering the board, however, any error would cause a number (usually 20) to either be removed from the database or be sent back for another round of measurement. In the end, the table of specimen codes would match all the (remaining) data files.

Then based on the board code table, the data file names were changed so that the file name represented the board code. For example,

1 0014 CLT\_0001.dat

CLT\_0001.dat → Landmark\_CLT\_0014.dat

If possible, the format of the data files was also changed to an easier to handle format for future research.

Landmark\_CLT\_0014.dat → Landmark\_CLT\_0014.txt

The strength of all the boards was recorded in one file, which contained two columns, one for board code and one for the registered strength. The database is a structure of directories that represent the related measurements as follows.

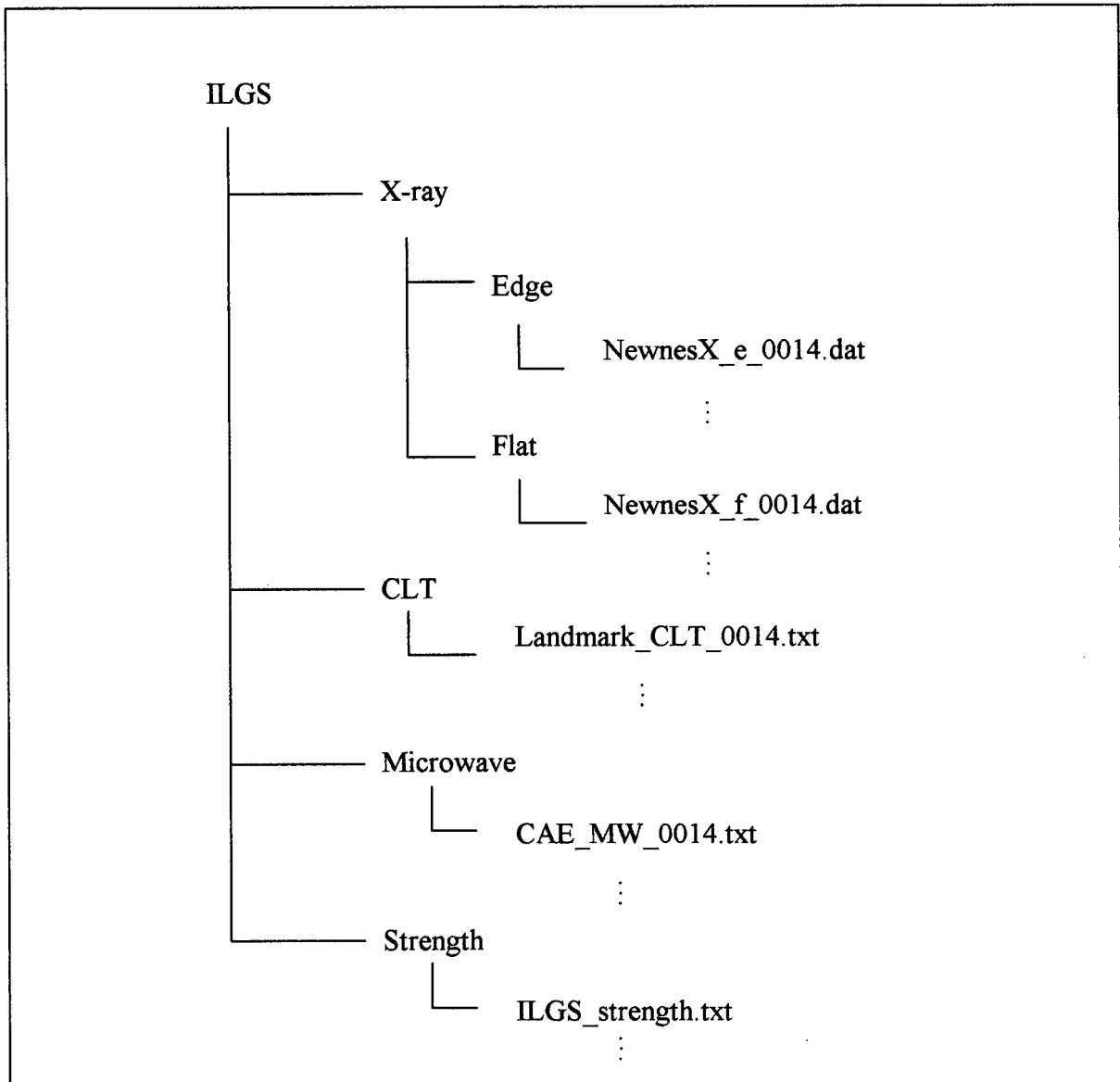


Figure 5.32. The database structure.

The X-ray files are binary files that contain 128x1400 matrices. Each element is represented by 16 bits. The CLT™ file is a text file that is a matrix of 4000x3. The first column is the location indicator, the second column is the first dynamic bending machine profile (cell 1), and the third column is the second dynamic bending machine profile. The microwave file is a text file that contains a matrix of seven columns. The first column is the location indicator, and other columns (as was said before) were the related grain angle, grain alignment, longitudinal signal amplitude, transverse signal amplitude, longitudinal signal phase, and transverse signal phase.

The strength file is a text file that contains a matrix of size two columns; the first column is the board code, and the second column is the tensile strength (MPa).

All the board codes of the specimens in the database are included in a specimen list text file (ILGS\_boards.txt) that is placed in the ILGS directory.

A data processing program reads the board codes from the specimen-list file, generates the name of all the related data files, and loads the files from the pre-specified directories of the database.

### 5.11. Specimen Characterization

In a pilot study in the Wood Science Department of UBC, the statistical features were applied to the raw data. In the work reported here the following features were computed. For one-dimensional profiles the average, minimum, maximum, standard deviation, variance, the standard deviation of the absolute value, Fisher skewness, Fisher kurtosis of the profile are the feature set of the profile. These features are compatible with those used in previous statistical analysis in the Wood Science Department of UBC. These features characterize the board. Fisher skewness ( $sk$ ) and Fisher kurtosis ( $k$ ) are defined as follows.

$$sk(x) = \frac{1}{N} \sum_{i=1}^N \frac{(x - \bar{x})^3}{\sigma^3} \quad (59)$$

$$k(x) = \frac{1}{N} \sum_{i=1}^N \frac{(x - \bar{x})^4}{\sigma^4} \quad (60)$$

In the above equations,  $x$  is the variable,  $N$  is the number of samples in the profile,  $\bar{x}$  is the average value of the variable, and  $\sigma$  is the standard deviation of the variable  $x$ .

For X-ray feature definition the geometrical features were produced. From the X-ray images of the board the knots were detected and a model board was generated. The coordinate of a knot is shown in Figure 5.33. For each knot the centers ( $x, y, z$ ) of the two

end ellipses and the diameter of the cone at the two centers were saved as the parameters of the knots. For each knot the characterizing features are as follows.

1.  $F_1$ : Volume
2.  $F_2$ : Axial deviation (average distance of the knot pixels from the center of the board),
3.  $F_3$ : Axial standard deviation (standard deviation of the knot pixels from the center of the board),
4.  $F_4$ : The ratio of small diameter to the large diameter,
5.  $F_5$ : Closest distance to a neighboring knot,
6.  $F_6$ : Knot area ratio (KAR: the ratio of a knot cross-section to the board cross-section)

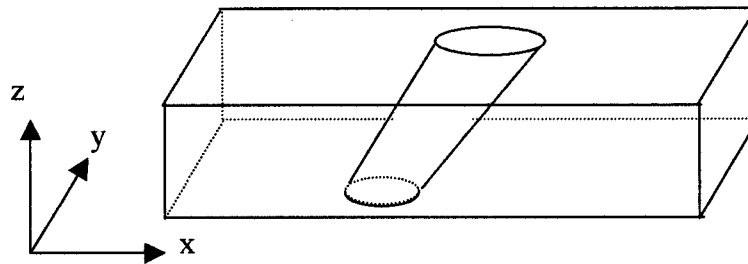


Figure 5.33. The coordinate of a knot in a board

These features were processed to produce a final set of features as follows.

1. The average of knot area ratio ( $F_6$ ) over all the knots of the board
2. The maximum of knot area ratio ( $F_6$ ) over all the knots of the board
3. The standard deviation of knot area ratio ( $F_6$ ) over all the knots of the board
4. The average of axial deviation ( $F_2$ ) over all the knots in the board

5. The standard deviation of axial deviation ( $F_2$ ) over all the knots in the board
6. The average of axial standard deviation ( $F_3$ ) over all the knots
7. The standard deviation of axial standard deviation ( $F_3$ ) over all the knots
8. The board distribution mode. This feature is equivalent to the density of clear wood of the board.
9. The standard deviation of the density distribution of the board
10. The variance of the density distribution of the board
11. The average of the distance between neighboring knots ( $F_5$ ). For each knot the neighboring knot is the closest knot.
12. The standard deviation of the neighboring knot distances ( $F_5$ ) over all the knots in the board.
13. The number of knots in a board
14. The ratio of knot volume to the board volume
15. The board volume is identified by the X-ray image
16. The board density, which is the summation of the density over the board. This feature is the summation of all the pixels of the X-ray image of the board. If two images exist, the average of this feature for two images is used.

### **Geometrical feature extraction program**

The flow chart of the knot detection program is shown in the following. This program is a part of geometrical feature extraction program that follows.



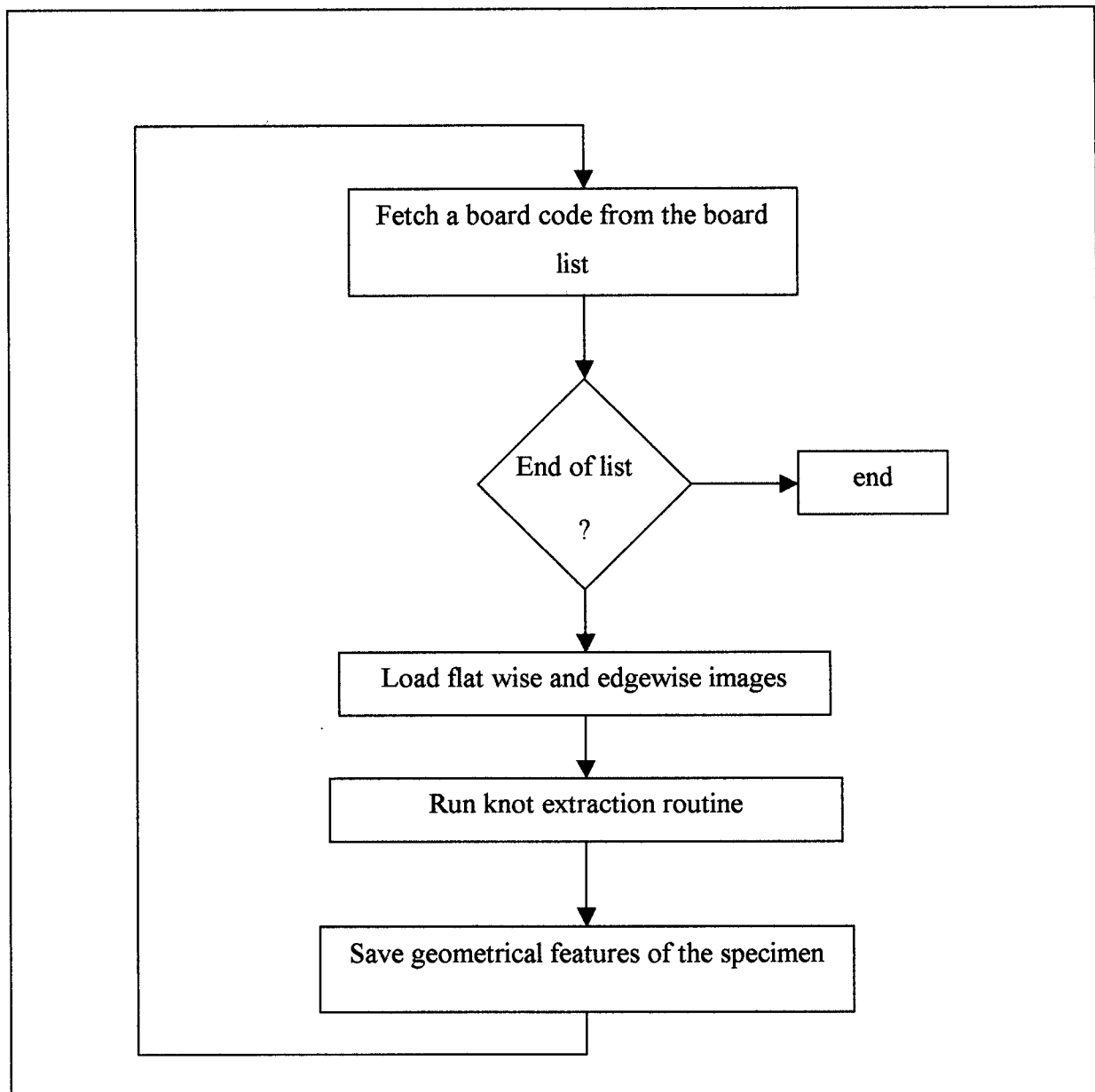


Figure 5.34. The general flow chart of geometrical feature extraction program

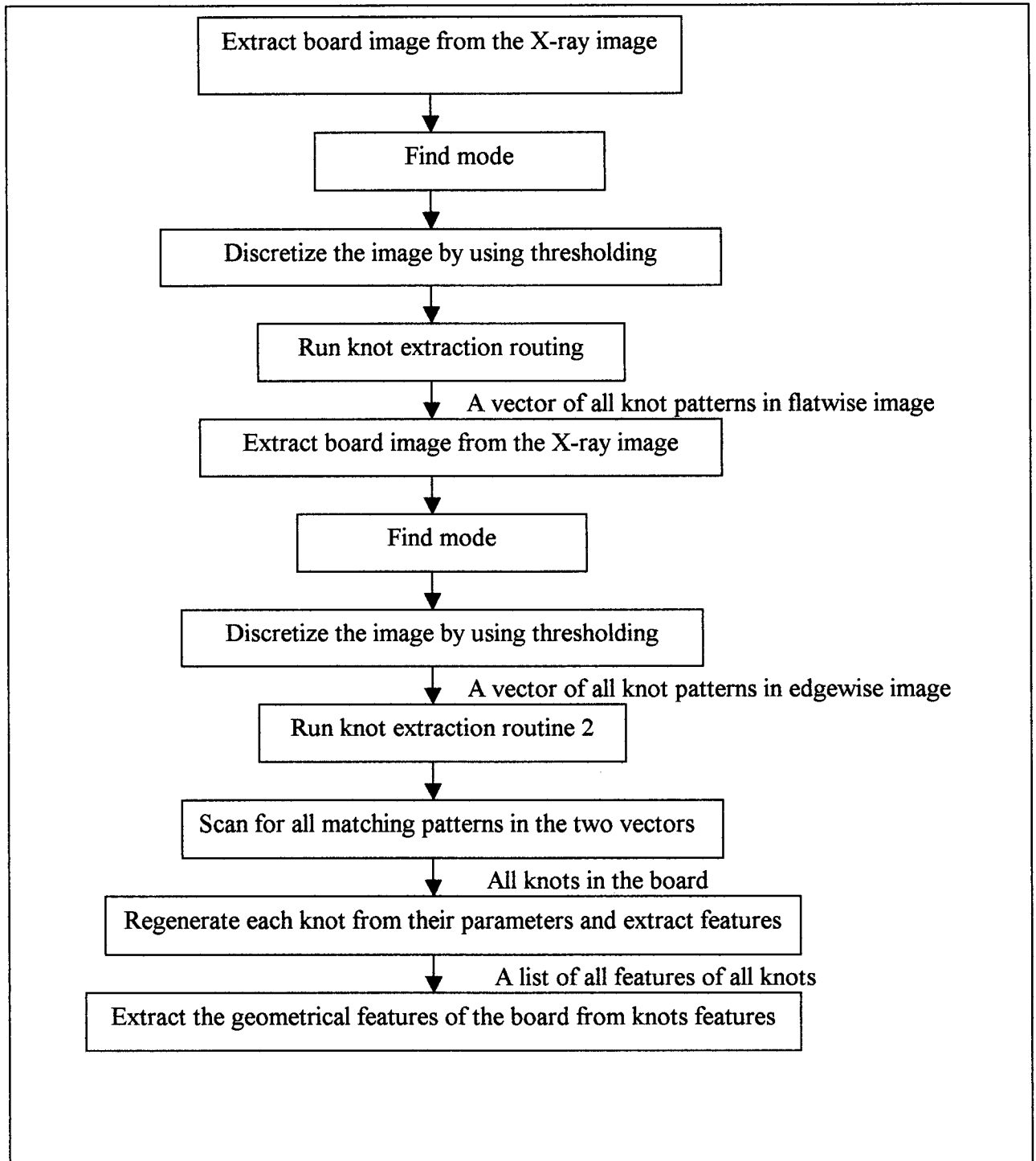


Figure 5.35. The flow chart of the knot extraction routine

Knot pattern extraction routine is as follows:

