QUALITY CONTROL METHODS FOR MONITORING THE VARIABILITY OF MOISTURE CONTENT IN KILN-DRIED LUMBER

by

CATALIN RISTEA

B.Sc., Transylvania University, Romania, 1994

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE in

THE FACULTY OF GRADUATE STUDIES
 THE FACULTY OF FORESTRY
 Department of Wood Science

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 2001

© Catalin Ristea, 2001
In presenting this thesis in partial fulfillment of the requirements for the degree of Master of Science 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 Wood Science
Faculty of Forestry

The University of British Columbia
Vancouver, Canada

Date ________________

Date ________________
ABSTRACT

Monitoring the variability of moisture content in lumber is a problem of utmost importance in kiln-drying. This thesis focused on determining the distribution of moisture content in lumber, and on the application of statistical process control principles to monitor the drying process through the use of quality control charts. Specific parameters of the charts for monitoring the process average and process dispersion were determined. The charts were developed on the assumption that the data under analysis has a Lognormal distribution. To test whether this assumption is valid or not for moisture content, graphical and numerical tests were employed as goodness of fit methods. Graphical methods utilized in this research were symmetry and probability plots, and empirical cumulative distribution function graphs. Numerical methods employed in this study were goodness of fit tests based on empirical distribution function statistics. Procedures were developed for both graphical and numerical goodness-of-fit methods, and also for construction of control charts for Lognormal variables. The methods presented were tested on a data set of actual Douglas-Fir (*Pseudotsuga menziesii*) lumber collected from a production facility in British Columbia, Canada. It was determined that the Lognormal distribution provided a good fit for the experimental data. Therefore, the proposed control charts for Lognormal data are to be used instead of the customary charts based on the Normality assumption.
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT ..................................</td>
</tr>
<tr>
<td>TABLE OF CONTENTS ..........................</td>
</tr>
<tr>
<td>LIST OF TABLES ..................................</td>
</tr>
<tr>
<td>LIST OF FIGURES ..................................</td>
</tr>
<tr>
<td>LIST OF PROCEDURES ...........................</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS ..................................</td>
</tr>
<tr>
<td>1 INTRODUCTION ...................................</td>
</tr>
<tr>
<td>1.1 OVERVIEW ....................................</td>
</tr>
<tr>
<td>1.2 PROBLEM DEFINITION ..........................</td>
</tr>
<tr>
<td>1.3 RESEARCH OBJECTIVES .......................</td>
</tr>
<tr>
<td>2 LITERATURE REVIEW AND BACKGROUND MATERIAL ....</td>
</tr>
<tr>
<td>2.1 OVERVIEW ....................................</td>
</tr>
<tr>
<td>2.2 MOISTURE CONTENT ............................</td>
</tr>
<tr>
<td>VARIATION IN MOISTURE CONTENT ..................</td>
</tr>
<tr>
<td>DISTRIBUTION OF MOISTURE CONTENT ...............</td>
</tr>
<tr>
<td>2.3 GOODNESS OF FIT METHODS .....................</td>
</tr>
<tr>
<td>GRAPHICAL ANALYSIS ..................................</td>
</tr>
<tr>
<td>NUMERICAL TESTS ..................................</td>
</tr>
<tr>
<td>2.4 PRINCIPLES OF STATISTICAL PROCESS CONTROL ......</td>
</tr>
<tr>
<td>CONTROL CHARTS ...................................</td>
</tr>
<tr>
<td>CONTROL CHARTS FOR LOGNORMAL VARIABLES ..........</td>
</tr>
<tr>
<td>SAMPLING AND SAMPLE SIZE .......................</td>
</tr>
<tr>
<td>STATISTICAL PROCESS CONTROL IN LUMBER DRYING ......</td>
</tr>
<tr>
<td>3 METHODS AND MATERIALS .......................</td>
</tr>
<tr>
<td>3.1 OVERVIEW ....................................</td>
</tr>
<tr>
<td>3.2 DATA COLLECTION .............................</td>
</tr>
<tr>
<td>DATA FOR DISTRIBUTION ANALYSIS ..................</td>
</tr>
<tr>
<td>DATA FOR CONTROL CHARTS ........................</td>
</tr>
<tr>
<td>DETERMINATION OF MOISTURE CONTENT ................</td>
</tr>
<tr>
<td>3.3 GRAPHICAL ANALYSIS OF DATA ..................</td>
</tr>
<tr>
<td>INVESTIGATION OF SYMMETRY ......................</td>
</tr>
<tr>
<td>PROBABILITY PLOTTING ............................</td>
</tr>
<tr>
<td>Normal Probability Plotting ....................</td>
</tr>
<tr>
<td>Lognormal Probability Plotting ..................</td>
</tr>
<tr>
<td>Weibull Probability Plotting ....................</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1. Absolute values of distances from sample median to percentiles</td>
<td>63</td>
</tr>
<tr>
<td>Table 2. Observed statistics and critical values for ECDF statistics for hypothesis testing</td>
<td>67</td>
</tr>
<tr>
<td>Table 3. Parameters of preliminary data for construction of control charts based on the Normal distribution</td>
<td>69</td>
</tr>
<tr>
<td>Table 4. Control limits and center lines for control charts based on Normal assumption</td>
<td>70</td>
</tr>
<tr>
<td>Table 5. Parameters of preliminary data for construction of control charts based on the Lognormal distribution</td>
<td>71</td>
</tr>
<tr>
<td>Table 6. Control limits and center lines for control charts based on Lognormal assumption</td>
<td>72</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Symmetric and skewed distributions</td>
</tr>
<tr>
<td>2</td>
<td>Empirical Cumulative Distribution of MC for Douglas-Fir, and the relation</td>
</tr>
<tr>
<td></td>
<td>of $10^{th}$ and $90^{th}$ percentiles about the median ($50^{th}$ percentile)</td>
</tr>
<tr>
<td>3</td>
<td>Plot of upper vs. lower observations for investigation of symmetry.</td>
</tr>
<tr>
<td>4</td>
<td>Normal probability plot for the moisture content data.</td>
</tr>
<tr>
<td>5</td>
<td>Lognormal probability plot for moisture content data.</td>
</tr>
<tr>
<td>6</td>
<td>Weibull probability plot for moisture content data.</td>
</tr>
<tr>
<td>7</td>
<td>Control chart for sample averages based on the Normality assumption</td>
</tr>
<tr>
<td>8</td>
<td>Control chart for sample standard deviations based on the Normality</td>
</tr>
<tr>
<td></td>
<td>assumption</td>
</tr>
<tr>
<td>9</td>
<td>Control chart for sample averages ($\bar{y}$ chart) based on Lognormal</td>
</tr>
<tr>
<td></td>
<td>assumption</td>
</tr>
<tr>
<td>10</td>
<td>Antilog $\bar{y}$ chart (in original scale) based on Lognormal assumption</td>
</tr>
<tr>
<td>11</td>
<td>Control chart for sample standard deviations ($S$ chart) based on Lognormal</td>
</tr>
<tr>
<td></td>
<td>assumption</td>
</tr>
<tr>
<td>12</td>
<td>Antilog $S$ chart (in original scale) based on Lognormal assumption</td>
</tr>
</tbody>
</table>
Figure 13. Control chart for sample averages based on the Normality assumption. ..................76

Figure 14. Antilog $\bar{y}$ chart (in original scale) based on Lognormal assumption. ..................76

Figure 15. Control chart for sample standard deviations based on the Normality assumption. ....77

Figure 16. Antilog $S$ chart (in original scale) based on Lognormal assumption. ....................77
## LIST OF PROCEDURES

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure 1: Constructing an ECDF plot</td>
<td>26</td>
</tr>
<tr>
<td>Procedure 2: Assessing the symmetry by the distance of percentiles to the sample median</td>
<td>26</td>
</tr>
<tr>
<td>Procedure 3: Assessing symmetry by plotting the upper half of the observations against the lower</td>
<td>27</td>
</tr>
<tr>
<td>Procedure 4: Normal probability plotting</td>
<td>31</td>
</tr>
<tr>
<td>Procedure 5: Lognormal probability plotting</td>
<td>32</td>
</tr>
<tr>
<td>Procedure 6: Weibull probability plotting</td>
<td>34</td>
</tr>
<tr>
<td>Procedure 7: Goodness-of-fit test of hypothesis for the Normal distribution</td>
<td>40</td>
</tr>
<tr>
<td>Procedure 8: Goodness-of-fit test of hypothesis for the Lognormal distribution</td>
<td>41</td>
</tr>
<tr>
<td>Procedure 9: Goodness-of-fit test of hypothesis for the Weibull distribution</td>
<td>42</td>
</tr>
<tr>
<td>Procedure 10: Construction of charts for sample means – or scale parameters – ((\bar{y}) chart) and the corresponding Lognormal control charts in original scale of measurements (antilog (\bar{y}) chart)</td>
<td>57</td>
</tr>
<tr>
<td>Procedure 11: Construction of charts for shape parameter – sample standard deviations - ((S) chart) and the corresponding Lognormal control charts in original scale of measurements (antilog (S) chart)</td>
<td>60</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to offer my sincere thanks to my advisor, Dr. Thomas C. Maness, for his constant encouragement and guidance throughout this research. Special thanks to my supervisory committee, Dr. Stavros Avramidis, Dr. Robert Kozak, and Dr. Ian Hartley, for many valuable suggestions and comments. I wish to express my gratitude to Dr. Michael Stephens who has made a substantial contribution to my understanding of goodness-of-fit testing.