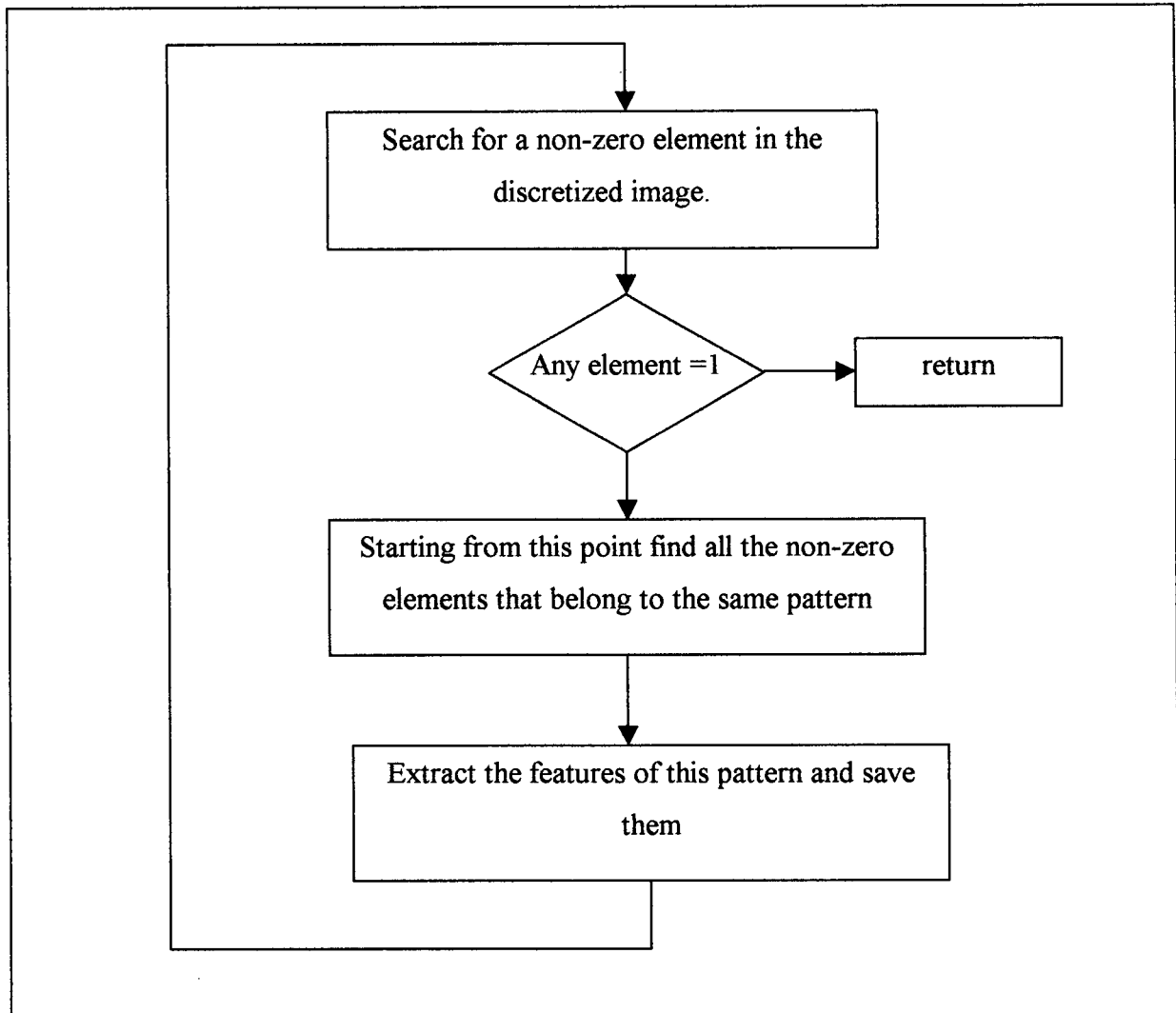


Figure 5.36. The flow chart of knot extraction routine 2

### Statistical feature extraction program

The feature extraction program for one-dimensional profiles is the same for grain angle and for dynamic bending machine profiles. Only the threshold levels are adjusted.

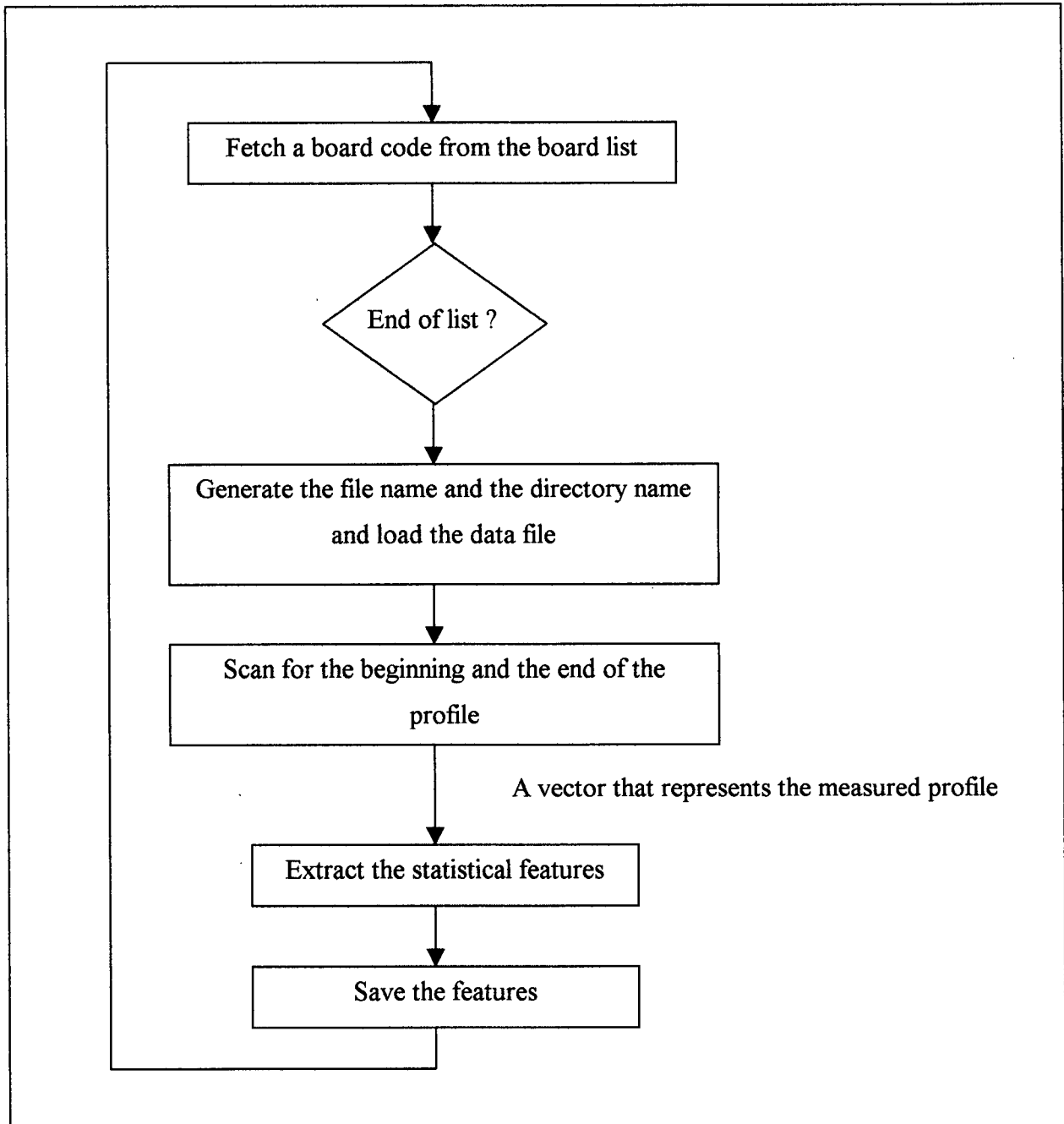


Figure 5.37. The flow chart of the statistical feature extraction program

The output of each feature extraction program is a matrix whose first column is the board code and other columns are the extracted features. Once all the feature files are generated one output file is produced that is the combination of all the extracted features. The first column is the board code, then all the features (dynamic bending machine, Microwave,

and Geometrical features), and the last column is the measured strength of the specimen. This file is used in all the learning algorithms of Chapter 6.

## ***Chapter 6. Strength Estimation; Empirical Learning for Small Sample Size***

The final stage of the strength estimation process is the empirical learning problem. Once the feature set is defined and the strength of each specimen is measured, a suitable machine should be chosen in order to produce the best predictive model. Due to the high cost of generating a learning set for wood strength estimation, the number of specimens in the training set is very limited. Therefore, a number of different learning machines need to be examined to discover what model works best.

The problem of learning (or function approximation) can be stated as finding the optimal parameters of the approximating function by minimizing a risk functional. The type of input variables (continuous or discrete) and the task (regression or classification) of the function should be defined. As were shown in the previous chapters the features of a board, which are the input variables of the approximating function, are numeric variables. The approximation problem is to find a regression function to estimate the strength of a new board.

In Section 6.1 a review of learning systems is presented in order to create the common ground for various learning machines for strength estimation, as described later in this chapter. In this section the principles and components of a learning method are discussed and the relationship between two accuracy measures (also called performance criteria), such as the mean square error and the coefficient of determination is presented. The coefficient of determination between measured and estimated strength is used in the standard of wood grading. Mean square error is usually recognized as the measure of accuracy in learning methods.

The problem of high dimensionality for feature space is discussed in Section 6.2. The high dimensionality of feature space is known as "the curse" or complexity of dimensionality ([54]).

In Section 6.3 a review of the theoretical upper bound ([66]) on the error of a learning system is presented. The upper bound of the error is a function of sample size and the learning capacity of the learning system. Therefore, in Section 6.3 the relationship between the sample-size, the learning-capacity of the learning machine, and the estimation error is presented. The learning-capacity of a learning machine is the factor that defines how suitable the machine is for the given task. The understanding that is developed in this section will later be used in the ASEC model (Section 6.6). A practical method for measuring the learning capacity of a classifier exists in the literature ([73]). The equation of the upper bound error is valid for both classifiers and regression estimators ([81]), assuming that the learning-capacity of the machine is known. A practical method of learning-capacity measurement for regression estimators is presented in Section 6.4. The learning capacity of regression estimators is defined based on the definition of the learning capacity of classifiers and thus, the presented capacity measurement method for regression estimators is based on the existing method for the learning-capacity measurement of classifiers.

Different learning machines that have been used for strength estimation are discussed in Section 6.5, and the ASEC algorithm is presented in Section 6.6. The existing learning machines of Section 6.5 are linear regression, K-nearest neighbor ([74]), the radial basis network, the multi-layer neural network ([75]), and SPORE ([21]).

### **Parametric Learning Machines**

The linear regression method is commonly used in practice and is a parametric learning machine. Parametric learning machines develop a model by assuming that the data distribution has a known parametric form. A parametric learning machine can result in constant error due to irresolvable biasing, because of the restriction imposed by the fixed structure of the approximating function.

### **Nonparametric Learning Machines**

Neural networks, K-nearest neighbor, SPORE, and ASEC, are non-parametric learning machines. These machines reduce the bias error by changing the complexity (see sub-Section 6.1) of the learning machine. A simple multi-layer neural network consists of

three layers (input, output, and middle layers) with one layer (middle layer) for non-linear mapping of the input to the output ([75]). The K-nearest neighbor algorithm and radial basis networks estimation methods are also called memory based learning machines and do not generate a model from the given data but predict the output for a new specimen by finding its closeness to the given samples.

SPORE is a learning method that was developed for learning conditions similar to that dealt with in this research. As explained in Chapter 4, a minimum set of features is not known. Therefore, there may potentially be a group of features that are poorly correlated with the tensile strength of a board. SPORE was developed specifically for this condition and for when a large number of features and noisy input and output exists with no *a priori* knowledge about the input-output (i.e., feature-strength) relationship.

The ASEC algorithm is developed in this thesis in order to include the *a priori* knowledge about the input-output (feature-strength) relationship in the learning process. In ASEC, the features and the input-output transformation are selected such that the estimation error is minimized. The *a priori* knowledge is included in the form of input-output transformations. ASEC uses subset selection ([82]) and space expansion ([83]) and learning system error estimation (Section 6.1 and [81]) in order to transform the given set of features to a new set of features with minimum size and best accuracy of estimation.

## **6.1. Strength estimation as an empirical learning problem**

In this section the problem of wood strength estimation is looked at as a learning problem, or equivalently, a function approximation problem. The problem of learning (or function approximation) can be modeled as finding the optimal parameters of the approximating function by minimizing a risk functional. The type of input variables and the task of the function should be defined. As were shown in the previous chapters the features of a board, which are the input variables of the approximating function, are numeric variables. The approximation problem is to find a predictive model to estimate the strength of a new board.



The process of data generation and the learning problem are shown in Figure 6.1. The sample space is the collection of boards randomly picked from a typical mill-run in the region. It is assumed to be a typical sample of the board production. The actual strength affecting factors are denoted by  $F$ . These factors may or may not be measurable. The feature set presented to the learning system is denoted by  $\hat{F}$ . These features were presented and discussed in Chapter 5. The measured strength of the board (the output of the destructive test) is denoted by  $S$ . The learning system estimates a dependence between  $\hat{F}$  and  $S$ . The estimate of the strength is the output of the learning system ( $\hat{S}$ ).

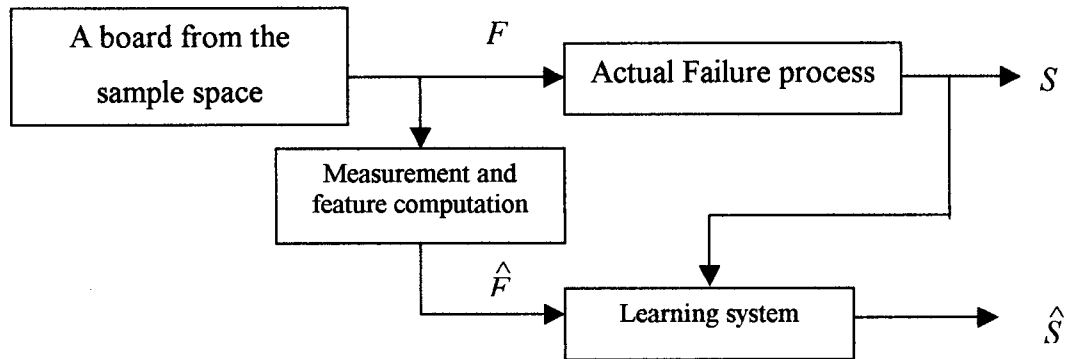


Figure 6.1 The learning problem

It is obvious that the input variables to the learning system are not the same as the actual factors of the failure process. This has been explained in detail in previous chapters. The features in  $\hat{F}$  are the features, among measurable features, that are reasonably related to board strength. The learning system uses  $S$  and  $\hat{F}$  to generate an estimate of the strength ( $\hat{S}$ ). The whole strength estimation problem is then formulated as follows:

$$S = \hat{S} + \varepsilon \quad (61)$$

In order to separate the analytical discussion of the function approximation problem from the wood strength estimation problem, the regular notation in the literature is used in the following. Therefore  $\hat{F}$  is denoted as  $x$  and  $S$  is denoted as  $y$ . The estimated strength,  $\hat{S}$ , is

denoted as  $f(x,w)$  (or  $\hat{y}$ ) which is the output of the approximating function and  $w$  denotes the set of parameters of the learning system.

The learning problem is stated as estimating the parameters of the approximating function in order to find the best mapping from  $x$  to  $y$  from a set of given  $(x,y)$ . A set of approximating functions,  $f(x,w)$ , are chosen to estimate the dependence. It is assumed that the dependence between  $x$  and  $y$  can be modeled by the chosen set of  $f(x,w)$ . The choice of the approximating function depends on the researcher and depends on the nature of the dependence between  $x$  and  $y$ . Different forms of  $f(x,w)$  were tested in order to find the best model.

A learning problem can be formulated as a minimization problem. The goal is to minimize a risk functional,  $R(w)$ , which determines the difference between the actual dependence between  $x$  and  $y$ , and the estimated dependence that is modeled by  $f(x,w)$ .

$$R(w) = \int L(y, f(x,w)) p(x,y) dx dy \quad (62)$$

A common empirical risk functional, which is an estimate of  $R(w)$ , is used for regression purposes as shown in equation (64).

$$L(y, f(x,w)) = (f(x,w) - y)^2 \quad (63)$$

The risk functional can be estimated as follows.

$$R_{\text{emp}}(w) = \frac{1}{n} \sum_{i=1}^n (f(x,w) - y)^2 \quad (64)$$

Where  $R_{\text{emp}}(w)$  denotes the empirical risk functional that will be minimized in the estimation problem. The parameter vector  $w$  is the vector of all the adjustable parameters of  $f(\cdot)$ .

Therefore the dependence between  $x$  and  $y$  is estimated as shown in equation (65).

$$y = f(x,w) + \varepsilon \quad (65)$$

Equation (65) is equivalent to equation (61). The difference between  $y$  and  $f(x,w)$  can be contributed by two factors: the inherent difference between the true dependence (between

$y$  and  $x$ ) and  $f(x,w)$ , and the measurement noise. In a small neighborhood in the feature space the risk functional can be written as in equation (66) ([72]):

$$E[(f(x,w) - y)^2] = E[ (f(x,w) - E[f(x,w)] )^2 ] + (y - E[f(x,w)])^2 \quad (66)$$

The function  $E(.)$  represents an average over all training samples. The first term of the right hand side of equation (66) is the variance of the approximating function. This term is related to the complexity of  $f(x,w)$  and will be explained later. The second term is the square of bias error, which is the actual error between the approximating function and the true dependence.

The **complexity** of  $f(x,w)$  defines how fast  $f$  changes with respect to the variation of its input,  $x$ . The complexity of  $f(x,w)$  can be measured with respect to a given set of data (as will be done later). However, the complexity of a function may be defined by using the number of its parameters. For example for a linear model, the complexity of the function is equal to the number of coefficients. Generally speaking, depending on whether the complexity of  $f(x,w)$  is low or high, the dominant estimation error can be bias error or variance error.

Informally speaking, if the complexity of  $f$  is limited, the resulting approximating function cannot follow all the variations of the measured output and the bias error will be dominant in equation (66). In this case, the variance error will be limited. An obvious form of bias error happens when the set of approximating functions cannot estimate the true dependence, for example, trying to approximate a second-degree dependence ( $y = x^2$ ) using a linear regression.

If the complexity of  $f(x,w)$  is too high, the variance error will dominate equation (66). The variance error is a result of having a limited number of specimens for training the learning machine. In this case, if the complexity of  $f(x,w)$  is too high, the approximating function follows all the variation of  $y$  that is caused by the noise.

Assuming that  $f(x,w)$  can model  $y$ , minimizing  $R_{\text{emp}}(w)$  as shown in equation (64) generates little bias error in training stage, but large error results due to the variance of the resulting function,  $f(x,w^*)$ . Therefore, a second factor to be minimized is the complexity of the estimating function  $f(x,w)$ . Therefore, finding the right parameters of

the approximating function in the learning problem means finding a compromise between the bias error and variance error.

### Model development

In order to solve the model development problem four stages are followed:

1. a set of approximating functions is chosen;
2. the *a priori* knowledge or assumptions are used to place constraints on the approximating function (in order to reduce the variance error);
3. an inductive principle is chosen in order to reach a solution and compare the results (Section 6.1.2);
4. a learning method (e.g. an optimization algorithm) is developed in order to construct the approximating function based on the inductive principle.

Each of the parts of this process is explained in the following.

The structure of the **approximating function**  $f(x, w)$  depends on the choice of the designer. The definition of the approximating function's complexity is related to *a priori* knowledge about the type of dependence. Basically,  $f(x, w)$  belongs to a wide class of functions. Most of the time the approximating function,  $f(x, w)$ , is generated by using a class of basis functions. Based on the type of the basis functions, the approximation method can be called universal or local. Also based on the freedom of the class of basis functions the approximation method is called parametric, semi-parametric, or non-parametric. In parametric form the shape of the approximating function is defined *a priori* (e.g. linear regression) while a non-parametric structure uses the data to find the right shape of the approximating function. In a semi-parametric form (also called dictionary method) there is a limited number of basis functions that a designer can choose from. For example, linear regression is a parametric structure of an approximating function, and a multi-layer neural network is a semi-parametric form of structure. An example of the universal approximator is  $x'$  that can construct the polynomials ( $f(x, w) = \sum_{i=1}^n w_i x^i$ ). The SPORE algorithm is a non-parametric structure that will be introduced in subsequent sections.

The *a priori* knowledge can be used to limit the set of approximating functions, or limit the optimization goals. For example, the early stopping rule in multi-layer neural networks limits the complexity of these learning machines.

There are different **inductive principles** that lead to defining the complexity term and minimizing the risk functional of equation (62). These principles are Bayesian inference, the penalization principle, the neural network early stopping rule, the minimum description length, and the structural risk minimization ([66]). As will be shown, we will use the penalization principle and the neural network early stopping rule. Structural risk minimization also overlaps with work in this project, but does not add to it, and therefore, is not studied in detail.

In the penalization inductive principle,  $R_{\text{emp}}(w)$  is defined as follows:

$$R_{\text{emp}}(w) = \frac{1}{n} \sum_{i=1}^n (f(x, w) - y)^2 + \lambda \Phi[f(x, w)] \quad (67)$$

where  $\lambda \Phi[f(x, w)]$  is the term that measures the complexity of the approximating function. The penalization functional,  $\Phi[f(x, w)]$ , is defined based on *a priori* knowledge of the data. The regularization parameter,  $\lambda$ , is derived from the data in order to make the learning problem data driven. This means that  $\lambda$  should be chosen so that minimizing  $R_{\text{emp}}(w)$  is equivalent to minimizing  $R(w)$ . In the neural network early stopping rule the complexity penalization term is controlled by limiting the repeated training steps.

The complexity of the approximating function can be defined in different ways. The penalty functional,  $\Phi[.]$ , can be parametric or non-parametric. If a function of the parameters of the approximating function defines the complexity of the approximating function, the complexity functional is called parametric. If the penalty functional restricts the shape of the approximating function it is called a non-parametric penalty functional.

For example,  $\sum w_i^2$  is a parametric penalty functional while  $\Phi[f] = \int \frac{|f(s)|^2}{G(s)} ds$  (for a given filter  $G(s)$ ) is a non-parametric penalty-functional. The complexity of the learning methods used in this research is explained with the learning methods.

For a given  $\Phi[f(x,w)]$ , finding the right  $\lambda$  is called complexity control or model selection. As mentioned above, the right  $\lambda$  is chosen so that minimizing the empirical risk functional is equivalent to minimizing the actual risk functional. In order to do so, an estimate of the actual risk functional should be obtained. Then for a series of  $\lambda$ , the approximating function is found by minimizing the empirical risk functional.

$$w^* = \underset{w}{\operatorname{argmin}} \{ \sum_i [f(x_i, w) - y_i]^2 + \lambda \Phi(f(x, w)) \} \quad (68)$$

The desired value of  $\lambda$  is the one that produces an approximating function,  $f(x, w^*)$ , which minimizes the estimated risk functional. Here  $w^*$  is the vector of the parameters of  $f(x, w^*)$ . The input-output relationship is estimated by the following equation:

$$\hat{y}_i = f(x_i, w^*) \quad (69)$$

In order to evaluate the risk functional two approaches can be taken: analytical and numerical. In an analytical approach, the actual risk functional is related to the empirical risk functional by the following equation:

$$R(w) = r\left(\frac{h}{n}\right) R_{\text{emp}}(w) \quad (70)$$

where  $r(\cdot)$  is called the penalization function.  $h$  is a measure of the degree of freedom of the data set and is similar to  $\lambda$ .  $n$  is the number of specimens. This approach is commonly developed for linear estimators.

## Cross-Validation

In the cross-validation technique the data set is divided into two sets, one for training the learning machine and the other for finding the accuracy of the developed learning machine. Therefore, the cross-validation technique uses the data for approximating the actual risk functional,  $R(w)$ . In this method the actual risk functional is estimated by using a portion of the data set not used for training. In this case, the data set is divided into two parts: one part for estimating the parameters of the estimation function, and the second part (validation set) for estimating the risk functional. In this case, different values of  $\lambda$

can be used for developing the approximating function and the estimation of risk functional is used for choosing the optimal regularization parameter ( $\lambda^*$ ).

An effective method of estimating the risk functional is k-fold cross validation. For example, in ten-fold cross validation, the specimens of the data set are randomly categorized into ten groups. Producing the approximation function and estimating the risk functional is repeated as follows. Each group in turn is considered as the validation set. The rest of the data set is used to develop the approximating function. For the group  $i$  the risk functional is then estimated as follows:

$$r_i = \frac{k}{n} \sum_j (f(x_j, w_i) - y_j)^2 \quad (71)$$

The estimation of the actual risk functional is as follows:

$$R(w) \approx \frac{1}{k} \sum r_i \quad (72)$$

The above learning method is good for understanding the learning process. However, in practice other methods can be followed that accomplish the same goal with less effort. For example, in training the neural networks the number of neurons in the middle layer and the training goal are two factors that can define the complexity of the network. Therefore, a two dimensional search can be completed by using the cross-validation technique in order to find the optimal values.

### **Risk functional**

The learning system usually minimizes the risk functional but it is customary (from a statistical point of view) to compare different random variables by using  $r^2$ , the coefficient of determination, defined by the following equation:

$$r(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} \quad (73)$$

The coefficient of determination,  $r^2$ , is related to the risk functional  $\sum (y_i - f(x_i))^2$  in the sense that the maximum of  $r^2$  happens when the minimum of  $\sum (y_i - f(x_i))^2$  is happening.

The following relationship between  $r^2$  and  $\sum (y_i - f(x_i))^2$  is the basis of the comparison (see [84], pp. 531).

$$r^2 = 1 - [\frac{1}{n} \sum (y_i - f(x_i))^2] / \sigma_y^2 \quad (74)$$

Therefore, the risk functional is a Euclidean norm of the estimation error and the approximating functions are  $L^2$  (Lebesgue integrable) functions.

The ten-fold cross validation technique can be used to estimate the risk functional. In this technique, the data set is divided into two parts, one for developing each model and one part for testing all the models as to how well they can predict the model. In order to do so the data set is randomly sorted and 10% of the data is separated for the final prediction test. The training data set is also sorted for 10 fold cross validation. Here, the data set is randomly grouped into ten sets. In each repeat of the training, one of the sets (without replacement) is picked for testing. The remaining nine data sets are grouped into the training data set used for model parameter estimation. The tenth set, which is the test set, is used to find out what the risk functional value is. The process is repeated ten times; therefore, ten values of the risk functional are obtained. The average of these estimated risk functionals determines how well the model works.

## 6.2. High dimensionality of the feature set

As mentioned above and seen in Chapter 5, the representative features are logically chosen such that they seem to be best correlated to strength. The feature set therefore would be high dimensional. The high dimensionality of the feature set makes it difficult to develop an estimator that can desirably estimate the tensile strength because the samples are not densely distributed in the feature space. This issue, which is called “the curse of dimensionality”, is a restricting factor. This problem is usually dealt with by placing restricting constraints on the approximating function  $f(x)$ . This means that  $f(x)$  will be of lower complexity, or of smoother function.



There are four issues that affect the function approximation process when dealing with high dimensional data as follows ([72],[21]). First, the number of samples needed for generating the same density of the data (in the feature space) increases exponentially with dimension. Second, as the dimensionality of the feature space increases, in order to cover a fraction of the feature space volume (e.g. a hyper cube) a greater portion of each axis is needed. Third, in the unit hyper-cube every point in the feature space is closer to an edge than to a neighboring point. Fourth, for every point in the feature space all other points look close to each other. Therefore every point can look like an outlier and the concept of interpolating between sampling points (that is usually associated with function approximation) can be deceptive.

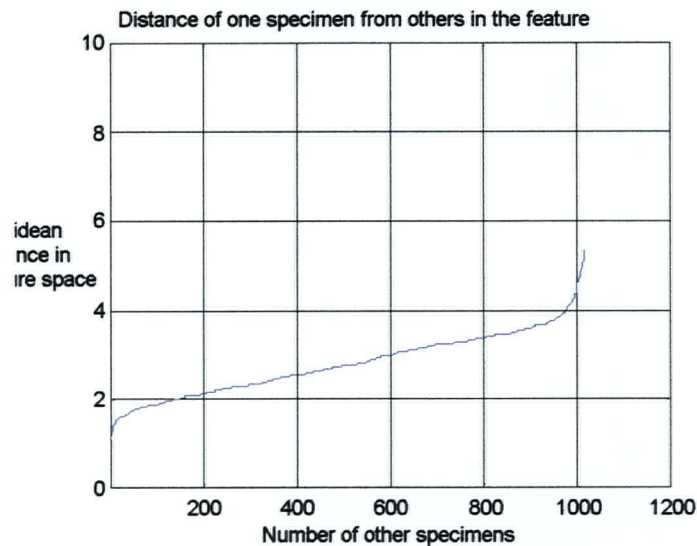


Figure 6.2 Distance of all other boards from a sample board. The Figure shows that the average of the distance is less than 3 while the distance from the closest point is more than 1.

Figure 6.2 demonstrates the high dimensionality of the feature space by showing the variation of distance from a randomly selected point. This Figure shows that the Euclidean distance of a point (in the feature space) from the neighboring points is close to 1, while the average distance of the point from all other points is between 2 and three. Since Figure 6.2 is produced from the feature set of this thesis, it is a demonstration of the

high dimensionality issue in the strength estimation problem. The number of features is 88 and the number of specimens is more than 1000.

### **6.3. The relationship between estimation accuracy, learning capacity, and sample size**

In order to get an estimate of the structure of the learning system, one can use upper bound error estimation based on the learning capacity of a learning machine. The learning capacity of a learning machine (also called the VC-dimension) is a measure of how well a machine can learn the training data. If the learning capacity of the machine is very low, the machine cannot learn the variations in the training data and the trained system will produce a bias error. If the learning capacity of the machine is too high, it will generate a variance error, or the model will simply over-fit its parameters. If the learning capacity of a learning machine is known, statistical learning theory provides the framework for analyzing bias error and variance error. The result of this theory tends to be conservative; that is, an over estimation of error is produced. Therefore, the cross validation technique is used for producing the estimation of error. The upper bound of error guides the design of the learning machine.

The learning capacity of a learning machine defines how much of the variance in the learning data can be learned by the machine. The learning capacity of a learning machine is roughly equal to the number of specimens (number of samples in the training set) that the learning machine can learn without generating any error. For a limited learning set size, this factor controls the estimation error. Small learning capacity causes bias error and large learning capacity causes variance error. An estimate of the learning capacity (or VC-dimension) of a learning machine can be measured for a classifier ([73]) and a regression estimator (see Section 6.4). This measurement is based on repeating the machine training-step numerous times and measuring the variance of error at each learning step. Therefore, this approach becomes impractical for cases where training a model is already difficult.

In order to find the optimum learning capacity of a learning machine, the designer can adjust a controllable learning factor such that, theoretically, by finding the bounds, or practically by measuring estimation error, the performance of the system is improved. In the case of polynomials, one of the controlling factors of the learning capacity of the machine is the number of terms of the function. For multi-layer neural networks the number of neurons of the hidden layer, for K-nearest neighbor the size of neighborhood (K), and similarly for radial basis functions the size of the neighborhood (the radius of the radial function) controls the learning capacity of the machine.

In the case of K-nearest neighbor and the radial basis function, the functions locally minimize the risk functional. But if K or radial function radius is kept constant, the overall estimation error can be used for finding the optimum K or radial function radius. If the estimation error is minimized at a level of the controllable learning factor, that level identifies the best structure of the learning system. For example, by increasing K in a K-nearest neighbor and using 10 fold cross validation for estimation error evaluation, a minimum point for the error curve can be found which defines the best learning capacity of the learning machine for the given data set.

An upper bound of the estimation error can be established based on the approximation error when the machine is being trained. The estimation error is a function of the number of training samples, learning estimation errors (error in learning the training set), and the learning capacity of the machine. The learning system learns the dependence between  $x$  and  $y$  for  $\lambda$  given samples by minimizing the following empirical risk functional:

$$Q(z_p, \alpha) = f(x_p, \alpha) - y_i \quad (75)$$

$$R_{emp}(\alpha) = \frac{1}{\lambda} \sum_{i=1}^{\lambda} Q(z_p, \alpha) \quad (76)$$

where the empirical data or the training data  $(z_i = (x_i, y_i))$  is the set of  $\lambda$  samples that was mentioned above:

$$z = \{z_1, \dots, z_\lambda\} \quad (77)$$

The empirical risk functional is an estimation (over a finite number of samples) of following risk functional:

$$R(\alpha) = \int Q(z, \alpha) dF(z) \quad (78)$$

Where  $F(z)$  is the cumulative distribution function.

If the estimated parameter set that results from minimizing the empirical risk functional ( $\alpha_\lambda$ ) is put to the test, it will produce  $R(\alpha_\lambda)$  estimation error. The best parameter set for the same learning machine could be obtained by minimizing equation (78). If the best parameter set is called  $\alpha_0$ , the following is the inherent error due to the limited number of samples.

$$\Delta(\alpha_\lambda) = R(\alpha_\lambda) - R(\alpha_0) \quad (79)$$

The estimated error at the training level, which is estimated by  $R_{\text{emp}}(\alpha)$ , should be adjusted to count for the limited number of samples. This difference was derived in [81] and is as follows.

$$R(\alpha_\lambda) < R_{\text{emp}}(\alpha) + B \varepsilon(\lambda) \left( 1 + \sqrt{1 + \frac{4R_{\text{emp}}(\alpha_\lambda)}{\varepsilon(\lambda)}} \right) \quad (80)$$

This bound holds with the probability of  $1-\eta$ . The two new parameters in inequality (80),  $B$  and  $\varepsilon(\lambda)$ , are defined as follows.

$$0 < Q(z, \alpha) < B \quad (81)$$

$$\varepsilon(\lambda) = 4 \frac{h(\ln \frac{2\lambda}{h} + 1) - \frac{1}{4} \ln(\eta)}{\lambda} \quad (82)$$

The learning capacity of the machine,  $h$ , is assumed to be known. Equations (80) to (82) define the learning error as a function of the VC-dimension ( $h$ ) and the number of specimens ( $\lambda$ ) in the database.

## **6.4. Extending the learning-capacity (VC-dimension) to regression estimators**

In this section the Vapnik-Chervonenkis dimension related to classifiers is generalized to cover regression problems. Since the VC-dimension is originally defined for classifiers, the equivalence of classifiers and regression estimators will be demonstrated first. Then, based on the VC-dimension of classifiers, the VC-dimension of regression estimators will be defined. Statistical learning theory can be used for estimating an upper bound for estimation variance.

The VC-dimension of classifiers was presented in [73]. The goal is to measure the dimensionality of a learning machine and derive an upper bound for the risk functional estimation error. In [73] a practical method was presented for measuring the corresponding dimension of a classifier. The analytical VC dimension for all learning models cannot yet be derived, thus there is a need to estimate it for any particular machine. The measurement is based on finding the classification error for a size of training samples. By varying the size of the training sample one can find the sample size that initiates the learning process.

In this section we will show that the same technique can be applied to regression estimators as well.

### **6.4.1. The equivalence of classifiers and regression estimators**

In this section it will be shown that under a general condition (on the output of the regression estimator) a classifier and a regression estimator can easily be transformed into each other.