Last but not least, I would like to thank my wife Anca and my son Alex for their unconditional love and constant support, without which this work would have not been possible.
QUALITY CONTROL METHODS FOR MONITORING
THE VARIABILITY OF MOISTURE CONTENT IN
KILN-DRIED LUMBER

1 INTRODUCTION

1.1 OVERVIEW

The variability of moisture content (MC) in solid wood is a problem of utmost importance in kiln-drying processes. Variation in final moisture content of lumber can cause serious problems in its subsequent processing and use. It is known that the moisture content of lumber varies considerably among boards in a kiln charge. This can happen because of natural variability in drying rate or initial moisture content, sapwood and heartwood, wet pockets in the lumber, and/or variability in drying conditions in various parts of the kiln. It is known that substantial MC variation can exist between package locations in the kiln, between row locations in the package, between kiln charges, and also between different kilns at the same mill.

The moisture content variation needs to be estimated at all drying stages: before the actual kiln drying begins to determine the actual setting of the kiln controls; during the kiln drying period to monitor MC distribution and/or change drying schedules; and finally, at the end of a kiln charge to determine whether the lumber has acceptable moisture variation or not. To
minimize lumber degrade during drying requires that the variation of moisture content be accurately monitored and controlled. This can be done using the quality tools of statistical process control (SPC).

The essential idea of SPC is to improve the quality of a process continuously through the constant application of statistical methods to process control. A major goal of SPC is to detect the occurrence of any assignable causes of disturbances in the process as soon as possible so that investigation of the process and corrective action can be taken before many nonconforming products reach the final stage of the process. This is usually done through the use of parametric control charts for variables.

The parametric control charts used in SPC are always based on the knowledge of the underlying distribution of the data under analysis, and the conventional assumption is that the data is approximately Normally distributed. If the process shows evidence of a significant departure from normality, then the control limits calculated may be entirely inappropriate. The moisture content of kiln-dried lumber does not appear to be well modeled by a normal distribution. It is, therefore, very important to determine the distribution of the moisture content of kiln-dried lumber.

There are a number of reasons why a process such as kiln drying, which is operating in a state of statistical control, yields non-normal distributions. To name a few:

a) Restraining variables at a fixed limit, e.g. preventing the kiln temperature from exceeding a certain set point, preventing the air relative humidity to exceed set parameters, etc.;

b) Measurement of a characteristic that has zero as a natural limit, e.g. moisture content.
1.2 PROBLEM DEFINITION

Although variability is an important parameter to be measured and controlled, many quality control methods suggested in literature are concerned with the average moisture content alone, such as sample estimation of average moisture content for a charge. Other methods are based on “go/no-go” decision criteria, such as acceptance sampling, and conformance tests. These quality control methods help lumber producers to comply with lumber standards, but they do not help to improve the consistency of drying processes. Other methods for consistency, although concerned with the variation of moisture content, did not find practical applicability because they are based on the incorrect assumption that moisture content has a Normal probability distribution.

The variation of moisture content during and after drying could be monitored with the tools of statistical process control (SPC), such as control charts for process averages and process dispersion. One obstacle in the development and use of these tools is the insufficient knowledge of the distribution of moisture content in kiln-dried lumber, which is not well modeled by the customary Normal distribution. Many authors suggest that the Lognormal distribution may offer a good model for the moisture content (McMahon 1961, Zwick and Cook 1985, and Maki 1991). However, suitable procedures for testing this assumption have not been documented yet and therefore are not used in industry.

This thesis presents a methodology for determining the distribution of moisture content by assessing how well a sample of data agrees with a given distribution as its population. A methodology is also developed for constructing control charts to monitor the moisture content of kiln-dried lumber, when the Lognormal is the appropriate underlying distribution. The methods
presented are tested on a data set of actual lumber collected from a production facility in British Columbia, Canada.

1.3 RESEARCH OBJECTIVES

Understanding how well a drying process is operating is very important to mill management. An ideal condition would be to know at all times what is happening in the kiln, with respect to the correct moisture content, its distribution and its variability. If the process is stable or in-control, it should behave in a predictable fashion, and the distribution of moisture content should not change. If the process becomes unstable or out-of-control, the parameters of the distribution of moisture content will change, and this situation should be detected immediately because it will lead to unacceptable quality levels of process performance. Two specific objectives were undertaken in this study, which will lead to methods that correctly determine the distribution of moisture content, and properly interpret this information using statistical process control principles. These objectives are as follows:

1) Develop a methodology for determining the underlying distribution of moisture content in kiln-dried lumber, which is essential for making valid assumptions used in statistical process control.

Conventional control charts for variables are based on the assumption that the quality characteristic of interest is approximately Normally distributed. It has been shown that the moisture content in wood has a non-normal probability distribution. Therefore, the underlying distribution of moisture content in wood must be understood and properly modeled. Moisture content typically has a skewed distribution that could be modeled
with the Lognormal or Weibull distributions, but this assumption was not well justified yet statistically. The approach will be to fit the skewed distribution of MC to other known distributions, such as Lognormal or Weibull;

2) Develop quality control charts that accurately and realistically monitor the variability of moisture content in kiln-dried lumber.

Using the moisture content distribution previously determined, appropriate control charts will be developed that monitor process average and dispersion.
2 LITERATURE REVIEW AND BACKGROUND

MATERIAL

2.1 OVERVIEW

This literature review will examine the concept of moisture content in solid wood, and current methods for testing the goodness of fit of experimental distributions of MC against given probabilistic distributions. General issues concerning statistical process control principles and their application to lumber drying are also discussed.

2.2 MOISTURE CONTENT

The wood of freshly cut trees, also called ‘green’ wood, contains a considerable amount of water. Most of this water has to be removed before the wood is used for various purposes. Also, wood loses or gains moisture in an attempt to reach a state of balance with the surrounding air. Therefore, the knowledge of the relations between wood and its moisture are very important for storage, fabrication and use (Simpson 1991).

The weight of moisture contained in a piece of wood is called the moisture content, and it is usually expressed as a percentage of its oven dry weight. By oven dry weight is meant the constant weight attained by wood samples dried at 103 ±2 °C. It is advantageous to define moisture content in terms of the oven dry weight of wood since the oven dry weight is a constant
value that may be determined at any time. Moisture in wood above the fiber saturation point is described by two types: free water, contained in voids and lumen of the structure, and bound water, contained in the cell-wall material. In most applications, free water has to be removed completely along with some of the bound water, before it can be used for different products. The water is usually removed by drying.

Wood is typically dried in large kilns, which are large chambers designed to provide and control the heat, humidity and air circulation necessary for proper drying. A regular kiln will dry most species of wood to any specified moisture content between 3 and 19 percent in a reasonable short time. One of the main objectives of kiln drying is to bring all the wood to about the same moisture content. This is a difficult task because wood dries at different rates: thicker lumber dries harder, sapwood dries faster than heartwood, quartersawn boards dry slower than flatsawn, etc.

**VARIATION IN MOISTURE CONTENT**

The variation of moisture content in kiln-dried lumber is due to three major factors: the inherent properties of wood as a non-homogenous material, the different rates at which wood dries, and the inconsistencies of drying conditions between different regions of the kiln.

Wood is a non-homogenous material, and the initial moisture content of green wood varies greatly and can range from over 200 percent to 30 percent. These variations occur mostly between different species, but also between the same specie, and even within the same tree (Simpson 1997). In softwood species, sapwood usually contains more water than heartwood. In many species, the butt logs of trees may contain more water than the top logs. Therefore, the
wood enters the kiln with a certain amount of variability of MC. However, Culpepper and Wengert (1980) showed that there is a poor correlation between incoming (green) moisture content and final moisture content – the wettest boards when green were not the wettest boards at the end of drying.

The rate at which wood dries depends on the relative humidity of the surrounding air, the steepness of the moisture gradient, and the temperature of the wood. Drying rate is also influenced by lumber thickness and grain direction. The lower the relative humidity, the faster water moves through capillary cavities. The steeper the moisture gradient (the difference in MC between the inner and outer portions of a board), the faster the drying rate. The higher the temperature of the wood, the faster water will be removed from wood. Drying time increases with thickness, at a rate that is more than proportional to thickness. Finally, flatsawn (also called flat grain) lumber dries faster than quartersawn (or vertical grain) lumber (from Simpson 1997).

The variation of MC in kiln-dried lumber is also caused by inconsistency of drying conditions within the kiln. Culpepper and Wengert (1980) showed that tremendous variation in final moisture content could result from the kiln itself. The identified causes for the variation were variation in air velocity from top to bottom of kiln, and heat supply partially blocked on one side of the kiln, causing over-drying on the other side.

The moisture content variation needs to be estimated at all drying stages (McMahon, 1961): before the actual kiln drying begins to determine the actual setting of the kiln controls; during the kiln drying period to monitor MC distribution and/or change drying schedules; and finally, at the end of a kiln charge to determine whether the lumber has acceptable moisture variation or not.
Variation in final moisture content of lumber can cause serious problems in its subsequent processing and use (Simpson, 1991). Rice and Shepard (1993), and Culpepper and Wengert (1980) showed that substantial MC variation exists between package locations in the kiln, between row locations in the package, between kiln charges, and also between different kilns at the same mill.

To correctly monitor the variation of moisture content it is necessary to know its underlying distribution or at least the best approximation of it.

**DISTRIBUTION OF MOISTURE CONTENT**

Knowledge of the underlying distribution of moisture content would be an advantage for producers of kiln-dried lumber. Parameters such as average MC or standard deviation could be properly estimated from samples. Also, meaningful and reliable parametric control charts and other statistical process control techniques could be employed. Many of the existing methods for parameter estimation, sampling, and control charting, assume that the moisture content obeys the normal law of probability. Many authors agree that the distribution of moisture content of kiln-dried lumber is not well modeled by the Normal distribution (McMahon 1961, Zwick and Cook 1985, Maki 1991, and Noghondarian 1997).

The distribution of MC has a finite left tail, because the moisture content cannot have values less than zero percent. In addition, the distribution is positively skewed – the right-hand side tail is longer (McMahon 1961, Maki 1991). As the lumber becomes drier in a kiln, the skewness of the distribution generally increases (Zwick and Cook 1985). All these considerations suggest that the Lognormal or the Weibull distributions may be capable to model
appropriately the moisture content in kiln-dried lumber. These assumptions should be rigorously checked with statistical methods, such as goodness of fit tests.

### 2.3 GOODNESS OF FIT METHODS

By goodness of fit techniques is meant the methods of examining how well a sample of data agrees with a given distribution as its population. A comprehensive review of these methods is given in D'Agostino and Stephens (1986). In the formal framework of hypothesis testing, the null hypothesis is that a given random variable follows a stated probability law (for example, the Normal distribution or the Lognormal distribution). The goodness of fit techniques applied to test the null hypothesis are based on measuring in some way the conformity of the experimental data to the hypothesized distribution. The techniques usually give formal statistical tests and the measures of consistency are test statistics.

There are many goodness of fit methods, such as graphical analysis, chi-square-type tests, tests based on the empirical distribution function, tests based on regression and correlation, tests based on the third and fourth sample moments, etc. (from D'Agostino and Stephens 1986). However, not all these techniques are well developed for testing all three distributions: Normal, Lognormal, and Weibull, especially when the distribution parameters are not known a-priori and they have to be estimated from samples.
**GRAPHICAL ANALYSIS**

Graphical methods are as exploratory techniques only and should always be used in conjunction with formal numerical techniques. Two graphical methods, which are based on the empirical cumulative distribution function (ECDF), are investigation of symmetry and probability plotting. D'Agostino and Stephens (1986) showed the procedure for investigating the symmetry by means of an ECDF plot and by plotting a scatter diagram of upper half of the ordered observations against the lower. Procedures for Normal and Lognormal probability plotting are given D'Agostino and Stephens (1986). Weibull probability plotting is showed in Dodson (1994).

**NUMERICAL TESTS**

The goodness of fit tests that are well developed for all three distributions under analysis are the tests based on ECDF statistics. They are more powerful than chi-square type tests (D'Agostino 1986), and their procedures and test statistics are better developed for the analysis of the three distributions under study, Normal, Lognormal, and Weibull distribution, especially for the case when the parameters of the hypothesized distribution are not known and have to be estimated from the sample.

The goodness of fit tests based on ECDF are founded on various measures of the discrepancy between the empirical cumulative distribution function $F_n(x)$ and the proposed theoretical cumulative distribution function $F(x)$. The statistics measuring this discrepancy are called ECDF statistics. To decide which statistics should be used, the following considerations were taken into account:
a) The parameters of the underlying distribution are not known. The distribution of moisture content in kiln-dried lumber is not well known, and it will have different parameters for a specific set of conditions (kiln, species, type of lumber, drying schedule, etc.). These parameters must be estimated from the sample, and the goodness of fit theory is not well developed for this situation, except for a few theoretical distributions;

b) Most of the procedures and test statistics for goodness of fit tests were developed for the following situations: $F(x)$ is completely specified; $F(x)$ is the normal distribution with one or more parameters unknown; $F(x)$ is the exponential distribution, with the scale parameter unknown; $F(x)$ is the extreme value distribution, with one or both parameters unknown; $F(x)$ is the Weibull distribution with all parameters unknown (Stephens 1974 and 1976, Chandra 1981, and Lockhart 1994);

c) The probability distributions that are similar with the distribution of moisture content data are the Lognormal and the Weibull. The Normal distribution needs to be considered also, because of its wide acceptance in statistical process control.

Given the above considerations, the following ECDF statistics are used for goodness of fit hypothesis testing: the Kolmogorov-Smirnov $D$ statistic, the Cramer-von Mises $W^2$ statistic, and the Anderson-Darling $A^2$ statistic. Their definitions and formulas are given later in the ‘Methods and Materials’ section.

Once the underlying distribution is determined, control charts can be constructed according to statistical process control principles.
2.4 PRINCIPLES OF STATISTICAL PROCESS CONTROL

If a lumber product is to meet customer requirements from the moisture content point of view, generally it should be produced by a drying process that is stable or repeatable. More precisely, the drying process must be capable of operating with little variability around the target or nominal value of the final moisture content. Statistical Process Control (SPC) is a powerful collection of problem-solving tools useful in achieving process stability and improving capability through the reduction of variability (Montgomery 1997). One of the most technically sophisticated tools of SPC is the control chart.

CONTROL CHARTS

The theory of variability is the basis for the construction and use of control charts. In kiln-drying processes, regardless of how well designed or carefully maintained they are, a certain amount of inherent or natural variability exists. This natural variability, also called background noise, is a stable system of "common causes", which are an inherent part of the process. A process that is operating with only common causes of variation is said to be "in statistical control" (Montgomery 1997).

Other kinds of variability may occasionally be present in the output of a kiln-drying process, namely in the moisture content of lumber. This variability in moisture content usually arises from three sources: improperly adjusted kiln equipment, operator errors, or defective raw material. Such variability is generally large when compared to the background noise, and it usually represents an unacceptable level of process performance. These sources of variability are called "assignable causes", and a process that is operating in the presence of such causes is said
to be “out of control”. A kiln can operate in the in-control state for a long time, producing acceptable moisture content levels in lumber. Eventually, however, assignable causes will occur, resulting in a “shift” to an out-of-control state, where a large proportion of the lumber does not conform to requirements.

A major objective of the control charts is to quickly detect the occurrence of assignable causes of process shifts, so that investigation of the process and corrective action may be undertaken before many nonconforming units are produced. Control charts may also be used to estimate the parameters of the kiln-drying process, and through this information to determine process capability. The control chart also provides useful information to improve the process. In general, the control chart is an effective tool in reducing variability as much as possible.

**CONTROL CHARTS FOR LOGNORMAL VARIABLES**

Conventional control charts for variables are based on the assumption that the underlying distribution of the data is Normal. As it was discussed earlier in the *Distribution of Moisture Content* section, the Lognormal may provide a good fit for the underlying distribution of moisture content. Lognormal control charts were discussed by Ferrell (1958), Morrison (1958) and Joffe and Sichel (1968).