#### **6.4.1.1. Learning problem statement**

A classification problem is defined on a set of  $(x_p, \omega_p)$  where  $x_i \in R^n$  is the input space and  $\omega_i \in \{0,1\}$  is the output space, which is the class of the given input. A classifier consists of a set of basis functions  $(f(.))$  with a set of parameters  $\alpha \in A$ .

The **classification problem** is the problem of learning the dependence between  $x_i$  and  $\omega_i$  for a given class of estimating functions  $f(x, \alpha)$ . In the empirical risk minimization principle the learning problem is equivalent to minimizing a risk functional on the given parameter space ( $\alpha \in \Lambda$ ) of the estimating functions ( $f(x, \alpha)$ ). The risk functional is a distance measure and is usually the mean squared error or a similar measure.

$$L^c(x, \alpha) = \frac{1}{n} \sum_{i=1}^n (f(x, \alpha) - \omega_i)^2 \quad (83)$$

The parameter set that produces the best classification fit is denoted by  $\alpha^*$  and is defined as follows:

$$\alpha^* = \min_{\alpha \in \Lambda} R_{\text{emp}}^c(\alpha) = \min_{\alpha \in \Lambda} \frac{1}{n} \sum_{i=1}^n (f(x, \alpha) - \omega_i)^2 \quad (84)$$

A **regression problem** is defined on a set of  $(x_i, y_i)$  where  $x_i \in R^n$  and  $y_i \in R$ . A regression estimator estimates the dependence of  $x_i$  on  $y_i$ . Similar to a classification problem, a regression problem minimizes a risk functional on the given set of estimating functions ( $f(x, \alpha)$ ). The risk functional is the same.

$$L^r(x, \alpha) = \frac{1}{n} \sum_{i=1}^n (f(x, \alpha) - y_i)^2 \quad (85)$$

The parameter set that produces the best regression fit is denoted by  $\alpha^*$  and is defined as follows:

$$\alpha^* = \min_{\alpha \in \Lambda} R_{\text{emp}}^r(\alpha) = \min_{\alpha \in \Lambda} \frac{1}{n} \sum_{i=1}^n (f(x, \alpha) - y_i)^2 \quad (86)$$

If one assumes that the same set of estimating functions is used for both regression and classification problems, then the question arises whether either problem could be stated by using the other. In other words, if one has a regression estimator, is there any way to use it for a classification problem? Or if one has a classifier, is there any way to use it for a regression problem? Solving this problem leads to the VC-dimension of a regression estimator if the VC-dimension of the corresponding classifier is known.

#### 6.4.1.2. Constructing a classifier by using a regression estimator

The regression estimator learns the *a posteriori* probability ( $p(y|x)$ ) at every point as will be shown in the following. For reconstructing the regression estimator, we need to construct a two-class classifier by using the regression estimator. In the case of a two-class classifier, the estimating function  $f(x, \alpha)$  can be used to define the *a posteriori* probability of  $p(C_1 | x)$  or  $p(C_2 | x)$  as follows. Establishing a classifier by using the *a posteriori* probability is the subject of sufficient statistics [79]:

$$C_1 = \{x_i | y_i < T\} \quad i=1, \dots, n \quad (87)$$

$$C_2 = \{x_i | y_i > T\} \quad i=1, \dots, n \quad (88)$$

Assuming  $D$  is the domain of  $x$  where the regression problem is defined, we can define the normalized regression estimator as follows:

$$f(x, \alpha) \cong \int_D p(x, y) dy = \int_{y < T} p(x, y) dy + \int_{y > T} p(x, y) dy \quad (89)$$

This leads to the following necessary and sufficient conditions:

$$f(x, \alpha) > T \Leftrightarrow \int_{y < T} p(x, y) dy < \int_{y > T} p(x, y) dy \quad (90)$$

and

$$f(x, \alpha) < T \Leftrightarrow \int_{y < T} p(x, y) dy > \int_{y > T} p(x, y) dy \quad (91)$$

Then the *a posteriori* probability of a class can be derived as follows:

$$p(C_1 | x) = p(y < T) = \int_{y < T} p(x, y) dy \quad (92)$$

and

$$p(C_2 | x) = p(y > T) = \int_{y > T} p(x, y) dy \quad (93)$$

Therefore, assuming the two conditions of (90) and (91) are valid, the two classes are defined as follows:

$$C_1 = \{x \mid f(x, \alpha) < T\} \quad (94)$$

and

$$C_2 = \{x \mid f(x, \alpha) > T\} \quad (95)$$

Note that in the above there is no assumption about the type of the estimation problem (classification or regression). Therefore, the learning problem is reduced to training the regression estimator to learn the dependence (either regression or classification). The classification problem then results from setting a decision level for distinguishing the two classes.

The two conditions of (90) and (91) can be questions when  $f(x, \alpha) \cong T$ . Therefore, in a classification problem, the decision level can be adjusted so that minimum classification error results. The nominal decision level for a classification problem is naturally  $T_{\text{nom}} = (A + B) / 2$  where  $A$  and  $B$  are the outputs related to either class.

#### 6.4.1.3. Constructing a regression estimator by using a classifier

In the previous section it was shown that a classifier could be constructed by using the regression estimator. A regression function can be constructed by defining the output of a classifier so that it satisfies the following condition:

$$\forall x_i, x_j, T \quad (y_i < T < y_j) \Rightarrow (\omega_i < \omega_j) \quad (96)$$

The threshold level  $T$  in equation (96) is an arbitrary variable that belongs to the range of the regression estimator ( $\min(y) < T < \max(y)$ ). It will be shown that the regression estimator can be constructed (as closely as desired) by using the set of classifiers defined by equation (96).

The regression estimator is developed on a training set as follows:

$$S_R = \{(x_i, y_i), i=1, \dots, n\} \quad (97)$$



The estimator is developed by using equations (85) and (86). A group of classification sets can be defined as follows. Assume the specimens are ordered such that  $y_i < y_{i+1}$ . By using a decision level the output of a classifier can be defined as follows:

$$\forall x_i, x_j, T \quad (y_i < T < y_j) \Rightarrow (\omega_i < \omega_j) \quad (98)$$

Then a set of input/output can be defined that represents a class, as follows:

$$S_{ci} = \{(x_j, \omega_j), j = 1, \dots, n, \omega_1, \dots, \omega_i = 1, \omega_{i+1}, \dots, \omega_n = 0\} \quad (99)$$

Then a group of  $n$  classifiers can be trained to learn the dependence in each  $S_{ci}$  by using equations (83) and (84). Each classifier ( $S_{ci}$ ) is also equivalent to a classifier of the previous section with the decision level of  $y_i < T_i < y_{i+1}$ .

Assuming that the range of the regression problem is limited and is divided into equally separated threshold levels, an upper bound for equation (85) can be reached as follows:

$$A = \min(y), B = \max(y) \quad (100)$$

$$\Delta_i = |T_i - T_{i+1}| = (B - A)/(n + 1) = \Delta \quad (101)$$

The output of the regression estimator can be approximated by the outputs of all the classifiers of  $S_{ci}$  ( $i = 1, \dots, n$ ).

$$\hat{y}_{ri} = T^* \cdot y_{ci} + \Delta/2 \quad (T^* = \max_i(T_i = 1), i = 1, \dots, n) \quad (102)$$

Where  $y_{ci}$  is the output of the classifier  $S_{ci}$ . The minimum mean squared error in this case will be as follows:

$$R_{emp}^{cr}(\alpha^*) = \min_{\alpha} L(x, \alpha) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{ri})^2 \quad (103)$$

The error in the regression estimator can be defined as follows:

$$\max(|\hat{y}_{ri} - f(x, \alpha)|) \leq \Delta/2 \quad (104)$$

The output of the regression model and the output of the regression estimator are related as follows:

$$\hat{y}_{ri} = f(x, \alpha) + \varepsilon_i \quad (105)$$

Where  $\varepsilon$  is the error (noise) in the estimation of the output of the regression estimator. This noise is uniformly distributed and is independent of  $y_i$  and  $f(x, \alpha)$  ([80]):

$$p(\varepsilon) = \begin{cases} \frac{1}{\Delta} & |\varepsilon| \leq \Delta/2 \\ 0 & |\varepsilon| > \Delta/2 \end{cases} \quad (106)$$

Properties of the quantization error follow:

$$E(\varepsilon) = 0 \quad (107)$$

and

$$\sigma^2 = \text{var}(\varepsilon) = \int_{-\Delta/2}^{\Delta/2} \varepsilon^2 \frac{1}{\Delta} d\varepsilon \quad (108)$$

$$\Rightarrow \sigma^2 = \frac{1}{12} \Delta^2 = \frac{1}{12} \frac{(B - A)^2}{(m + 1)^2}$$

where  $m$  is the number of classifiers in the model. The variance of noise approaches zero as the number of classifiers approaches infinity.

In the following we will show that  $R_{\text{emp}}^{\sigma}(\alpha)$  is an upper bound for  $R_{\text{emp}}^r(\alpha)$ :

$$\begin{aligned} R_{\text{emp}}^{\sigma}(\alpha) &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{ri})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (y_i - f(x_p, \alpha) - \varepsilon_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (y_i - f(x_p, \alpha))^2 - \frac{2}{n} \sum_{i=1}^n (y_i - f(x_p, \alpha)) \varepsilon_i + \frac{1}{n} \sum_{i=1}^n (\varepsilon_i)^2 \end{aligned} \quad (109)$$

$$= R_{emp}^{cr}(\alpha) - \frac{2}{n} \sum_{i=1}^n (y_i - f(x_i, \alpha)) \cdot \varepsilon_i + \frac{1}{n} \sum_{i=1}^n (\varepsilon_i)^2$$

In the two extreme cases where one of  $R_{emp}^{cr}(\alpha)$  or  $R_{emp}^r(\alpha)$  is zero, the other one will be equal to  $\sigma^2$ .

Therefore the distance of the two empirical cost functionals is as follows:

$$|R_{emp}^{cr}(\alpha) - R_{emp}^r(\alpha)| = | - \frac{2}{n} \sum_{i=1}^n (y_i - f(x_i, \alpha)) \cdot \varepsilon_i + \frac{1}{n} \sum_{i=1}^n (\varepsilon_i)^2 | \quad (110)$$

$$\leq | \frac{2}{n} \sum_{i=1}^n (y_i - f(x_i, \alpha)) \cdot \varepsilon_i | + | \frac{1}{n} \sum_{i=1}^n (\varepsilon_i)^2 |$$

$$\cong | \frac{2}{n} \sum_{i=1}^n (y_i - f(x_i, \alpha)) | \cdot | \frac{2}{n} \sum_{i=1}^n \varepsilon_i | + | \frac{1}{n} \sum_{i=1}^n (\varepsilon_i)^2 |$$

Loosely accepting the following equation:

$$\sigma^2 \cong \frac{1}{n} \sum_{i=1}^n (\varepsilon_i)^2 \quad (111)$$

we will have the following bound:

$$\sup_{\alpha} \{ |R^{cr}(\alpha) - R^r(\alpha)| \} \leq \sigma^2 \quad (112)$$

Similarly, the risk functionals are related as follows:

$$R^{cr}(\alpha) = \int_A^B (y - f(x, \alpha) - \varepsilon)^2 p(y) dy \quad (113)$$

$$= \int_A^B (y - f(x, \alpha))^2 p(y) dy - 2 \int_A^B (y - f(x, \alpha)) \cdot \varepsilon p(y) dy + \int_A^B \varepsilon^2 p(y) dy$$

Which results in the distance of the two risk functionals, as follows:

$$\sup_{\alpha} \{ |R^{cr}(\alpha) - R^r(\alpha)| \} \leq \sigma^2 \quad (114)$$

Where  $\sigma^2$  is the variance of  $\varepsilon$ . This shows that  $R_{emp}^{cr}(\alpha)$  is an upper bound for  $R_{emp}^r(\alpha)$  and  $R^{cr}(\alpha)$  is an upper bound for  $R^r(\alpha)$ . If the number of classifiers approaches infinity the bound  $R^{cr}(\alpha)$  approaches  $R^r(\alpha)$ .

#### 6.4.2. The VC-dimension of a regression estimator

In the above the construction of each learning system by using the other has been explained. In order to measure the VC-dimension of a given regression estimator we will construct a classifier from it and reconstruct the regression estimator from the classifiers. In the process, the regression estimator is represented by a finite number of classifiers whose VC-dimension can be measured ([73]). Then by equation (114) the upper bound for the empirical risk functional of the regression estimator will be derived.

In order to find the VC-dimension of the regression estimator we should derive an upper bound for  $R^r(x, \alpha)$ . It is obvious that if the maximum of VC-dimension of the regression estimator in the range  $[A, B]$  is known we will have the following:

$$\sup_{\alpha} \{|R^r(\alpha) - R_{emp}^{cr}(\alpha)|\} \leq \phi\left(\frac{h_{\max}}{\lambda}\right) \quad (115)$$

Therefore, with  $\sigma$  tolerance, an upper bound for the risk functional can be obtained by using  $m$  classifiers. The maximum estimation error for the regression estimator can be derived as follows:

$$\sup_{\alpha} \{|R^{cr}(\alpha) - R_{emp}^{cr}(\alpha)|\} + \sup_{\alpha} \{|R_{emp}^{cr}(\alpha) - R_{emp}^r(\alpha)|\} \quad (116)$$

$$+ \sup_{\alpha} \{|R^{cr}(\alpha) - R^r(\alpha)|\} \leq \phi\left(\frac{h_{\max}}{\lambda}\right) + 2\sigma^2$$

$$\sup_{\alpha} \{|R^{cr}(\alpha) - R_{emp}^{cr}(\alpha)|\} + \sup_{\alpha} \{|R_{emp}^{cr}(\alpha) - R_{emp}^r(\alpha)|\} \quad (117)$$

$$+ \sup_{\alpha} \{|R^{cr}(\alpha) - R^r(\alpha)|\} \leq \sup_{\alpha} \{|R^{cr}(\alpha) - R_{emp}^{cr}(\alpha) + R_{emp}^{cr}(\alpha) - R_{emp}^r(\alpha) + R^r(\alpha) - R^{cr}(\alpha)|\}$$

Therefore, the upper bound for maximum risk functional error can be written as follows:

$$\sup_{\alpha} \{|R(\alpha) - R_{emp}(\alpha)|\} \leq \phi\left(\frac{h_{\max}}{\lambda}\right) + 2 \cdot \sigma^2 \quad (118)$$

As the number of classifiers increases,  $\sigma^2$  approaches zero and  $h_{\max}$  approaches to the VC-dimension of the regression estimator ([81]).

#### 6.4.3. Corollary

An upper bound not related to the assumed distribution of  $\varepsilon$  can be obtained by placing  $\Delta^2/4$  instead of  $\sigma^2$ :

$$\sup_{\alpha} \{|R(\alpha) - R_{emp}(\alpha)|\} \leq \phi\left(\frac{h_{\max}}{\lambda}\right) + \frac{\Delta^2}{2} \quad (119)$$

#### 6.4.4. Measuring the VC-dimension at a point

In order to measure the VC dimension of the regression estimator for the given set of data, the regression data set is transformed to a classification problem as follows:

$$(x_i, \omega_i) \quad (120)$$

where, the classification output ( $\omega_i$ ) is defined as follows:

$$\begin{cases} \omega_i = 1, & \text{if } y_i \geq T \\ \omega_i = 0, & \text{if } y_i < T \end{cases} \quad (121)$$

In order to measure the VC dimension of the regression estimator, the variance in the estimation for a data set with variable set size should be measured. Therefore, a number of different test set sizes ( $\lambda$ ) are selected. For each set size, 20 different data sets (from the original data set given in equation (120)) are produced:

$$Z_i^{2\lambda} = \{(x_1^i, \omega_1^i), (x_2^i, \omega_2^i), \dots, (x_\lambda^i, \omega_\lambda^i), \dots, (x_{\lambda+1}^i, \omega_{\lambda+1}^i), \dots, (x_{2\lambda}^i, \omega_{2\lambda}^i)\} \quad (122)$$

The set size ( $\lambda$ ) is between 1 and 100, and 20 equally spaced numbers are selected.