Ferrell (1958) proposes a control chart for monitoring the “geometric midrange”, which he defines as the Lognormal equivalent of the arithmetic midrange from a Normal distribution (half the sum of the largest and smallest data values). However, he uses an equivalent to the ‘3-sigma’ limits for the geometric midrange without showing what the distribution of this parameter
actually is. This control chart is developed *only* for the two-parameter Lognormal. He also does not present any chart for monitoring process dispersion.

Morrison (1958) suggests two control charts for monitoring the sample mean and the sample ratio, respectively. He states that the control limits for a Normal distribution are calculated by processes of addition, subtraction, multiplication, and division. Hence, the corresponding processes for a Lognormal distribution will be multiplication, division, evolution, and involution. The ratio of sample values in a Lognormal distribution corresponds to the range within samples in a Normal distribution. Therefore, the factors used as multipliers of the average range in calculating Normal distribution control limits should be used as indices of the average ratio in the Lognormal distribution. The formulas for calculating the Lognormal limits are given for charts that monitor the sample mean and the sample ratio. However, these limits apply *only* to the two-parameter Lognormal. Also, the measuring units of the control charts are not in the original scale of measurement of data, so the interpretation of these charts by mill personnel would be rather difficult.

Joffe and Sichel (1968) used the arithmetic mean to sequentially testing hypotheses about the mean of a Lognormal distribution. The procedure is essentially an acceptance sampling plan for attributes, and it can be used as an inspection of the final product rather than a real-time monitoring. Also, the procedure applies *only* to the two-parameter Lognormal.

To use the control charts in practical applications, sample sizes must be established along with a method for taking the samples.
SAMPLING AND SAMPLE SIZE

In general, the parameters of a process are unknown; furthermore, they can change over time. Therefore, estimates for these parameters are obtained by statistical inferences based on samples taken from that population. A sample is a portion or a subset of the entire population of measurements under consideration (Milton 1986). The method of sampling used in goodness-of-fit tests is random sampling (D’Agostino and Stephens 1986). Also, Montgomery (1997) clearly advises that samples be randomly selected for SPC procedures, otherwise false conclusions will be derived. Random samples can be taken either with replacement or without replacement. In practical applications, the sampling is done typically without replacement, because the ratio of sample size versus population size is negligible.

From a sample or a collection of samples, inferences are made about the entire population. Therefore, large sample sizes would be ideal; however, the financial cost usually associated with sampling will favour small samples. Consequently, the practitioner will be concerned with the minimum sample size that will provide accurate information. The minimum sample size for goodness of fit tests varies from a situation to another.

For empirical distribution function plots, the typical sample size is at least 100 measurements (D’Agostino and Stephens 1986). For probability plots, the same authors recommend a sample size greater than 50. For a discussion about sample size in probability plotting, see also Daniel and Wood (1971). The minimum sample size for numerical goodness-of-fit tests is not well defined in literature. However, Stephens (1974), Lockhart (1994), and Cohen et al. (1985) illustrated applications of goodness of fit tests with sample sizes ranging from 15 to 100.
Only small sample sizes are required for SPC charts (relative to the size of the population), so sampling without replacement is acceptable for the case of moisture content. For example, a typical kiln (as the one used to dry the lumber for this study) holds over 35,000 lumber pieces (with a nominal size of 50.8mm x 101.6mm x 1.83m, or 2"x4"x6"). For control charts, sample sizes as small as 5 are often used, and therefore the ratio of the sample size vs. population size is negligible (Montgomery 1997).

**STATISTICAL PROCESS CONTROL IN LUMBER DRYING**

Quality control methods have been successfully used and are now ordinary in most high-tech industries. However, the application of statistical process control in the forest products industry, specifically in lumber drying, is delayed well behind other industries. There are three primary reasons for this:

a. Technology development in the past has been based on achieving high production rates of low valued products. Product quality was not a market driver;

b. Wood is a non-homogenous material, and even when it is kiln-dried under controlled conditions, the moisture content does not stabilize or become constant throughout the wood pieces. Due to this large variability, Statistical Process Control (SPC) methods based on inspecting product do not work because of the large volume of product that must be inspected;

c. The mathematical distribution of the moisture content in dried wood is not well understood. Therefore, SPC methods borrowed from other industries do not work well.
Although variability is an important parameter to be measured and controlled, many quality control methods suggested in literature are concerned with the average moisture content alone, such as sample estimation of average moisture content for a charge (Fell and Hill 1980, Rasmussen 1988, and Simpson 1991). Other methods are based on “go/no-go” decision criteria, such as acceptance sampling (Bramhall and Wellwood 1976, and Bramhall and Warren 1977), and conformance tests (Cheung 1994). These quality control methods help lumber producers to comply with lumber standards such as NIST (1999), WWPA (1998), and NLGA (1998), but they do not help to improve the consistency of drying processes.

Other methods for consistency, although concerned with the variation of moisture content, did not find practical applicability because they are based on the incorrect assumption that moisture content has a Normal probability distribution (Pratt 1953 and 1956, Bramhall 1975, and Maki and Milota 1993). This assumption considers the asymmetry and the positive skewness of the MC distribution as “defects” or out-of-control situations, when in fact these are typical outcomes of the drying process (Zwick and Cook 1985, and Maki 1991).

McMahon (1961) proposed Lognormal probability plotting in conjunction with frequency distribution analysis as quality control methods. These techniques were used to estimate the average moisture content and the percentage of MC above or below given limits. The methods are only approximate however, and the possible procedural errors are quite large. Also, the author fails to consider the third parameter of the Lognormal distribution – the threshold.
3 METHODS AND MATERIALS

3.1 OVERVIEW

The study begins with the collection of MC measurements from kiln-dried lumber, from a process that is thought to be in-control. Graphical and numerical methods are used to determine the underlying distribution of MC using goodness-of-fit analysis for three known distributions: Normal, Lognormal, and Weibull. The parameters of the distribution are estimated and appropriate control charts are constructed to monitor process average and dispersion.

The graphical methods employed in this paper are investigation of symmetry and probability plotting. The first method provides a general estimation with respect to the symmetry and skewness of the moisture content distribution. The second method offers a visual determination of how well the data follows each of the three mathematical distributions.

To complement the goodness-of-fit analysis, formal mathematical tests are carried out for each distribution using test statistics based on the empirical distribution function (ECDF).

Generic quality control charts are developed for Lognormal distributed data, which are designed to monitor the process average and dispersion. Specific control limits and charts are then constructed using the collected data. Subsequent samples, collected from later kiln charges, are plotted on the proposed charts, and a comparison is made with conventional charts based on the normality assumption.
3.2 DATA COLLECTION

The lumber used in this study was kiln-dried Douglas-Fir with a nominal section size of 50.8mm x 101.6mm (2"x4"), and a nominal length of 1.83m (6 ft). The lumber came from two different kiln charges that had the same species, dimensions, and drying conditions. The lumber from the first charge was used for distribution analysis and determination of parameters for control charts, and the lumber from the second charge was used for plotting data on the control charts. It is important to mention that the two charges were obtained from a kiln that was operating under normal conditions, using a typical drying schedule. By normal conditions it is meant the typical operating situations of the kiln, without any known errors in the drying schedule and level of drying parameters. Technical details about the kiln are given in Appendix A, and the drying schedule is given in Appendix B.

DATA FOR DISTRIBUTION ANALYSIS

The first kiln charge contained approximately 36,480 boards. Two hundred and twenty-five of these boards were randomly selected without replacement from the first kiln charge and each board was cut up in sections of 1-foot length each, resulting in a total number of 1350 specimens. The ‘1-foot’ length was chosen because of practical applications of thesis’ results. Specifically, ‘1-foot’ is the longitudinal dimension of the measuring field of an in-line moisture meter, and it is intended to apply the results of this research to this type of instruments. The
moisture content of these specimens was determined by oven-drying as described below. Six hundred and seventy-five measurements were selected from the total of 1350 by random sampling with replacement, and they constituted the experimental data for the analysis of the underlying distribution. The procedure of sampling the 675 MC measurements was replicated two more times, to verify the validity of the conclusions drawn from the first sample.

**DATA FOR CONTROL CHARTS**

To determine the control limits and centerlines of the control charts, 20 samples were successively selected from the 1350 measurements of the first charge. Each sample contained 5 randomly selected MC measurements, resulting in a total of 100 measurements. The minimum sample size and the minimum number of samples are recommended by Montgomery (1997). The sampling procedure was 'simple random sampling with replacement'.

To plot moisture content data on the control charts, 225 boards were randomly chosen from the second kiln charge. The charge contained approximately 36,480 boards, and the kiln was operating under typical conditions. The boards were cut in 1-ft long specimens, and their MC was determined by oven-drying. Twenty samples were successively selected from the total of 1350, each sample consisting of 5 MC measurements. The sampling procedure was again 'simple random sampling with replacement'.
DETERMINATION OF MOISTURE CONTENT

The moisture content was determined using the oven-dry method as described in ASTM Standard: D 4442 – 92 (ASTM 1997). To prevent wood moisture loss or gain, the specimens were weighed within three days after the lumber exited the kiln. The initial mass \(A\) of the specimens was determined by weighing them with a scale, which had a minimum readability of 1 mg. The specimens were oven-dried in a forced-convection oven, which maintained a temperature of 103 ± 2 °C. The specimens were considered completely dry when the mass loss in a 3-hr interval was equal to zero, as read by the scale. After oven-drying, the specimens were immediately weighed again, thus obtaining the oven-dry mass \(B\). The moisture content of each sample was calculated using the formula:

\[
\%MC = \frac{A - B}{B} \times 100
\]

(1)
3.3 GRAPHICAL ANALYSIS OF DATA

Two graphical analysis methods were used in this research: investigation of symmetry, and probability plotting. These methods can reveal relationships present in the data, and departures from the assumed models and statistical distributions. D'Agostino and Stephens (1986) recommend that graphical methods precede and complement the numerical methods that are presented in the following section.

As an exploratory technique, the objective of the graphical analysis is to uncover characteristics of the data such as symmetry, skewness, and outliers. This is accomplished by constructing different graphs based on the empirical cumulative distribution function (ECDF).

The ECDF is defined for a set of \( n \) independent observations \( X_1, \ldots, X_n \) with a common distribution function \( F(x) \). The observations are ordered from smallest to largest as \( X_{(1)}, \ldots, X_{(n)} \). The empirical cumulative distribution function, \( F_n(x) \), is defined as

\[
F_n(x) = \begin{cases} 
0, & x < X_{(1)} \\
\frac{i}{n}, & X_{(i)} \leq x < X_{(i+1)} \quad i = 1, \ldots, n-1 \\
1, & X_{(n)} \leq x
\end{cases}
\]  

(2)

\( F_n(x) \) is a step function that takes a step of height \( 1/n \) at each observation. This function estimates the distribution function \( F(x) \). At any value of \( x \), \( F_n(x) \) is the proportion of
observations that is less than or equal to \( x \), while \( F(x) \) is the theoretical probability of an observation that is less than or equal to \( x \).

The use of the ECDF plot does not depend upon any assumptions concerning the underlying parametric distribution and it has some definite advantages over other statistical devices (D’Agostino and Stephens 1986):

a) It is invariant under monotone transformations with regard to quantiles;

b) Its complexity is independent of the number of observations;

c) It supplies immediate and direct information regarding the skewness and bimodality of the underlying distribution;

d) It is an effective indicator of outliers;

e) It supplies robust information on location and dispersion;

f) It does not involve grouping difficulties that arise in using for example, a histogram;

g) It can be used effectively in censored samples.

There is, however, one serious potential drawback with the use of ECDF plots. They can be sensitive to random occurrences in the data and exclusive reliance on them can lead to false conclusions. This is especially true if the sample size is small (D’Agostino and Stephens 1986). Therefore, they should be used in combination with other methods, such as formal numerical tests.
INVESTIGATION OF SYMMETRY

Figure 1 contains plots of three distributions to illustrate different situations one can encounter in attempting to determine if a distribution is symmetric or skewed. The three distributions are the Normal (which is symmetric), the negative exponential (which is "positively skewed": its upper tail is longer than its lower tail, its upper percentage points are farther from the median than are the lower), and the Johnson unbounded curve (which is "negatively skewed": its lower tail is longer than its upper tail, its lower percentage points are farther from the median than are the upper).

![Density and cumulative plots](image)

**Figure 1. Symmetric and skewed distributions**

The ECDF plot is obtained by plotting \( i/n \) as ordinate against the \( i^{th} \) ordered value of the sample as abscissa.

The procedure for constructing the ECDF plot is as follows:
**Procedure 1**: Constructing an ECDF plot

1. Order the observations $X_{(i)}$ ($i=1,...,n$) from smallest to largest;

2. Calculate the ECDF: $F_n(x) = \frac{i}{n}$ for each $X_{(i)}$;

3. Plot $F_n(x)$ versus $X_{(i)}$.

The ECDF plot can be used to assess the skewness of the distribution. If the distribution has positive skewness, the portion of the ECDF for $i/n$ values close to 1.0 (i.e. greater than 0.9) will usually be longer and flatter (almost parallel to the horizontal axis) than the rest of the ECDF. Similarly, if the distribution has negative skewness the long flat portion will lie in the lower end of its ECDF ($i/n$ values lower than 0.1).

The symmetry of a distribution can be assessed by the distance of percentiles to the sample median, and by plotting the upper half of the ordered observations against the lower.

If a distribution is symmetric, then the distance between the median ($50^{th}$ percentile) and any percentile $P$ below the median ($0<P<50$) is equal to the distance from the median to the $(100-P)^{th}$ percentile. The procedure of assessing the symmetry by the distance of percentiles to the sample median is as follows:

**Procedure 2**: Assessing the symmetry by the distance of percentiles to the sample median

1. Order the observations $X_{(i)}$ ($i=1,...,n$) from smallest to largest;
2. Determine the values of \( X(i) \) corresponding to the following percentiles: 10\(^{th}\), 20\(^{th}\), 30\(^{th}\), 40\(^{th}\), 50\(^{th}\), 60\(^{th}\), 70\(^{th}\), 80\(^{th}\), and 90\(^{th}\). The 50\(^{th}\) percentile is the median, and the corresponding value is \( X(50) \);

3. Distances in absolute value from each percentile to the median are calculated by subtracting \( X(50) \) from all \( X(i) \)'s obtained in step 2;

4. These distances are compared for the following pairs of percentiles: 10\(^{th}\) and 90\(^{th}\), 20\(^{th}\) and 80\(^{th}\), 30\(^{th}\) and 70\(^{th}\), 40\(^{th}\) and 60\(^{th}\). For the distribution to be symmetrical, these distances have to be equal for each of the pairs considered;

Symmetry can also be assessed by a scatter diagram of the upper half of the ordered observations against the lower. This graph is obtained by plotting \( X(n+1-i) \) on the vertical axis versus \( X(i) \) on the horizontal axis, for \( i \leq \frac{n}{2} \). A "-1" slope indicates symmetry; a negative slope exceeding 1 in absolute value indicates positive skewness; and a negative slope less than unity in absolute value indicates negative skewness.

The procedure for assessing symmetry by plotting the upper half of the observations against the lower, is as follows:

**Procedure 3:** Assessing symmetry by plotting the upper half of the observations against the lower

1. If the \( X(1), X(2), ..., X(n) \) represent the ordered observations, plot \( X(n) \) versus \( X(1), X(n-1) \) versus \( X(2) \), and in general \( X(n+1-i) \) versus \( X(i) \) for \( i \leq \frac{n}{2} \);
2. Calculate the slope of the plotted data. A slope equal to -1 indicates symmetry. A negative slope exceeding unity in absolute value indicates positive skewness, and a negative slope less than 1 in absolute value indicates negative skewness;

**PROBABILITY PLOTTING**

A major problem with the use of the ECDF plot in attempting to judge visually the correctness of a specific hypothesized distribution is due to the curvature of the ECDF and CDF plots. The CDF (Cumulative Distribution Function) plot is a cumulative plot of the hypothesized distribution function. It is usually very hard to judge visually the closeness of the curved ECDF plot to the curved CDF plot. It is much easier to determine if a set of points deviates from a straight line. A probability plot offers such an opportunity to reach a decision based on visual inspection, because if the hypothesized distribution is the true underlying distribution, then it will be a straight-line plot.

To construct a probability plot, the ordered observations are plotted against functions of the ranks. In general, a probability plot is a plot of \( z_i \) versus \( x_{(i)} \).

\[
z_i = G^{-1}(F(x_{(i)})) \text{ for } i = 1, ..., n
\]  \hspace{1cm} (3)

where: \( G^{-1}(.) \) is the inverse transformation which here transforms \( F(x) \) (the CDF of the hypothesized distribution) into the corresponding standardized value \( z \);

\( x \) represents the observed values of the random variable \( X \);

\( n \) is the sample size.
In equation (3) the standardized $F(x)$ can be calculated using automated functions in spreadsheets, or it can be substituted with $p_i$:

$$F(x_{(i)}) = p_i = \frac{i - 0.5}{n} \quad (4)$$

Once the points are plotted the major task is to judge if the plotted data form a straight line. If they don’t, the task is then to decide what are the properties of the underlying distribution or data that cause this non-linearity.