The maximum variation in estimated empirical risk functional must be measured. The empirical risk functional of a set of size  $\lambda$  is defined as follows:

$$v_{\lambda}(\alpha) = \frac{1}{\lambda} \sum_{i=1}^{\lambda} |\omega_i - f(x_i, \alpha)| \quad (123)$$

The maximum variation in the measured empirical risk functional is defined as follows:

$$\xi(Z_i^{2\lambda}) = \sup_{\alpha \in \Lambda} [v_1^{\lambda}(Z_i^{2\lambda}, \alpha) - v_2^{\lambda}(Z_i^{2\lambda}, \alpha)] \quad (124)$$

The expected value of  $\xi(Z_i^{2\lambda})$  is a monotonically decreasing function of the VC dimension of the learning machine:

$$E\{\sup_{\alpha \in \Lambda} [v_1^{\lambda}(Z_i^{2\lambda}, \alpha) - v_2^{\lambda}(Z_i^{2\lambda}, \alpha)]\} \approx \Phi\left(\frac{\lambda}{h}\right) \quad (125)$$

where  $h$  is the VC dimension of the learning machine. The expected value is approximated by an average over  $N$  ( $=20$ ) test sets:

$$\xi(\lambda) = \frac{1}{N} \sum_{i=1}^N \xi(Z_i^{2\lambda}) \quad (126)$$

The VC dimension of the learning machine ( $f(x, \alpha)$ ,  $\alpha \in \Lambda$ ) is the integer parameter that makes the best fit between  $\Phi\left(\frac{\lambda}{h}\right)$  and  $\xi(\lambda_i)$ :

$$h^* = \operatorname{argmin}_h \sum_{i=1}^N \left[ \xi(\lambda_i) - \Phi\left(\frac{\lambda_i}{h}\right) \right]^2 \quad (127)$$

The largest deviation in the empirical risk is estimated by using the opposite class of half of the members of  $Z^{2\lambda}$  as follows:

$$\xi(Z^{2\lambda}) = \frac{1}{\lambda} \sum_{i=1}^{\lambda} [\varpi_i - f(x_i, \alpha)]^2 + \frac{1}{\lambda} \sum_{i=\lambda+1}^{2\lambda} [\omega_i - f(x_i, \alpha)]^2 \quad (128)$$

where the opposite class ( $\varpi$ ) is defined as follows:

$$\begin{cases} \varpi_i = 0 & \text{if } \omega_i = 1 \\ \varpi_i = 1 & \text{if } \omega_i = 0 \end{cases} \quad (129)$$

The deviation function ( $\Phi(\frac{\lambda}{h})$ ) has the following simplified form that was suggested in [73]:

$$\begin{cases} \Phi(\frac{\lambda}{h}) = 0 & \text{if } \frac{\lambda}{h} \leq 0.5 \\ \Phi(\frac{\lambda}{h}) = d \frac{\ln(2\lambda/h) + 1}{(\lambda/h) + d - 0.5} & \text{if } \frac{\lambda}{h} > 0.5 \end{cases} \quad (130)$$

The constant  $d$  is empirically derived ([73]) and is equal to 0.39.

### *Example*

In this example the VC dimension of a regression estimator will be measured. The original dependence between input and output of the samples is a norm function and the estimator function is a linear regression estimator. The regression problem will be transformed to a classification problem by using a threshold level and the VC dimension of the learning machine will be measured by following the steps in [73]. The VC dimension of a polynomial is known to be equal to the number of its coefficients. It will be shown that the measured VC dimension is very close to the known VC dimension of the linear regression estimator.

The function to be studied is the norm function of ten independent variables as follows:

$$y = \sum_{i=1}^N x_i^2 \quad (131)$$

Where, the independent variables ( $x_i$ ) are independent random variables in [0,10]. The number of variables ( $N$ ) is changed and the VC dimension is measured. A linear regression estimator is used to learn the input-output relation:

$$y = c_0 + \sum_{i=1}^N c_i x_i^2 \quad (132)$$

The VC dimension of this linear regression estimator is equal to  $N + 1$ . A sample set of size ten thousand is produced, by generating one thousand random points for each variable ( $N$ ) and the related output from equation (131):

$$(x_i, y_i) \quad i = 1, \dots, 1000 \quad (133)$$

Where the input variable is a vector of size 10.

The threshold level is equal to the average of the output:

$$T = \frac{1}{\lambda} \sum_{i=1}^{\lambda} y_i \quad (134)$$

For solving equation (127), it was assumed that  $h$  is within the range of [1,100] and a "look-up table" was generated for  $\Phi(\frac{\lambda}{h})$ . The solution to equation (127) was found by searching for the best  $h$  that minimized the error. Table 4 shows that the measured VC dimension is very close to the number of coefficients of the linear regression estimator. Changing the threshold level ( $T$ ) does not affect the measured VC dimension.

**Table 4: The measured VC dimension of a linear regression estimator.**

Number of variables	Number of coefficients	Measured VC dimension
5	6	6
10	11	10
15	16	16
20	21	21
25	26	25
30	31	31
35	36	37
40	41	42
60	61	61



## The VC-dimension measurement algorithm

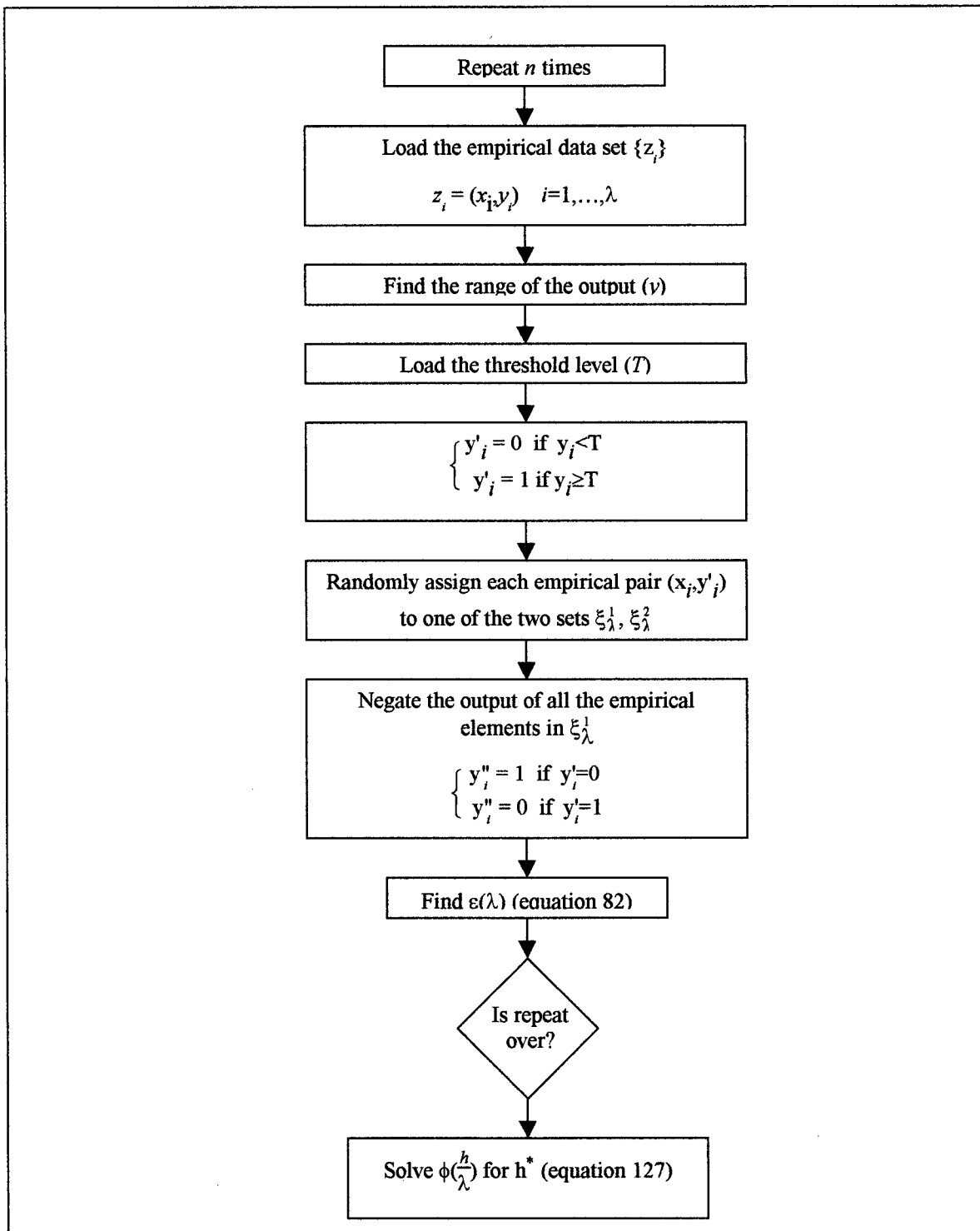


Figure 6.3 The block diagram for estimating the VC-dimension

This routine works especially well when  $2h^* \leq \lambda \leq 10h^*$ . If an initial value for  $h^*$  is known, the range of  $\lambda$  can be defined readily. The routine works well when an assumption for  $h^*$  is made and the range of  $\lambda$  is obtained. Then this range of  $\lambda$  is used to find a new  $h^*$  and this loop continues until the sequence of  $h^*$  approaches to a constant value.

### Significance of VC-dimension calculation

In order to get the theoretical distribution of the error of the risk functional, the VC-dimension of a regression estimator is needed. Obtaining an estimate of the risk functional ( $R(\alpha)$ ) requires having a sufficient number of specimens. This condition is not always practically possible. The VC-dimension of a regression estimator (either measured or estimated) provides a means for obtaining the error distribution of the (regression) estimator. This error distribution is used in every step of ASEC for finding the size of the selected subset.

For grading purposes, the practical significance of the VC-dimension is that one can estimate the grading improvement for any increase in the number of database specimens. Assuming that the VC-dimension of the regression estimator ( $h$ ) is known, or is measured, at a threshold level (or grading level) and assuming that the number of specimens increases from  $\lambda_1$  to  $\lambda_2$ , one can calculate the expected improvement in grading as follows:

$$R_1(\alpha) = R_{emp}^1(\alpha) + \varepsilon\left(\frac{h}{\lambda_1}\right) \quad (135)$$

$$R_2(\alpha) = R_{emp}^2(\alpha) + \varepsilon\left(\frac{h}{\lambda_2}\right) \quad (136)$$

For the same learning problem and the same learning machine, the empirical risk does not change much by changing the number of the specimens, thus:

$$R_{emp}^1(\alpha) \cong R_{emp}^2(\alpha) \quad (137)$$

Therefore the improvement in the risk functional of the regression problem is as follows.

$$R_1(\alpha) - R_2(\alpha) \cong \varepsilon\left(\frac{h}{\lambda_2}\right) - \varepsilon\left(\frac{h}{\lambda_1}\right) \quad (138)$$

By using equations 74 and 138, one can obtain the improvement in the coefficient of determination ( $r^2$ ).

## 6.5. Strength estimation by using existing methods

Each of the learning machines used for strength estimation is explained in the following.

### 6.5.1. Linear regression

As a parametric model, a linear regression is the first model to be used for checking the dependence between features,  $x$ , and the output,  $y$ . This model generally fits a polynomial to the given data set. The approximating function is shown in equation (139):

$$f(x, w) = \sum_{i=1}^N w_i x_i \quad (139)$$

where  $N$  is equal to the number of features and the complexity of the approximating function. Therefore, one can test the effect of the complexity of the approximating function on the estimation, at the lower end of the complexity, by using this model. The higher degree terms in equation (139) can be introduced as new features; therefore, an estimation of the parameters of  $f(x, w)$  can simply be formulated as follows:

$$f(x, w) = \sum_{i=1}^N w_i x_i = w_1 x_1 + w_2 x_2 + \dots + w_n x_n = y_i \quad (140)$$

or,

$$Xw = Y \quad (141)$$

$$w = X^{-1}Y \quad (142)$$

The capital letters denote the vectors of the same variables that lower case letters represent.  $X^{-1}$  is the pseudo-inverse of  $X$ , which is  $(X^T X)^{-1} X^T$ . The potential problem with

the polynomial fit is that it generates a higher dimensional feature space than already exists. For example, for the strength estimation of wood, about 1000 specimens are collected and the number of features is about 100. A third degree polynomial creates 200 new features (the original features to the power of two and three). Multiplying the number of features in the polynomial learning method rapidly adds to the problem of high dimensionality of the feature space. Also, since the form of the function is defined at the beginning, the bias error has a lower bound ([66]).

### 6.5.2. K-Nearest Neighbor and radial basis functions

These two networks are local networks. As mentioned before, the local networks do well when the samples are not dense. These methods generate a prediction at a given point of the input space without generating a hyper surface in the input space.

These two methods are adaptive networks, called kernel-based networks, in the sense that  $f(x)$  is a local function ([66],[74]). The radial basis network can be phrased as follows:

$$y = \sum_{i=1}^N w_i h_i(\|x - x_i\|) \quad (143)$$

Where  $h_i(\cdot)$  is usually a bell-shaped function ( $h_i(x) = e^{-x^2}$ ).  $\|x - x_i\|$  is the Euclidean distance of the input  $x$  from the location of the radial basis function  $x_i$ . The complexity of the approximating function (see page 125) is inversely proportional to the radius of  $h(\cdot)$ . Since the radius is the same for all  $h_i(\cdot)$  a line search can find the optimum radius. Therefore, for a given radius,  $w_i$  can be found by an optimization technique (similar to a training algorithm for a neural network) and the risk functional value is computed. An optimization technique can find the radius that minimizes the risk functional.

Similarly a K-nearest neighbor uses the average of  $y_i$  for K neighbors of the point. The larger K is the smoother the function will be, therefore, the complexity of the function is inversely proportional to K. The choice of the distance is important in this case.

In order to construct the approximating function, the risk functional for the neighborhood the kernel function covers should be minimized. An alternative would be to use similar kernel function over the entire space of the input variables. Therefore, the summation of

all the local risk functionals is the empirical risk functional. The optimal parameter set is the one that minimizes the empirical risk functional.

In a K-nearest neighbor machine, the size of neighbors controls the learning capacity of the machine. Therefore, by increasing the number of neighbors and finding the neighborhood size where the maximum estimation accuracy occurs, the best structure of the machine can be found. It was seen that the neighborhood of size 5 results in the most accurate estimation from 10 fold cross validation ([72]).

In a radial basis function machine the diameter of the radial function controls the learning capacity of the learning machine. By keeping the radial function diameter constant, one can maximize the overall estimation accuracy and find the right structure of the learning machine. By using 10 fold cross validation for accuracy approximation and by doing a search on radial function diameter, the best diameter was obtained.

### 6.5.3. Multi-layer neural networks

Multi-layer neural networks are semi-parametric adaptive networks. These networks usually consist of three fully connected layers. The first layer is the input (features). The second layer consists of the neurons and the third layer is the output layer, which is a linear neuron that sums the scaled output of all the neurons of the second layer. These networks are robust with respect to the training set and perform very well when dense samples are available, even for high dimensional inputs ([66],[75]). This network is a universal approximator; that is, given enough neurons the network can approximate the continuous function that appears in applications ([76],[77]).

Multi-layer neural networks with different neuron activation functions are trained by using back propagation. The network can be modeled as follows:

$$y = \sum_{i=1}^M h_i(g(x)) \quad (144)$$

$$g(x) = b + \sum_{i=1}^N a_i x_i \quad (145)$$

where  $M$  is the number of neurons,  $N$  is the number of features,  $b$  is the bias, and  $a_i$  are the scale factors of each feature.  $h_i(.)$  is the activation function of the related neuron.  $g(.)$

is a linear function that generates the input to the neuron. A network consists of the necessary number of neurons at its middle layer and a linear neuron ( $h_i(x) = x$ ) as the output of the network. The typical forms of  $h_i(\cdot)$  are linear, sigmoid, or radial basis. The design of the network consists of finding the right number of neurons and activation functions.

The common neurons are sigmoid functions, the logistic function and the arc-tangent function, as defined by equations (146) and (147) in the following. The choice of either one of these functions does not affect the result much ([72]):

$$h(x) = \frac{1}{1+e^{-x}} \quad (146)$$

$$h(x) = \frac{1}{\pi} \tan^{-1}(x) + \frac{1}{2} \quad (147)$$

Networks of more than three layers are also being used. The extra layer(s) adds to the complexity of the network. By limiting the number of neurons and limiting the norm of the parameter vector,  $\|w\|$ , one can control the complexity of the network. The first usually is choosing the structure of the network, and the second is choosing the early stopping rule.

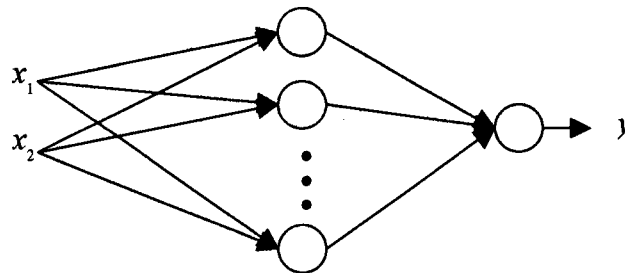


Figure 6.4 A multi-layer neural network.

In this case, only a three-layer network is used. The hidden layer consists of symmetric sigmoidal neurons and the output layer is a linear neuron. The best accuracy was obtained by 5 hidden neurons. Adjusting the accuracy goal at the training level did not improve the results.

#### 6.5.4. SPORE

This method of function approximation was developed for a set of problems with the same conditions as the problem at hand (Ph.D. thesis of Grudic, [21]). This method is a nonparametric adaptive network that works well even when the number of features is large. There are different versions of the SPORE algorithm but what is applied in this research is SPORE-1, as is explained in the following.

The network consists of a cascade of blocks of two variable polynomials with adjustable degrees. The output of each block is fed to the next block until the end of the cascade. The output function is a linear combination of all the blocks (Figure 6.5).

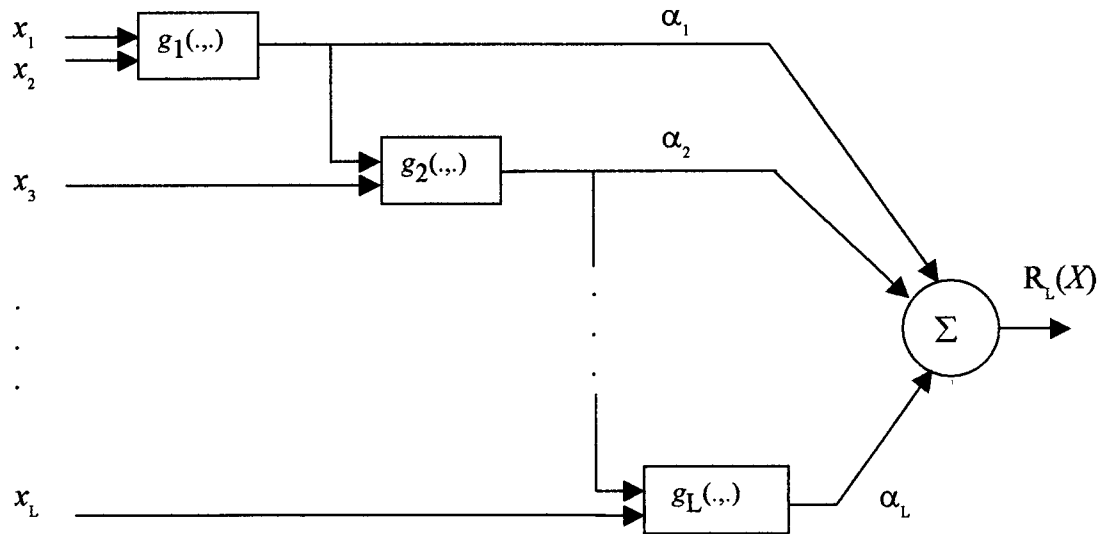


Figure 6.5 The cascade form of SPORE-1.

Each polynomial ( $g(.,.)$ ) is defined as follows:

$$g(u,v) = \sum_{i,j \in \{0, 1, \dots\}}^{i+j \leq K} a_{ij} u^i v^j \quad (148)$$

The network is developed in block sections, a few blocks at a time, and progresses until all the variables are included. The number of blocks in a block section is more than a constant number 'depth1'. The order of the input variables is random. The parameters of each block in the block section is estimated for minimizing the following cost function.

$$\sum_{q \in \Gamma_T^B} (g(.,.) - y_q)^2 \quad (149)$$

where the set  $\Gamma_T^B$  is a bootstrap sample from the training set. In Figure 6.5, the output scale of each block,  $\alpha_p$ , is proportional to the inverse of the mean square error of the block. When the change in the cascade error in the last 'depth1+1' blocks is less than a constant factor  $\varepsilon$  the construction of the new blocks stops. Then a search in the block set takes place to find a subset of the blocks that minimizes the risk functional (equation (149)) more than the block set does. If such a subset exists, the blocks that are in the block set but not in the subset, are removed and variables are freed for later approximation improvement. The same process continues by using a block set of size 'depth2' from the remainder of the variables. This second block set is constructed to learn the error of the first block set. It can be seen that  $\kappa$  defines the complexity of the network and the learning algorithm is a comprehensive search among different combinations of the variables.

In this model the learning capacity of the machine has to be controlled by adjusting its parameters. The model finds the best structure by trial and error. Therefore, no notion of learning capacity is used for this machine.

## 6.6. Alternating Space Expansion and Contraction (ASEC)

As was implied through this thesis and in many exploratory estimation problems, not only the input-output dependence ( $f(x,w)$ ) is not known but also the representative feature set ( $x$ ) could be arbitrarily defined. In ASEC method both the parameters of the estimating function ( $f(x,w)$ ) and the set of input features ( $x$ ) are optimized in order to improve the estimation accuracy. Two iterations of ASEC are shown in Figure 6.6.



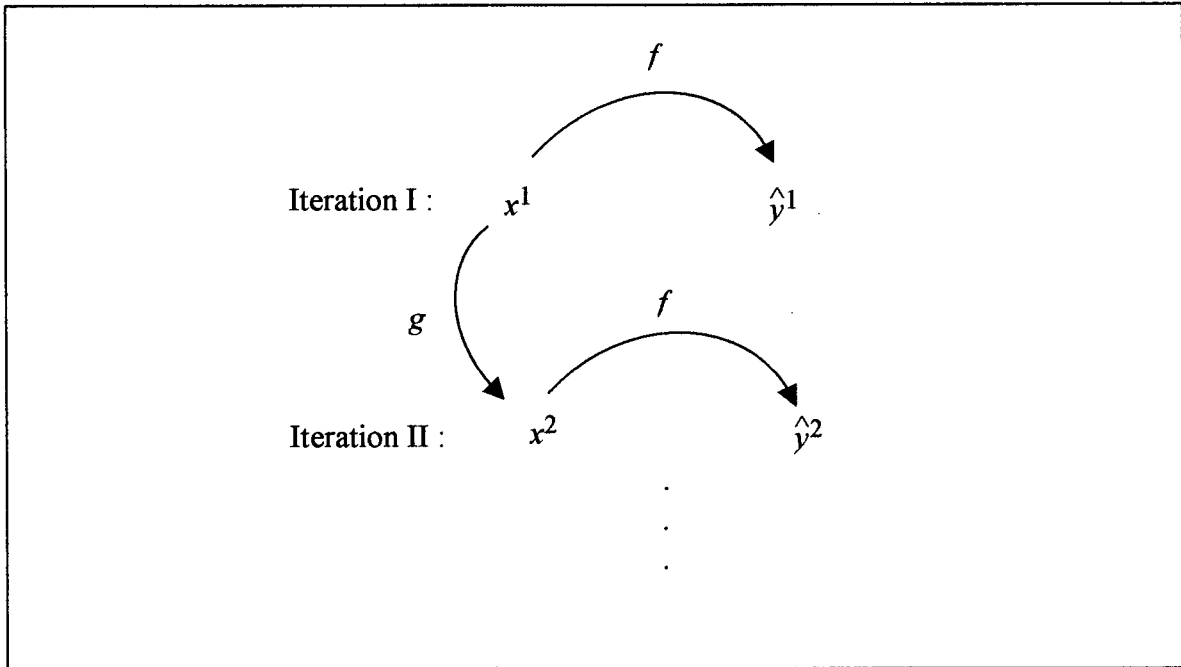


Figure 6.6 Two iterations of ASEC

In a field like the strength of material, the *a priori* knowledge usually exists in the form of relationships between strength affecting factors (represented by  $g$  in Figure 6.6), and between them and the strength of material (represented by  $f$  in Figure 6.6).  $f$  and  $g$  do not have to be the same in different iterations.

For producing the new feature set ( $x^{j+1}$ ), the existing feature set ( $x^j$ ) is transformed by the nonlinear transformation ( $g$ ) to a large set of features.  $x^j$  is a small subset of the generated feature set that produces the smallest estimation error.

If the set of transformations ( $g$ ) is similar to a universal approximator, its convergence could be guaranteed for a large number of specimens. A good example of a universal approximator is a generalized polynomial, as follows.

$$\sum C_{i1} x_i + \sum C_{ij} x_i x_j + \sum C_{i2} x_i^2 + \dots \quad (150)$$

If the individual functions in  $g$ , are exponential transformations of the given feature set ( $x^j$ ), then the terms of equation (150) could be used to check if all the necessary terms are included in  $g$ .

Although the polynomial model is a general approximator, it is not the final model when a small number of measured specimens are available. For example, for estimating  $y = \tan(x)$  with a given number of samples and a given accuracy  $r^2$ , a polynomial with a relatively large number of terms should be generated. The accuracy of the estimation depends inversely on the number of coefficients because of the limited number of specimens. Therefore, a range of nonlinear dependencies that can reasonably exist should be included in the expansion guide.

The problem of wood features to tensile strength was modeled by using the following. An approximation model is selected. In our case the model is as follows:

$$y_i = \sum c_i x_i \quad (151)$$

The feature transformation ( $g(x)$ ) was done by two sets of nonlinear transformations: a set of such nonlinear transformations of a single feature; and a set of cross-multiplication of features. The nonlinear transformations of a single feature (see Appendix) is shown in the following:

$$\left\{ \frac{1}{x}, |x|, \text{sign}(x)\sqrt{|x|}, x^2 \right\} \quad (152)$$

The cross multiplication of features investigate the inter-relation of the features and is shown in the following:

$$\{x_i x_j\} \quad (153)$$

The contraction method is a selection method that identifies a subset of the given feature set so that the best estimation (minimum error) can be obtained. The Gram-Schmidt method of space spanning is used for this purpose, as follows.

1. The closest vector to the output is selected.
2. All other features (including the output) are replaced by their orthogonal projection on the first feature.

Steps 1 and 2 are repeated to generate the subset of given size.

In order to include the variation in the feature set, a group of training feature sets (e.g. 10) called  $T_p$  is generated in order to demonstrate the direction of feature vectors and their variations. For each training feature set ( $T$ ) the minimum subset is obtained and the table of consequent features is generated.

**Table 5: The parallel feature sets.**

	iteration 1	iteration 2	iteration 3	...
$T_1$	$f_{11}$	$f_{12}$	$f_{13}$	...
$T_2$	$f_{21}$	$f_{22}$	$f_{23}$	...
$T_3$	$f_{31}$	$f_{32}$	$f_{33}$	...

The selected subset (based on Table 5) is then as follows:

$$\{\{f_{11}, f_{21}, f_{31}\}, \{f_{12}, f_{22}, f_{32}\}, \{f_{13}, f_{23}, f_{33}\}\} \quad (154)$$

The features are then eliminated from the set one by one such that a subset with highest accuracy is obtained. This step eliminates the features whose contribution to the accuracy of the model can be compensated for by the other features.

Based on the given contracted feature set, the model of equation (151) was developed and the accuracy of the estimated and is compared to that of the previous model (accuracy of model 0 is equal to 0).

If improvement results another iteration will be attempted. If no improvement results, the process will be terminated.

The number of iterations in Gram-Schmidt selection can be obtained from statistical learning theory. In this theory there is an inherent estimation error associated with the generalization ability of the learning system. The generalization ability (measured by learning capacity or VC-dimension) for a polynomial can be estimated by the number of coefficients, which is the number of Gram-Schmidt iterations.

A starting point for the number of iterations (for example 10% of the training set size) is considered and the subset is generated. The learning accuracy of the estimator can be obtained as shown in the following Figure:

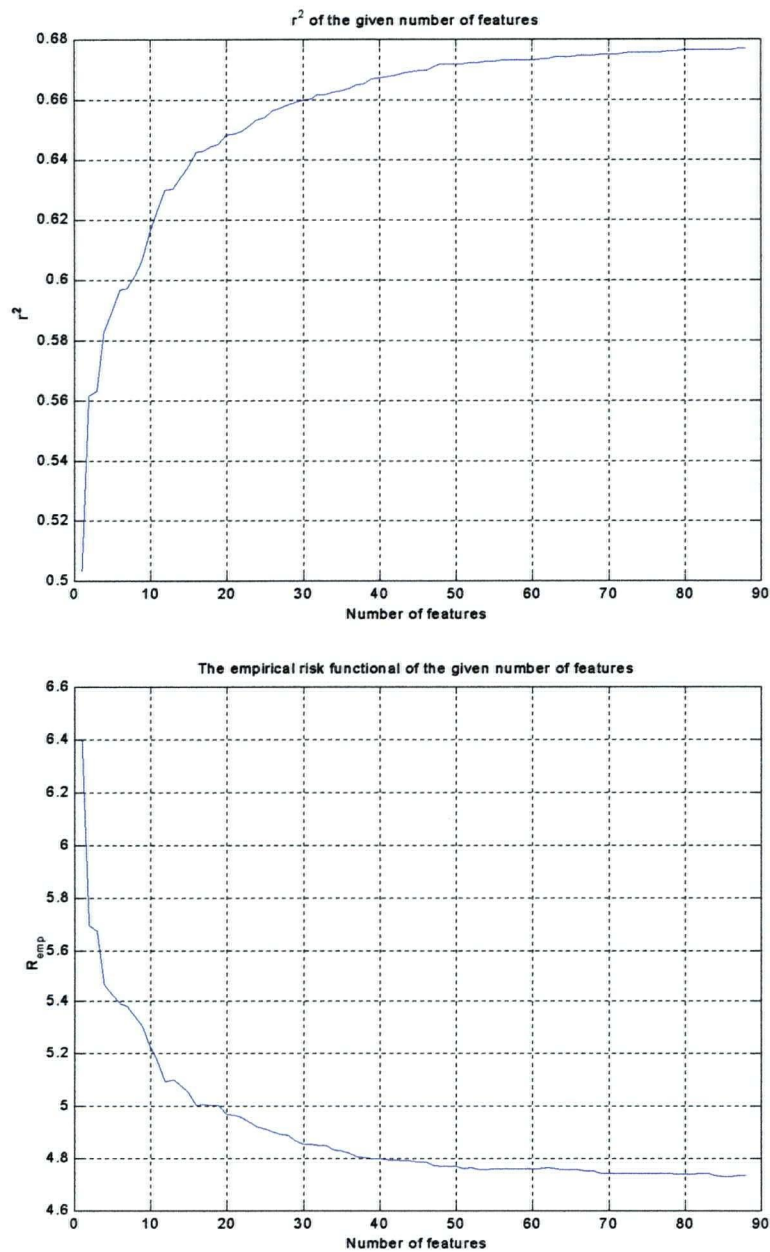


Figure 6.7 Estimation accuracy from the contracted space (top) and the empirical risk (bottom).

The overall (lower bound) estimation accuracy is obtained by subtracting the error from the estimation accuracy.

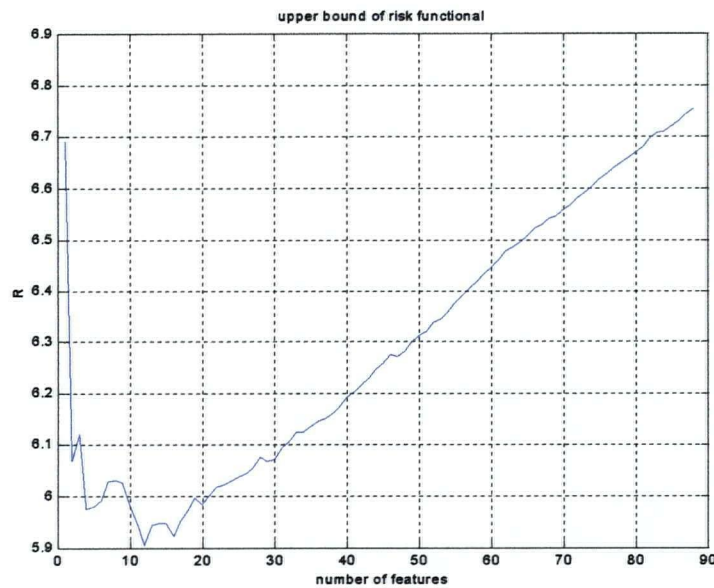


Figure 6.8 The curve of the upper bound of the risk functional for the given number of features.

The optimum feature set size 'T' is the basis number in the contraction step (step 3). 'T' is a minimum number of iterations because the statistical learning theory provides a conservative measure of performance. The actual number of iterations is a multiple of 'T' (between 3 and 6 times) and a more realistic iteration number can later be obtained by using the cross validation technique.

The cross validation technique works as follows. Each training set is again divided into a group of cross validation sets and will be called second sets.

The second training sets and the second testing sets are used to find the realistic 'T'. Each second set generates one trajectory of features for the higher level set. Different trajectories are different rows of the feature table in 'Table 1'. From each trajectory a diagram of estimation accuracy is obtained whose maximum is at the iteration number. The average of these profiles is used for finding the iteration number. By doing so, the iteration number, which is one of the learning parameters, is produced without the effect

of the test set. Once this process is repeated for all 'n' scenarios, a set of estimation accuracy and an optimum iteration number will be obtained:

$$\{R_1, I_1\}, \{R_2, I_2\}, \{R_3, I_3\}, \dots, \{R_n, I_n\} \quad (155)$$

One can use the average of these values as the outcome of one cycle of ASEC, as follows:

$$\{R_p, I_p\} \quad (156)$$

Different trajectories are representative of the variation of the feature vector directions. If the variation is large enough to get closer to a different vector, the projection vector at the "ith" level switches to the new vector, thus a new trajectory results. If the two projection vectors are very close the variation in the feature vector can be ignored. A measure of this closeness is the condition number of the transfer matrix of equation (151). Therefore, if by adding a projection vector to 'A' the condition number of 'A' jumps (beyond some threshold level so that it can produce the numerical error), the projection vector will be removed from the table and replaced by the previous projection vector (Figure 6.9).

Although the process of space spanning for the trajectory with the replaced projection vector should be repeated to account for the change, it can be ignored because the induced change is very small.

Finally, the minimum subset will be as follows:

$$\{\{f_{11}, f_{21}, f_{31}, \dots, f_{m1}\}, \{f_{12}, f_{22}, f_{32}, \dots, f_{m2}\}, \dots, \{f_{1l}, f_{2l}, f_{3l}, \dots, f_{ml}\}\} \quad (157)$$

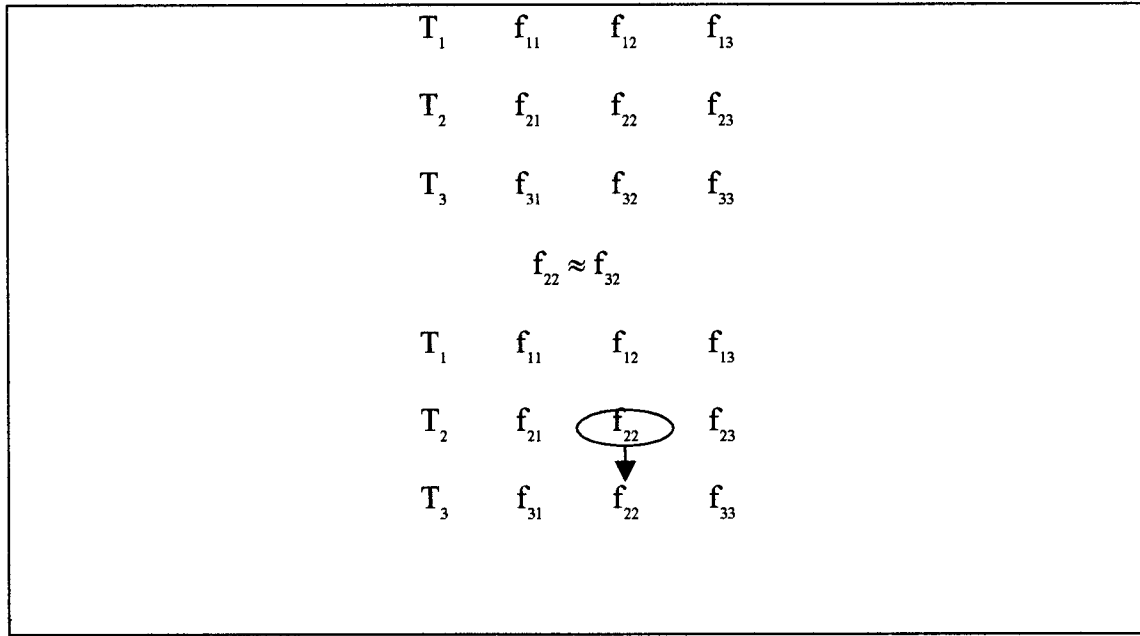


Figure 6.9 Replacement of close features in the trajectory table.