Methods for probability plotting are constructed for three hypothesized distributions: Normal, Lognormal, and Weibull. The last two distributions were chosen because of their characteristics - finite left tail and a positive skew – which could mimic the moisture content data.

Normal Probability Plotting

Normal probability plotting is the plotting of data in order to investigate the goodness of fit of the data to the Normal distribution.

The Normal distribution has the probability density function

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right] \quad \text{for} \quad -\infty < x < \infty \quad (5)$$

where $\mu$ is the mean and $\sigma$ is the scale parameter.
The cumulative distribution function is

\[ F(x) = \Phi \left( \frac{x - \mu}{\sigma} \right) \]  \hspace{1cm} (6)

where the function \( \Phi \) is the cumulative distribution function of the standard normal variable:

\[ \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp \left( -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right) dx \]  \hspace{1cm} (7)

The \( z \) of equation (4) is approximated by:

\[ z = sign(F_n(x) - 0.5)(1.238t(1 + 0.262t)) \]  \hspace{1cm} (8)

where:

\[ t = \left[ -\ln[4F_n(x)(1 - F_n(x))] \right]^{1/3} \]  \hspace{1cm} (9)

and

\[ sign(F_n(x) - 0.5) = \begin{cases} +1 & \text{if } F_n(x) - 0.5 > 0 \\ -1 & \text{if } F_n(x) - 0.5 < 0 \end{cases} \]  \hspace{1cm} (10)

where \( F_n(x) \) is the empirical cumulative distribution function.

This approximation is given in Hamaker (1978) and appears to be of sufficient accuracy for plotting.
Caution should be exercised when using normal probability plots for samples smaller that 25, for they can show substantial variation and nonlinearity even if the underlying distribution is normal (Daniel and Wood, 1971).

The procedure for Normal probability plotting is as follows:

**Procedure 4: Normal probability plotting**

1. Compute $p_i$ with equation (4), then obtain $z_i$, the inverse of the standard normal cumulative distribution, for each $i = 1, ..., n$. This can be done by using an automated function in spreadsheet software, or by using the approximation formulas (8) through (10);

2. Plot the values of $z_i$ versus $x_i$;

3. If the plotted points form a straight line, then the Normal distribution is the true underlying distribution.

**Lognormal Probability Plotting**

The lognormal distribution has usually two parameters: scale and shape, and a lower bound of zero. However, a third parameter – threshold (or location) – has to be taken into consideration, because the moisture content in wood has a positive lower bound. For kiln-dried lumber, this lower bound is related to the lowest equilibrium moisture content (EMC), which depends on the relative humidity and temperature of the air in the kiln. The three parameter lognormal distribution has the probability density function
\[ f(x) = \frac{1}{x-\theta} \exp \left[ -\frac{1}{2} \left( \ln(x-\theta) - \mu \right)^2 \right] \] for \( x > \theta \)

(11)

where: \( \theta \) is the threshold parameter, \( \mu \) is the scale parameter, and \( \sigma \) is the shape parameter.

The cumulative distribution function is

\[ F(x) = \Phi \left( \frac{\ln(x-\theta) - \mu}{\sigma} \right) \] for \( x > \theta \)

(12)

An important property of the lognormal distribution is that the random variable \( Y = \ln(X - \theta) \) has a normal distribution with mean \( \mu \) and standard deviation \( \sigma \).

Probability plots for this distribution can be constructed by plotting \( z \) from equation (8) on the vertical axis, against \( \ln(x-\theta) \). It is important to note that in order to obtain a straight line, the threshold parameter (or an estimate of it, see Cohen 1951, Cohen et al. 1985, or Crow and Shimizu 1988 for estimation procedures) has to be subtracted from all the data. The subtraction of the threshold parameter is often ignored (Maki 1991, and Noghondarian 1997), thereby leading to incorrect conclusions about the underlying moisture content distribution. To draw correct conclusions from probability plots, they should always be complemented by numerical tests for goodness of fit.

The procedure for Lognormal probability plotting is as follows:

**Procedure 5:** Lognormal probability plotting
1. Estimate the value of the threshold parameter from the data set (estimation procedures are given in Cohen 1951, Cohen et al. 1985, and Crow and Shimizu 1988). Subtract this value from all the data, and then take the natural logarithm; the results will be plotted on the horizontal axis;

2. Calculate $z_i$ using equations (8) through (10);

3. Plot $z_i$ against $\ln(x_i - \theta)$;

4. If the plotted points form a straight line, then the Lognormal distribution is the true underlying distribution.

**Weibull Probability Plotting**

The Weibull distribution has the probability density function

$$f(x) = \frac{c}{\sigma} \left( \frac{x-\theta}{\sigma} \right)^{c-1} \exp \left[ - \left( \frac{x-\theta}{\sigma} \right)^c \right] \text{ for } x > \theta, c > 0$$  \hfill (13)

where: $\theta$ is the threshold parameter, $\sigma$ is the scale parameter, and $c$ is the shape parameter.

The cumulative distribution function is

$$F(x) = 1 - \exp \left[ - \left( \frac{x-\theta}{\sigma} \right)^c \right] \text{ for } x > \theta$$  \hfill (14)
Various methods are available for Weibull analysis (Nelson and Thompson, 1971). D’Agostino (1986) showed that Weibull probability plotting could be achieved simply by plotting $z$ on $\ln(x - \theta)$, where

$$z = \ln(-\ln(1 - F_n(x)))$$

(15)

where: $F_n(x)$ is the Weibull cumulative distribution function.

Before a probability plot can be constructed, an estimate of $F_n(x)$ is needed. The most common estimate is the median rank, a nonparametric estimate of the Weibull cumulative distribution function. Dodson (1994) shows that the median rank is well approximated by the expression:

$$F_n(x_i) = \frac{i - 0.3}{n + 0.4}$$

(16)

where: $i = 1, \ldots, n$ is the sample number;

$n$ is the sample size.

Similar to the three-parameter lognormal distribution, the threshold value must be subtracted from all the data before a Weibull plot will produce a straight line.

The procedure for Weibull probability plotting is as follows:

**Procedure 6: Weibull probability plotting**
a. Estimate the value of the threshold parameter from the data set (for estimation procedures see Cohen 1965 or Dodson 1994). Subtract this value from all the data, and then take the natural logarithm; the results will be plotted on the horizontal axis;

b. Calculate $z_i$ using equations (15) and (16);

c. Plot $z_i$ against $\ln(x_i - \theta)$;

d. If the plotted points form a straight line, then the Weibull distribution is indicated as the true underlying distribution.
3.4 GOODNESS OF FIT TESTS

INTRODUCTION

The general test of fit is a test of

\[ H_0 : \text{a random sample of } n \text{ X-values comes from } F(x; \omega) \]  

(17)

where \( F(x; \omega) \) is a continuous distribution and \( \omega \) is a vector of parameters.

In our case, \( \omega \) is not specified and as a result the null is a composite hypothesis. The alternative hypothesis is also composite – it gives no information on the distribution of the data, and simply states that \( H_0 \) is false. The major focus is on the measure of agreement of the data with the null hypothesis; in fact, it is usually hoped to accept that \( H_0 \) is true (D’Agostino, 1986). The fact that it is usually hoped to accept the null hypothesis and proceed with other analyses as if it were true, sets goodness of fit testing apart from most statistical testing procedures. In usual tests of hypotheses, the null must be rejected in order to prove a point. Here, because the alternative is very vague, the appropriate statistical test will often be by no means clear and no general theory of Neyman-Pearson type (in which the two competing hypotheses have to be precisely defined, along with the exact power and significance levels) appears applicable. Therefore, many different and elaborate procedures have been developed to test the same null hypothesis, without any one emerging as superior. To name a few: tests of chi-squared type, tests based on ECDF (empirical distribution function) statistics, tests based on regression and correlation, and moment techniques.
Given the above considerations, the following ECDF statistics will be used for goodness of fit hypothesis testing: the Kolmogorov-Smirnov $D$ statistic, the Cramer-von Mises $W^2$ statistic, and the Anderson-Darling $A^2$ statistic.