### Example

The given feature set consisted of 88 features (geometrical features from X-ray images and statistical features from others). As mentioned above the two function sets of  $\{\frac{1}{x}, |x|, \text{sign}(x)\sqrt{|x|}, x^2\}$  and  $\{x_i x_j\}$  (cross multiplication) were used in order to generate a new feature set from any given feature set. The approximation model was a linear regression model, i.e.,  $y_i = \sum c_i x_i$ . The Gram-Schmidt space spanning method was followed to generate a sequence of features, for example, one row of Table 5. Ten rows were produced. If the coefficient of determination between two features was more than  $r^2 = 0.9$ , one of them would be replaced with the other. Figure 6.7 shows the improvement in estimation as more terms were added to the feature set (see (154)). Using the inequality of equation (80), as an equality for  $\eta = 0.5$  (representing the expected value), and  $B = (\frac{y_{\max} - y_{\min}}{2})^2$  (see equation (81)) the diagram of 0 was obtained.

The cross-multiplication of features produced a large number of new features. In order to avoid memory overload, an iteration of the ASEC consisted of two cycles; in each only one of the transforming function sets were applied. Once the sequence of features was

produced, the risk functional was estimated by using ten-fold cross validation. The result is shown in the following:

**Table 6: The computed  $r^2$  after each iteration of ASEC**

	$r^2$
The estimation accuracy after the first iteration	0.6448
The estimation accuracy after the second iteration	0.6565
The estimation accuracy after the third iteration	0.7093

The estimation accuracy did not improve after three iterations. The final feature-set size was 38.

### **The block diagram for an iteration of the ASEC**

An iteration of ASEC consists of expanding a given set of features to a new set of features (usually with more features) and then contracting the expanded set in order to produce a new feature set with fewer features. The new feature set includes new and more effective features.

To expand a feature set, a set of predetermined functions can be used. The generality of this method is not compromised because if such a set of functions is not available, any set of universal approximator functions can be used. A given set is contracted by selecting a series of features, by using the Gramm-Schmidt theorem.



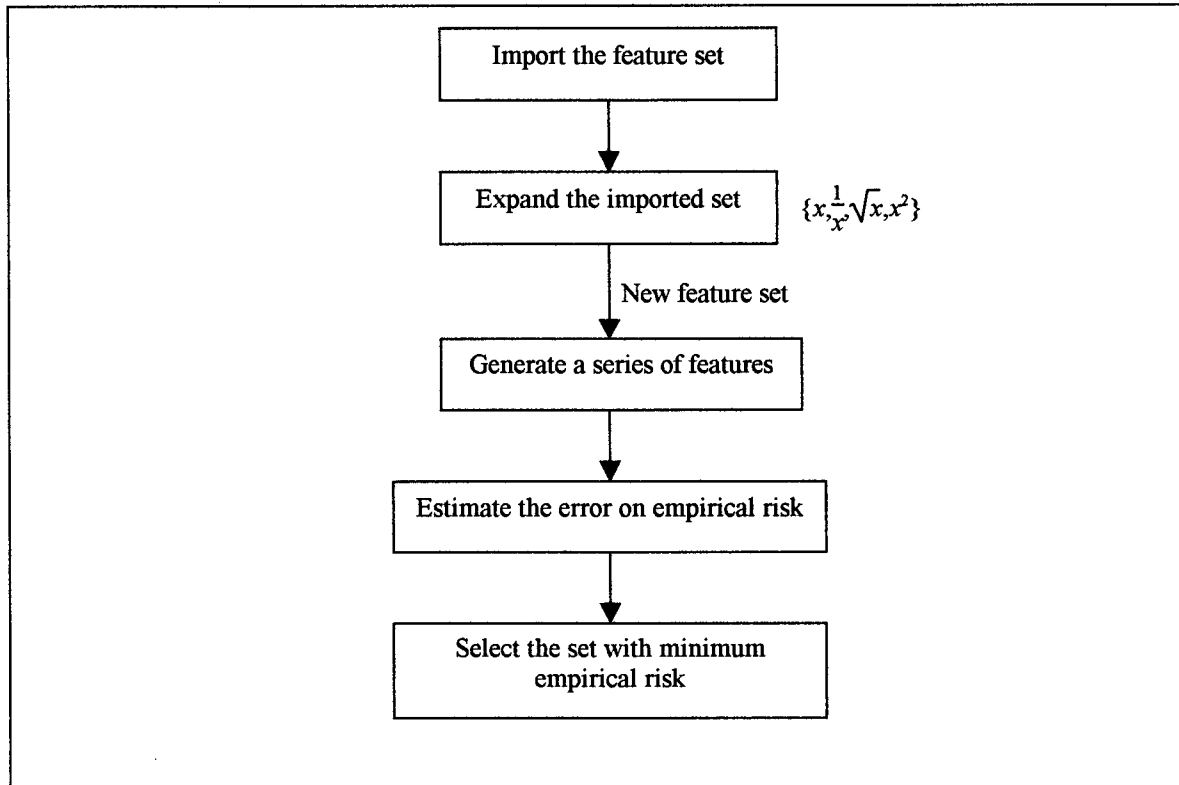


Figure 6.10 The flow chart of one cycle of the ASEC

This cycle is repeated to the point where it no longer improves the estimation accuracy.

## 6.7. Estimation accuracy of different learning machines

All the learning machines were tested using the cross validation technique. The ten fold cross validation technique is a commonly used test. In this test, each specimen is randomly assigned to one of ten sets. In ten repetitions, each set, in order, is used as the testing set and the rest of the ten sets are combined into the training set. The learning machine is trained using the training set and the estimation accuracy is tested by the test set.

For generating the ten sets, a random number was assigned to each specimen. The specimens were sorted with respect to the random numbers. Starting from the smallest random number, one tenth of the specimens were removed and assigned to the first set. Other sets were separated in the same fashion. In cases where the learning capacity of the

learning machine had to be determined, the best structure was determined by using the first training set and the same structure was used for the rest of the accuracy testing.

The number of neighbors controls the learning capacity of the K-nearest neighbor. To test the K-nearest neighbor the number of neighbors was increased (starting from 1) and the cross validation technique was applied. For the multilayer neural network, the number of hidden neurons and the early termination factor control the learning capacity. Starting with one hidden neuron (and increasing), the training was repeated. The early stopping level was changed to increase the cross validation accuracy. Since the steepest descent algorithm was used for neural network training, the estimation accuracy did not necessarily reach as low as the stopping level. Therefore, the training process was repeated by using another initial value for the parameters. Dr. Grudic, who developed the algorithm, trained the SPORE learning machine. This machine adjusts its learning by doing repeated searches.

For the ASEC algorithm the number of coefficients is taken as equivalent to the learning capacity of the machine. The algorithm reached maximum accuracy in five iterations, two non-linear mappings and three cross-multiplication mappings. The expected accuracy diagram was produced, for any iteration, and the optimum number of features (that minimized the estimation error) was found. The best number of parameters was usually approximately 20. Two to three times this number was transferred to the next iteration. The increased number of features normally did not affect the results. Too many features slowed down the algorithm to an impractical rate. Table 7 shows the strength estimation accuracy of the specimens using different learning machines.

**Table 7: The accuracy of estimation by using different learning machines.**

Learning machine	$r^2$ by using 10-fold cross validation
Linear regression	0.60
K-nearest neighbors	0.31, k=15
Multi-layer neural networks	0.68
SPORE	0.73

ASEC	0.71
------	------

An estimate of the contribution of different features is also studied here using a linear regression model. In order to see the contribution of the measurement machines to the strength estimation accuracy, one or more measurement systems are excluded from the input (the profiles). The cross validation technique is not used here because the model is not being tested. Therefore, the entire set of specimens is used for developing the regression model and then testing the accuracy of the estimation. The purpose of this test is to show that all the measurements are contributing to final estimation accuracy and, no one of them can be replaced by a combination of other measured properties. Table 8 shows the accuracy of the estimation for the given set of features.

**Table 8: The accuracy of the estimation for different feature sets.**

included features for estimation	$r^2$ at learning
X-ray	0.4059
MOE	0.5625
SOG	0.5275
X-ray and MOE	0.6101
X-ray and SOG	0.6011
MOE and SOG	0.6572
X-ray and MOE and SOG	0.6805

## 6.8. Discussion of the results

Table 7 shows two properties of the learning problem for the database of this thesis. By comparing the accuracy of linear regression and nonlinear transformations (Neural

Networks, ASEC, and SPORE) it becomes clear that higher order self and cross products of features lead to improved performance.

The poor estimation accuracy from K-nearest neighbor shows the effect of high dimensionality of the feature space. As was pointed out in Section 6.2, high dimensionality of a feature set (in this case, 88 features) makes the concept of 'neighborhood' ineffective. Especially if some dimensions are inherently noisy as they are in these types of measurements.

The ASEC algorithm and the SPORE algorithm followed two different approaches for improving the strength estimation. In the ASEC algorithm, the given feature set is being transformed (by using *a priori* knowledge) into a more effective set of features, while the SPORE algorithm used the given feature set (88 features) and found the best hyper-surface that maps them to the output (measured strength). The closeness of the estimation accuracy by the two algorithms shows the equal validity of the two approaches. However, the estimation accuracy of ASEC could be improved by searching for better transformations and more complex estimation functions (a simple linear estimator was used in this research).

The final feature set that was produced by ASEC included 38 features. This shows that the original feature set (with 88 features) includes redundant information. Using a smaller (and more effective) feature set reduces the VC-dimension of some of the estimators (e.g. linear regression) and thereby improves the estimation accuracy for those models. This can be seen in Table 6, which shows that the error is minimized at 38 features.

## ***Chapter 7. Conclusions and Future Directions***

### **7.1. Conclusions**

The problem of estimating the tensile strength of lumber is the focus of this thesis. This problem was posed as a multi-sensory measurement problem and a small-sample learning problem. It was shown that the regression approximation was successful in estimating strength, and that regular industrial estimation accuracy can be significantly improved by using a multiple sensory system.

The measurement means were those most commonly used in industry, so their speed could be adjusted to match speeds found in industry. The principles and practical limitations of each measurement means was studied in order to make sure each did not produce the significant effect on estimation accuracy. A study for calibrating the dynamic bending machine, the machine commonly used for wood grading, was performed. The result of this study showed that the flat-wise dynamic MOE that was measured by the dynamic bending machine was highly correlated ( $r^2 = 0.92$ , see Section 2.4.1) to flat-wise static MOE. The significance of this finding is that the type of MOE measuring machine used would not affect the accuracy of strength estimation.

A conic model of knots and a knot detection algorithm was developed. The estimation accuracy of this model was demonstrated through simulation ( $r^2 = 0.81$  for smaller diameter and  $r^2 = 0.98$  for larger diameter). This model was developed for single knots, which comprise the majority of the defects of boards. It was demonstrated that this model could be used for producing a simple representation of a board consisting of clear wood and single knots. Also, it was shown that the model and the related detection algorithm could be used for detecting the same defects in other measured profiles, such as the CCD camera image of a board.

Characterizing a board consisted of transforming the measured profiles into a set of previously studied statistical features and geometrical features as proposed in this thesis.

The measured profiles, except for the X-ray images, were transformed to their statistical moments. From the two X-ray images of a board the geometrical features of the board were calculated. The knots locations were shown to match the location of the actual knots.

A brief survey of different learning techniques was presented and the problem of strength estimation as an empirical learning problem was discussed. Because of the high cost of measuring such properties, the problem of lumber strength estimation was considered as a small-sample learning problem. It was shown that classification VC-dimension could be extended to the VC-dimension of a regression estimator. During this demonstration it was shown that a classifier could be constructed by using a regression estimator, and *vice versa*. This concept established the very basis of this research, which is to improve the grading accuracy (a classification problem) by improving strength estimation accuracy (a regression problem).

The ASEC learning model was proposed in order to produce an optimal feature set from the original feature set. This model allows a researcher to incorporate *a priori* knowledge about input-output dependence. If such dependence does not exist, a generic mapping can be used. The accuracy of strength estimation when using this model, along with a simple relationship between knot-size and stress-distribution, showed that this model performs better than most existing learning models.

## **7.2. Future Directions**

Throughout the course of this research, a high speed of computation has been a goal since the real time applicability of the process is a focus. Therefore, the solutions presented here were attempted without any recursive approach. Any improvement in estimation accuracy by using gained by recursive algorithms (for detection) could be further investigated. In particular, the knot detection algorithm could be repeated by adjusting the size and location of a partial cone in the space in order to maximize the similarity of its orthographic projections to the detected knot patterns from the X-ray images.

A more substantial model of clear wood properties for the board model (see Figure 4.8) could be developed. For example, the gradual variations in clear wood density could be

added to the model. Also grain angle, and MOE could be included with the properties of wood in every element of the model.

Any performance improvement for the ASEC algorithm by modifying its transformations could be investigated. In essence, the effect of other *a priori* knowledge in the form of feature transformations (for example higher order terms in Irwin's equation) as well as non-linear estimators are to be investigated.

Table 8 indicates that all the measurement means contributed to the estimation accuracy of the tensile strength of wood. The estimation accuracy is likely to be improved by adding new measurement means, such as ultrasound or visual sensing, in order to gather relevant information that is not presently available.

Other analyses, such as stress analysis, could be used for generating more effective features from the existing measurements.

The estimation improvement due to knowing basic wood properties could also be examined.

## ***Appendix: Board failure; initiation and propagation of a fracture***

Although fracture mechanics was not used in this research for analyzing the tensile strength of lumber, it provides the understanding of how the failure of wood takes place. In selecting the representative features of lumber, one needs an understanding of the failure process because it sometimes plays a leading role. This Appendix is intended as a brief examination of the wide area of fracture mechanics.

In the literature, a crack is defined as a split in material that grows (propagates) when applied stress increases. A fracture is the growth of a crack due to applied stress. These two terms (crack and fracture) are loosely used interchangeably here because the meaning is conveyed by the text and because it is not essential to the adopted methodology of this thesis.

The process of failure due to fracture can be caused by applied force or by material status. "Fatigue" and "creep" are two examples where fracture is caused by the feature of the applied force and by the feature of the material state. In fatigue, the cyclic pattern of the applied force causes the fracture but in creep the state of material (hot metal, for example) causes the fracture in low stress levels. Also, fast fracture propagation and slow fracture propagation are distinguished in the literature. The tensile strength of lumber being estimated in this thesis occurs where slow monotonic loading is applied to a test specimen. There are three major modes of failure, opening (mode-I), shear (mode-II), and tearing (mode-III). Mode-I of failure happens when the crack progresses across the board while the board is under tensile stress. Mode-II of failure happens when the crack is progressing along the board. Mode-III is not usually considered for the failure of a board. The failure of a specimen in most cases is an opening mode of failure (mode I), as will be explained later. However, the propagation of fracture in wood (across and along the grain) suggests that failure mode must be a combination of mode-I and mode-II ([56]).

There are numerous micro and macro fractures that already exist in wood. High stress or low strength zones create the condition in which the existing fractures can grow. In



theory, the fracture can start at any point in/on the board because micro fractures are randomly distributed in a board. In practice, it is almost certain that one of the defects of wood will be the location of fracture initiation, and in most cases, a knot initiates a fracture ([11], [12]). The tip of a fracture creates a high stress zone that deforms the material and causes its failure. The stress concentration at the tip of the fracture is much more than that of other natural defects of wood. Therefore, a fracture has a good chance of propagating until it stops at some point. It was observed that a crack ends in another defect (like knot) or when the crack propagation direction goes along the board. There are three types of fracture: opening, sliding, and tearing. Once the failure starts at a point it propagates step wise, that is, the fracture propagates consequently along the grains and across them until complete failure takes place. In practice a combination of opening and sliding takes place in the failure of wood. The sliding mode usually takes much more tensile stress to propagate. This is why in some cases one crack practically stops progressing. If another crack initiates with stress level close to that of the first crack, then this second crack gets a chance to propagate.

In grown knots there is density difference and high grain variation with respect to clear wood. In some cases of knots, only a high grain angle exists and that happens where the knot is related to the bottom of a branch (where the branch wood starts by high grain deviation of the trunk wood). Variation in the density of a knot with respect to clear wood causes high stress zones, which can cause the failure of clear wood. In high grain angle parts of a knot the tensile strength of wood drops significantly such that it can fail under regular tensile stress that clear wood, such as, trunk wood, with regular grain angle can stand.