**ECDF STATISTICS**

The **Kolmogorov-Smirnov** statistic assesses the discrepancy between the empirical distribution $F_n(x)$ and the estimated hypothesized distribution $F(x)$. Specifically, it is based on the largest vertical difference between $F_n(x)$ and $F(x)$, and it is computed as the maximum of $D^+$ and $D^-$. $D^+$ is the largest vertical difference between the ECDF and the distribution function when the ECDF is greater than the distribution function. $D^-$ is the largest vertical distance when the ECDF is less than the distribution function.

\[ D^+ = \max_i \left( \frac{i}{n} - U_{(i)} \right) \]  \hspace{1cm} (18)

\[ D^- = \max_i \left( U_{(i)} - \frac{i-1}{n} \right) \]  \hspace{1cm} (19)

\[ D = \max(D^+, D^-) \]  \hspace{1cm} (20)

where $U_{(i)} = F(x_{(i)})$ is the cumulative distribution function value at $x_{(i)}$, the $i^{th}$ ordered value.

The **Cramer-von Mises** statistic ($W^2$) is defined as
\[ W^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x) \]  
\hspace{1cm} (21)

and it is computed as

\[ W^2 = \sum_{i=1}^{n} \left( U_{(i)} - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n} \]  
\hspace{1cm} (22)

The *Anderson-Darling* statistic \( A^2 \) is defined as

\[ A^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \{F(x)[1 - F(x)]\}^{-1} dF(x) \]  
\hspace{1cm} (23)

and it is computed as

\[ A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} \{2i - 1\} [\ln(U_{(i)}) + \ln(1 - U_{(n+1-i)})] \]  
\hspace{1cm} (24)

**GENERAL PROCEDURE FOR TESTS OF HYPOTHESIS**

The general null hypothesis is that the input data values are a random sample from a specified distribution.

To test whether the null hypothesis is true or false, first determine from appropriate tables the statistic ("critical" value) that will cause us to reject the null, then calculate the test statistic ("computed" value) using relations (18) - (20), (22) and (24). Note that modified forms of the statistics might be necessary to be calculated, see the following section for details. If the
"computed" is greater than the "critical", then the null is rejected and the conclusion is that the MC data is not a random sample from the specified distribution.

**TESTS FOR NORMAL DISTRIBUTION**

For a test of Normality, the hypothesized distribution is a Normal distribution function with parameters $\mu$ and $\sigma$ estimated by the sample mean and standard deviation (see equations (4), (5), and (6)). The critical values for $D$, $W^2$, and $A^2$ are given by Stephens (1974). The author has shown how these goodness of fit statistics can be modified so that the critical values for $n = \infty$ can be used for all sample sizes. The following modified, more powerful statistics for $D$, $W^2$ and $A^2$ are recommended:

\[
\tilde{D} = D \left( \sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}} \right) \quad (25)
\]

\[
\tilde{W}^2 = W^2 \left( 1 + \frac{0.5}{n} \right) \quad (26)
\]

\[
\tilde{A}^2 = \left( A^2 + \frac{4}{n} - \frac{25}{n^2} \right) \quad (27)
\]

For different levels of significance, the "critical" values of the ECDF statistics are found in Stephens (1974), "table 1A.3: Modifications for a test of normality $\mu$ and $\sigma$ unknown".

The procedure for tests of hypothesis for the Normal distribution is as follows:
**Procedure 7:** Goodness-of-fit test of hypothesis for the Normal distribution

1. Estimate $\mu$ and $\sigma$ from the sample of $n$ observations using equations (45) and (46);

2. Order the observations, then calculate $F_n(x)$ and $F(x)$ for each observation;

3. Calculate $D$, $W^2$, and $A^2$ using equations (20), (22) and (24). Then compute the modified statistics using (25) through (27);

4. From Stephens (1974), find the critical values for each statistic, for the significance level desired;

5. For each ECDF statistic, reject the null hypothesis if the “computed” statistic exceeds the “critical” value.

**TESTS FOR LOGNORMAL DISTRIBUTION**

For a test of whether the data are from a Lognormal distribution (given by (11) and (12)), the hypothesized distribution is a Lognormal distribution function with parameters $\theta$, $\mu$ and $\sigma$. First, the threshold parameter $\theta$ is estimated from the sample (see Cohen 1951, Cohen et al. 1985, or Crow and Shimizu 1988 for estimation procedures). Then, the scale $\mu$ parameter is estimated from the sample after the logarithmic transformation of the data, $\ln(x-\theta)$. The sample mean of the transformed sample is used as the estimate for the scale parameter. The shape parameter $\sigma$ can be calculated with the formula given by Cohen, 1951. Other sources (SAS 1999) suggest that the shape parameter $\sigma$ can be estimated by the sample standard deviation of the logarithmic-transformed data. The test is therefore equivalent to the test of
normality on the transformed sample, and the same modified statistics from formulas (24) and (25) can be used. The critical values for $D$, $W^2$, $A^2$ are given by Stephens (1974) for different significance levels. The same formulas (24) through (26) for the modified statistics are to be used.

The procedure for tests of hypothesis for the Lognormal distribution is as follows:

**Procedure 8: Goodness-of-fit test of hypothesis for the Lognormal distribution**

1. Estimate the location parameter $\theta$ from the sample (for estimation procedures see Cohen 1951, Cohen et al. 1985, or Crow and Shimizu 1988). Transform the data by $\ln(x - \theta)$. Compute the sample mean and standard deviation of the transformed sample, which will give $\mu$ and $\sigma$ respectively;

2. Calculate $F_n(x)$ and $F(x)$ for each observation from the transformed data;

3. Calculate $D$, $W^2$, and $A^2$ using equations (20), (22) and (24). Then compute the modified statistics using (25) through (27);

4. From Stephens (1974) find the critical values for each statistic, for the significance level desired;

5. For each ECDF statistic, reject the null hypothesis if the "computed" statistic exceeds the "critical" value.
TESTS FOR WEIBULL DISTRIBUTION

For a test of whether the data are from a Weibull distribution, the hypothesized distribution is a Weibull distribution function (see formulas (13) and (14)) with parameters $\theta$, $c$ and $\sigma$ estimated by the maximum-likelihood method (see Cohen 1965 or Dodson 1994 for estimation procedures). In this case, the Kolmogorov-Smirnov $D$ statistic is not as powerful as the other two statistics. Therefore, only the Cramer-von Mises $W^2$ statistic, and the Anderson-Darling $A^2$ statistic will be used for the tests of hypothesis. The critical values for $W^2$ and $A^2$ are obtained by interpolation of simulated values given by Lockhart and Stephens (1994).

The procedure for tests of hypothesis for the Weibull distribution is as follows:

Procedure 9: Goodness-of-fit test of hypothesis for the Weibull distribution

1. Estimate the location, shape and scale parameters from the sample, using the methods presented in Cohen (1965) or Dodson (1994);

2. Calculate $F_n(x)$ and $F(x)$ for each observation for the transformed data;

3. Calculate $W^2$ and $A^2$ using equations (22) and (24);

4. From Lockhart and Stephens (1994) find the critical values for each statistic by interpolation, for the significance level desired;

5. For each ECDF statistic, reject the null hypothesis if the “computed” statistic exceeds the “critical” value.
3.5 CONTROL CHARTS FOR LOGNORMAL DATA

The methodology for estimating distribution parameters is presented for the three-parameter Lognormal distribution, and the control charts for Lognormal data are constructed. The proper variables to be monitored on control charts are also discussed.

THREE-PARAMETER LOGNORMAL DISTRIBUTION

The lognormal distribution function has usually two parameters, scale and shape, and a lower bound of zero. The probability density function for the two-parameter lognormal distribution is

\[ f(x) = \frac{1}{x\sqrt{2\pi} \sigma} \exp \left[ -\frac{1}{2} \left( \frac{\ln(x) - \mu}{\sigma} \right)^2 \right] \text{ for } x > 0 \] (28)

where \( \mu \) is the scale parameter, and \( \sigma \) is the shape parameter.

However, a third parameter – threshold or location – has to be taken into consideration, because the moisture content in wood has a positive lower bound, always greater than zero. For kiln-dried lumber, this lower bound is related to the lowest equilibrium moisture content (EMC), which depends on the relative humidity and temperature of the surrounding air in the kiln. The three parameter lognormal distribution has the probability density function

\[ f(x) = \frac{1}{x-\theta\sqrt{2\pi} \sigma} \exp \left[ -\frac{1}{2} \left( \frac{\ln(x - \theta) - \mu}{\sigma} \right)^2 \right] \text{ for } x > \theta > 0 \] (29)
where $\theta$ is the threshold parameter, $\mu$ is the scale parameter, and $\sigma$ is the shape parameter.

The cumulative distribution function for the three-parameter case is:

$$F(x) = \Phi\left(\frac{\ln(x - \theta) - \mu}{\sigma}\right) \text{ for } x > \theta$$

(30)

The three parameter Lognormal distribution has the mean (from Aitchinson and Brown 1957)

$$\alpha = \theta + \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

(31)

and variance

$$\beta^2 = \exp(2 \mu + \sigma^2) \cdot (\exp(\sigma^2) - 1)$$

(32)

It is important to note that a change in the value of the parameter $\theta$ affects only the location of the distribution, and it does not affect the variance or the shape (Johnson et al. 1994).