The normally used tool for analyzing the initiation and propagation of fractures in material is the linear elastic finite elements. Critical stress intensity factor, or fracture toughness ([57]) is the concept on which linear elastic fracture mechanics is based. It is assumed that once a fracture is initiated, it can progress almost regardless of body and loading geometry (given certain specific general conditions ([33],[58],[59])).

Once a crack is created in the body of the board, it generates its own high stress zone. The stress distribution around a crack tip and its propagation is the subject of many theories.

Based on stress distribution around a fracture, different theories were developed that predict the failure of material ([33], [60], [61], [62]). Griffith ([39]) developed a fracture theory based on the decrease in material potential energy ( $U_p$ ) and fracture surface energy ( $U_s$ ) as follows:

$$\frac{\partial U_p}{\partial a} + \frac{\partial U_s}{\partial a} = 0 \quad (158)$$

where  $U_p$  and  $U_s$  are as follows:

$$U_p = -\frac{\pi \sigma^2 \alpha^2 t}{E} \quad (159)$$

$$U_s = 4at\gamma_s \quad (160)$$

Equation (158) states that when the stress level reaches its critical level,  $U_p$  will be transformed to  $U_s$ . Therefore the critical stress level can be obtained as follows:

$$\sigma_c = \sqrt{\frac{2E\gamma_s}{\pi\alpha}} \quad (161)$$

$\sigma$  is the stress level,  $E$  is modulus of elasticity,  $t$  is thickness,  $\alpha$  is fracture surface, and  $\gamma_s$  is the elastic surface energy per unit thickness. Based on this theory, if the critical stress in the material exceeds the critical level, the material fails and the crack propagates.

A theory that distinguishes different modes of failure was published in [57]. Based on this theory, the stress distribution around the tip of the fracture can be modeled as follows:

$$\sigma_{ij} = \frac{1}{\sqrt{r}} (K_I f_I'(\theta) + K_{II} f_{II}'(\theta) + K_{III} f_{III}'(\theta)) + \text{higher terms} \quad (162)$$

where  $r$  and  $\theta$  are the coordinates of the material element in a two dimensional polar coordinate system.  $r$  and  $\theta$  are measured with respect to fracture tip and fracture line.  $K_p$

$K_{Ic}$  and  $K_{IIc}$  are stress intensity factors for opening, sliding, and tearing modes of failure. For a two-dimensional tensile stress problem only an opening mode of failure exists. For a fracture of length  $2a$  critical stress is found from the following equation.

$$K_{Ic} = \sigma_c \sqrt{\pi a} \quad (163)$$

$K_{Ic}$  is a material property. Once the stress intensity factor of any type exceeds the critical level the fracture of that type takes place.

The mixed mode of failure that takes place in a material such as wood was explored in ([63]). The failure criterion in this case is as follows:

$$\frac{K_I}{K_{Ic}} + \left(\frac{K_{II}}{K_{IIc}}\right)^2 = 1 \quad (164)$$

$K_I$  and  $K_{II}$  are the mode-I and mode-II stress intensity factors, and  $K_{Ic}$  and  $K_{IIc}$  are the fracture toughness for those failure modes, which are material properties (constant). A theory initially used in fracture analysis routines ([12]) was based on maximum stress failure theory. A second theory that combines mode-I and mode-II of failure was later used for wood ([63]).

The information from fracture mechanics contributes to the estimation problem in two different ways. Although the effect of knots on strength is the subject of other research, geometrical features will be defined in order to include this knowledge (in a qualitative way) in the feature set. For example, knowing that the knots are usually the initiating points of fractures, the distance between two knots is included as a feature. Also, the form of any dependence of strength to the feature set can be incorporated into the learning system as *a priori* knowledge. For example, equation (162) states that the stress level produced due to the existence of a fracture diminishes by  $\frac{1}{\sqrt{r}}$ . In ASEC model (Chapter 6) this information is used for generating new features from the given set of features by four different mappings of  $\frac{1}{x}$ ,  $|x|$ ,  $\text{sign}(x)\sqrt{x}$ , and  $x^2$ .

## References

- [1] "VisionSmart X-ray machine manual", VisionSmart INC, Edmonton, Alberta, Canada.
- [2] "Grain Angle Measurement with the Metriguard Model 510", Metriguard Inc, Pullman, WA 99163.
- [3] "Model 5100 slope-of-grain indicator operation manual", Metriguard, INC, PO Box 396, Pullman, Washington 99163, 1977.
- [4] Logan J. D., "Bending Stress in the Metriguard Model 7200 HCLT and the CLT", Metriguard INC.
- [5] Blass H., Gard W., "Machine Strength Grading of Timber", Pacific Timber Engineering Conference, Australia, pp. 598-603, 1994.
- [6] Perstorper M., "Strength and Stiffness Prediction of Timber Using Conventional and Dynamic Methods"
- [7] Lam F., Foschi R. I., Barrett J. D., Q. Y. He, "Modified algorithm to Determine Localized Modulus of Elasticity of Lumber", Wood Science and Technology, Springer-Verlag, pp. 81-94, 1993.
- [8] Bechtel F. K., Allen J. R., "Non-destructive Testing Methods for Lumber", United States Patent, 1990.
- [9] McDonald K. A., Cramer S. M., Bendtsen B. A., "Research Progress in Modeling Tensile Strength of Lumber from Localized Slope of Grain", Proceedings of the Sixth Symposium on the Nondestructive Testing of Wood, Washington State University, Pullman, WA, 1987.
- [10] Cramer S. M., McDonald K. A., "Predicting Lumber Tensile Stiffness and Strength with Local Grain Angle Measurement and Failure Analysis", Wood and Fiber Science, 21(4), pp. 393-410, 1989.
- [11] McGowan, W. M., "Parallel-to-grain tensile properties of visually graded 2 by 6 -inch Douglas-fir", Information Report VP-X-46, Forest Products Laboratory, Vancouver, BC.
- [12] Shi Y., "Multiple Knot Failure Mechanisms in Structural Lumber", Ph. D. thesis dissertation, Civil and Environmental Engineering Department, the University of Wisconsin, Madison, 1996.
- [13] Courchene T., "The Effect of Edge Knots on the Strength of Western SPF MSR Lumber", MSc. thesis dissertation, The University of British Columbia, Vancouver, 1996.
- [14] Schajer G. S., "Method for Estimating The Strength of Wood", United States Patent, 1990.
- [15] Newnes INC., "X-rays help measure lumber strength on-line", Wood Technology, pp. 41-43, June 1997.
- [16] Logan J. D., "X-ray Lumber Grading V.S. Traditional Methods, Review of a recent article", Metriguard INC., Pullman, WA, 1998.
- [17] Quin F. J., Steele P. H. Shmulsky R., "Locating Knots in Wood with an Infrared Detector System", Forest Products Journal, vol. 48, no. 10, pp. 80-84, 1998.

- [18] Cho T. H., Conners, R. W., Araman P. A., "A Computer Vision System for Automated Grading of Rough Hardwood Lumber Using a Knowledge-Based Approach", Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 1990.
- [19] Rouger F., "Application of a Modified Statistical Segmentation Method to Timber Machine Strength Grading", Wood and Fiber Science, 28(4), pp. 438-449, 1996.
- [20] Breiman L., Friedman J. H., Olshen R. A., Stone C. J., "Classification and Regression Trees", Wadsworth International Group, 1984.
- [21] Grudic G., "Nonparametric Learning from Examples in Very High Dimensional Spaces", Ph.D. Thesis dissertation, The University of British Columbia, 1997.
- [22] National Lumber Grading Authority, "North American Export Standard for Machine Stress-Rated Lumber", 1986.
- [23] Biernacki J. M., Lam F., Barrett J. D., "Benefits of Improved Strength and Stiffness Prediction of MEL and MSR Lumber", Faculty of Forestry, The University of British Columbia, 1997.
- [24] Hoag M. L., Krahmer R. L., "Polychromatic X-ray attenuation characteristics and wood densitometry applications", Wood and Fiber Science, v. 23(1), pp. 23-31, 1991.
- [25] Bertin E. P., "Principles and practice of X-ray Spectrometric Analysis", Plenum Press, Newyork, 1970.
- [26] Victoreen J. A., "Calculation of X-ray mass-absorption coefficient", Journal of Applied Physics, V. 20, pp. 1141-1147, 1949.
- [27] Schirato R. C., Polichar R. M., Reed J. H., "Development of monolithic CdZnTe arrays with improved energy and spatial resolution", pp.47-56, SPIE, v. 2278 X-ray and UV Detectors, 1994.
- [28] Shen J., Schajer G., Parker R., "Theory and Practice in Measuring Wood Grain Angle Using Microwaves", IEEE Transaction on Instrumentation and measurement, pp. 803-809, vol. 43, No. 6, December 1994.
- [29] Shen J., "Wood Property Measurements Using Microwaves", Ph.D. dissertation, The University of British Columbia, 1995.
- [30] Shen J., Schajer G., Parker R., "A New Probe for Characterizing Elliptically Polarized Microwave Fields", Journal of Microwave Power and Electromagnetic Energy, pp.55-62, vol. 28, No. 4, 1993.
- [31] Torgovnikov G. I., "Dielectric Properties of Wood and Wood-Based Materials", Springer Verlag, 1993.
- [32] Saboksayr H., Lau W., "Landmark CLT™ Calibration", UBC internal report, May 1999.
- [33] Broberg, K. B., "Cracks and Fracture", Academic Press, 1999.
- [34] Weibull, W., "A Statistical Theory of the Strength of Material", Proceedings of Swedish Institute of Engineering Research (Ingénjörsvetenskapsakademiens), Stockholm, 1939.
- [35] Weibull, W., "Investigating into Strength Properties of Brittle Material", Svenska teknologf m reningens f m rlag, Stockholm, 1938.

- [36] McGowan W. M., "Parallel-to-grain Tensile Properties of Visually Graded 2x6-inch Douglas-Fir", report # VP-X-46, Forest Products Laboratory, Department of Fisheries and Forestry of Canada, Vancouver, B.C., 1968.
- [37] Boatright S. W., Garret G. G., "The Effect of Microstructure and Stress State on the Fracture of Behavior of Wood", *Journal of Materials Science*, vol. 18, pp. 2181-2199, 1983.
- [38] Barrett, J. D., "Fracture Mechanics and the Design of Wood Structures", *Philosophical Transactions of Royal Society of London*, v. 299, p. 217-226, 1981.
- [39] Griffith, A. A., "The Phenomenon of Rupture and Flow in Solids", *Philosophical Transactions of Royal Society of London*, v. 221, p. 163-197, 1921.
- [40] Smith, I., Hu, L. J., "Factors Affecting Mode I Fracture of Plantation Grown Red Pine", *Wood Science and Technology*, v. 28, p. 144-157, 1994.
- [41] Wagner, H. D. "Statistical Concepts in the Study of Fracture Properties of Fibrous Composites", *Application of Fracture Mechanics to Composite Materials*, v. 6, Elsevier Science Publisher, Amsterdam, 1989.
- [42] Stanzl-Tschegg S. E., Tan D. M., Tschegg E. K., "New Splitting Method for Wood Fracture Characterization", *Wood Science and Technology*, vol. 29, pp. 31-50, 1995.
- [43] Bazant, Z. P., Xi, Y., "Statistical Size Effect in Quasi-Brittle Structures Part I: Is Weibull Theory Applicable?", *Journal of Engineering Mechanics*, v. 117, p. 2609-2622, 1991.
- [44] Bazant, Z. P., Xi, Y., "Statistical Size Effect in Quasi-Brittle Structures Part II: Nonlocal Theory", *Journal of Engineering Mechanics*, v. 117, p. 2623-2640, 1991.
- [45] Haygreen J. G., Bowyer J. L., "Forest Products and Wood Science, an introduction", The Iowa States University Press, 1982.
- [46] Hankinson, R. L., "Investigation of Crushing Strength of Spruce at varying angles or grain, U.S. Air Service Information Circular #259, Chief of Air Service, Washington, D.C., 1921.
- [47] "Wood Engineering Handbook", Forest Products Laboratory, Prentice Hall, 1990.
- [48] Dabholkar A. Y., "Analysis of Wood With Knots and Cross grain", Ph.D. dissertation, Colorado State University, 1980.
- [49] Goodman J.R., Bodig J., "Orthotropic Strength of Wood in Compression", *Wood Science*, vol. 4(2), pp. 83-94, 1971.
- [50] Norris C. B., "Strength of Orthotropic materials subject to combined stresses", USDA Forest Products Lab, Report No. 1816, Forest Products Laboratory, Madison, WI, 1962.
- [51] Cote W. A., Hanna R. B., "Ultrastructural Characteristics of Wood Fracture Surfaces", *Wood and Fiber Science*, vol. 15(2), pp. 135-163, 1983.
- [52] Lam F., Varoglu E., "Variation of Tensile Strength Along the Length of Lumber, Part 1: Experimental", *Wood Science and Technology*, pp. 351-359, 1991.
- [53] Lam F., Varoglu E., "Variation of Tensile Strength Along the Length of Lumber, Part 2: Model Development and Verification", *Wood science and Technology*, pp. 449-458, 1991.

- [54] Friedman J. H., "An Overview of Predictive Learning and Function Approximation", in From Statistics to Neural Networks, Cherkassky V., Friedman J. H., Wechsler H., pp. 1-61, Springer-Verlag, 1994.
- [55] Bury K., "Statistical Distributions in Engineering", The University of British Columbia, 1995.
- [56] Ashby M. F., Easterling K. E. Harrysson R., Maiti S. K., "The Fracture and Toughness of Woods", Proceedings of the Royal Society of London, vol. A 398, pp. 261-280, 1985.
- [57] Irwin G. R., "Analysis of Stresses and Strains Near the End of a Crack Traversing a Plate", Journal of Applied Mechanics, vol. 24, pp. 361-364, 1957.
- [58] Barenblatt G. I., "The Formation of Equilibrium Cracks During Brittle Fracture. General Ideas and Hypotheses. Axial-symmetric Cracks". Journal of Applied Mathematics and Mechanics, vol. 23, pp. 622-636, 1959.
- [59] Barenblatt G. I., "Equilibrium Cracks Formed During Brittle Fracture. Rectilinear Cracks in Plane Plates". Journal of Applied Mathematics and Mechanics, vol. 23, pp. 1009-1029, 1959.
- [60] Kolosov, G., Doctoral dissertation, Dorpat, Estonia, 1909.
- [61] Kolossoff, G., Uber einige Eigenschaften des ebenen Problems der Elastizitatstheorie. Zeitschrift fur Mathematik und Physik, vol. 62, pp. 384-409, 1914.
- [62] Inglis C. E., "Stresses in a Plate due to the Presence of Cracks and Sharp Corners", Transactions of the Institution of the Naval Architects, vol. 1, p. 219-230, 1913.
- [63] Wu E. M., "Application of Fracture Mechanics to Anisotropic Plates", Journal Of Applied Mechanics, pp. 967-974, 1967.
- [64] Cramer S. M., Goodman J. R., "Model for Stress Analysis and Strength Prediction of Lumber", Wood and Fiber Science, 15(4), pp. 338-349, 1983.
- [65] Lam F., Varoglu E., "Grading MSR Lumber for the Optimization of Market Value",
- [66] Cherkassky V. S., Mulier F., "Learning from Data", John Wiley & Sons, 1998.
- [67] Vapnik, V., "The Nature of Statistical Learning Theory", New York: Springer Verlag, 1995.
- [68] Castleman K. R., "Digital Image Processing", Prentice-Hall, 1979.
- [69] Horn P., "Robot Vision", McGraw-Hill, 1986.
- [70] Dabholkar A. Y., "Analysis of wood with knots and cross grain", Ph. D. Dissertation, Colorado State University, Fort Collins, USA, 1980.
- [71] Daugherty R. L., Franzini J. B., "Fluid Mechanics with Engineering Applications", McGraw-Hill Book Company, 1977.
- [72] Friedman J. H., "An Overview of Predictive Learning and Function Approximation", in From Statistics to Neural Networks, V. Cherkassky, J. H. Friedman, and H. Wechsler (eds.), NATO ASI Series F, 136, New York: Springer Verlag, 1994.
- [73] Vapnik V., Levin E., Le Cun Y., "Measuring the VC-dimension of a Learning Machine", Neural Computation, 6, 851-876, 1994.

- [74] Hand D. J., "Kernel Discriminant Analysis", Research Studies Press, 1982.
- [75] Zurada, J. M., "Introduction to Artificial Neural Systems", West Publishing Co.,
- [76] Hornik K., Stinchcombe M., White H., "Multilayer Feedforward Networks are Universal Approximators", Neural Networks, Vol. 2, pp. 359-366, 1989.
- [77] Funahashi K., "On the Approximate Realization of Continuous Mappings by Neural Networks", Neural Networks, Vol. 2, pp. 183-192, 1989.
- [78] Breiman L., Friedman J. H., Olshen R. A., Stone C. J., "Classification and Regression Trees", Wadsworth Inc., 1984.
- [79] Van Trees H. L., "Detection, Estimation, and Modulation Theory", part I, John Wiley and Sons, 1968.
- [80] Gersho A., Gray R. M., "Vector Quantization and Signal Compression", Kluwer Academic Publishers, 1991.
- [81] Vapnik V., "Statistical Learning Theory", John Wiley and Sons, 1998.
- [82] Miller A. J., "Subset Selection in Regression", Chapman and Hall, 1990.
- [83] Kreyszig E., "Introductory Functional Analysis with Applications", John Wiley & Sons, 1989.
- [84] Devor J. L., "Probability and Statistics for Engineering and the Sciences", third edition, Brooks/Cole Publishing Co, 1991.