There is a very important connection between Lognormal and Normal distributions. The Lognormal distribution in its simplest form may be defined as the distribution of a variable whose natural logarithm obeys the normal law of probability. In other words, if a variable $X$ is distributed Lognormal with a threshold parameter $\theta$, a scale parameter $\mu$, and a shape parameter $\sigma$, then the variable $Y = \ln(X - \theta)$ has a normal distribution with mean $\mu$ and standard deviation $\sigma$. This property of the Lognormal is very useful in quality control work,
because the methods of statistical process control are well known and widely applied for the Normal case.

Although many of the properties of the Lognormal may immediately be derived from those of the Normal distribution, there are certain features of the former that differ from anything arising in Normal theory. One example is that the mean and the variance of the Lognormal distribution are not parameters of the distribution, contrasting with the Normal case, where the parameters of the distribution are the mean and the variance. Another example is that, unlike the Normal distribution, the Lognormal is not uniquely determined by its moments (Heyde 1963). Also, when the threshold value is unknown and it has to be estimated from the sample, this complicates the estimation procedures developed for the two-parameter case (Aitchison and Brown 1957).

WHAT TO MONITOR ON CONTROL CHARTS

Control charts usually monitor process average and process dispersion. The process average is typically checked with a control chart for means, and the dispersion is usually monitored with a control chart for the standard deviation, or for the range. For Normally distributed quality characteristics, the mean and the standard deviation are also the parameters of the distribution, and are measures of central tendency and dispersion, respectively. For moisture content in kiln-dried lumber, whose distribution seems to be better modeled by the Lognormal, the mean, standard deviation, or variance are not parameters of the distribution. The question that arise for Lognormal variables is, should the mean and standard deviation be monitored, or the threshold, scale and shape parameters?
One approach would be to monitor the average and the standard deviation of the moisture content, which is a standard practice in quality control applications. For large samples, conventional control charts for sample averages could be constructed regardless of the Lognormal assumption, because of the central limit theorem, which makes sample averages to be approximately Normally distributed. However, the distributions of the standard deviation and the variance of a Lognormal variable are not well known, so conventional methods such as “3-sigma” or probability limits cannot be employed here. Another reason for not monitoring the “average” is that, for Lognormal distributions, a good measure of central tendency is not the arithmetic mean, but rather the geometric mean. This subject will be detailed a latter section.

This paper proposes a second approach, which is to monitor directly the parameters of the distribution. This consideration is based on the close relationship between the Lognormal and the Normal distributions. As it was discussed earlier, if a variable $X$ is distributed Lognormal with a threshold parameter $\theta$, a scale parameter $\mu$, and a shape parameter $\sigma$, then the variable $Y = \ln(X - \theta)$ has a normal distribution with mean $\mu$ and standard deviation $\sigma$. Monitoring the mean and standard deviation of the Normal variable $Y$ would actually control the scale and shape parameters of the Lognormal variable $X$. The threshold parameter would not be monitored directly with a specific chart. However, inferences could be made about a change in the threshold value from the other two charts. For example, if a shift in the scale parameter occurs, it could be an indication that the threshold value also changed, especially if the shape remains the same. Similarly, if a shift occurs in the shape parameter and the scale stays the same, it could imply that the threshold value changes also.
Suppose that the moisture content $X$ is distributed Lognormal with known threshold, scale, and shape parameters. If $x_1, x_2, ..., x_n$ is a sample of size $n$, we transform it to normality, by subtracting the value of $\theta$ from all $x_i$ and then taking the natural logarithm:

$$\ln(x_i - \theta) = y_i$$  \hspace{1cm} (33)

According to the definition of the Lognormal distribution, the new variable $Y$ is distributed Normal with mean $\mu$ and standard deviation $\sigma$. The average of the transformed sample is:

$$\bar{y} = \frac{y_1 + y_2 + ... + y_n}{n}$$  \hspace{1cm} (34)

It is known that $\bar{y}$ is normally distributed with mean $\mu$ and standard deviation $\frac{\sigma}{\sqrt{n}}$, for moderately large sample sizes (Montgomery 1997). The conventional “3-sigma” control limits, and the center line for a control chart for sample means $\bar{y}$ are:

$$\begin{align*}
\text{Center Line}_N &= \mu \\
\text{UCL}_N &= \mu + 3\frac{\sigma}{\sqrt{n}} \\
\text{LCL}_N &= \mu - 3\frac{\sigma}{\sqrt{n}}
\end{align*}$$  \hspace{1cm} (35)

The subscript “N” in equations (35) refers to the fact that the control limits are developed for Normal distributed data.
In practice, we usually will not know a priori $\theta$, $\mu$ and $\sigma$, the parameters of the Lognormal distribution. Therefore, they must be estimated from preliminary samples taken when the process is thought to be in-control. In the following section we show how to estimate the parameters of the underlying Lognormal distribution of moisture content.

**PARAMETER ESTIMATION**

The knowledge of the threshold parameter $\theta$ is critical for the choice of methods employed to estimate the other two lognormal parameters. If the threshold is known, the estimation methods are well developed and straightforward. If the threshold is *not* known, and it has to be estimated from previous data, the methods involved in parameter estimation are much more complex.

The threshold is said to be 'known' when it can be determined a priori by reference to the generating system. The minimum moisture content that the lumber can possibly have during drying is given by the equilibrium moisture content (EMC) set up in the kiln. This assumes, of course, that all the lumber entering the kiln has moisture content greater than the initial EMC. The temperature and relative humidity of air in the kiln determine the EMC. However, in order to force water out of the wood, the drying schedules are maintained in a way that does not allow the wood to attain the EMC set by the schedule, so the threshold parameter is related to EMC, but this relationship is not known. Therefore, the threshold parameter cannot be determined solely by reference to the generating system, and will have to be estimated from samples.

The fact that $\theta$ is unknown creates complications in the estimation of Lognormal parameters. A great deal of research concerning the three-parameter Lognormal was published
(see Crow and Shimizu 1988, p.113, for an exhaustive discussion). In general, global maximums of the likelihood function are used to estimate distribution parameters. However, Heyde (1963) demonstrated that the three-parameter Lognormal is not uniquely determined by its moments and this raised questions about moment estimators. Also, Hill (1963) has shown that global maximum likelihood estimators lead to inadmissible estimates regardless of the sample. Therefore, other estimators have to be used that at least produce reasonable estimates. As an alternative to global maximum likelihood estimators (MLE), local maximum likelihood estimators (LMLE) are generally accepted for the estimation of the threshold, scale, and shape parameters, especially for moderately large sample sizes (Cohen 1951, Cohen et al. 1985, and Crow and Shimizu 1988).

Given a sample of moisture content data, the first parameter that can be estimated is the threshold. For an ascending ordered sample \( x_1, x_2, ..., x_n \), Cohen (1951) proved the following local maximum likelihood estimator for \( \theta \):

\[
F(\theta) = \left[ \sum_{i=1}^{n} \left( \frac{1}{x_i - \theta} \right) \right] \left[ n * \sum_{i=1}^{n} \ln(x_i - \theta) - n * \sum_{i=1}^{n} \ln^2(x_i - \theta) + \left( \sum_{i=1}^{n} \ln(x_i - \theta) \right)^2 \right]
\]

\[ -n^2 \sum_{i=1}^{n} \frac{\ln(x_i - \theta)}{x_i - \theta} = 0 \]

(36)

It is convenient to use the "trial and error" technique with linear interpolation to solve this equation. A first approximation \( \theta_1 < x_j \) is chosen and \( F(\theta_1) \) is evaluated. If \( F(\theta_1) \) is zero, then no further calculations are required. Otherwise it is continued until a pair of values \( \theta_i \) and \( \theta_j \) is found in a sufficiently narrow interval such that \( F(\theta_i) > 0 > F(\theta_j) \) or \( F(\theta_i) < 0 < F(\theta_j) \).
and the final estimate \( \hat{\theta} \) is found by interpolation. When solving (36) for \( \hat{\theta} \), only values for which \( 0 < \hat{\theta} < x_i \) are accepted.

After obtaining the local maximum likelihood estimate \( \hat{\theta} \) for the threshold, the estimates for the other two parameters, scale and shape, can be determined with the following relations (from Cohen 1951):

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i - \hat{\theta})
\]

(37)

\[
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \ln^2(x_i - \hat{\theta}) - \left[ \frac{1}{n} \sum_{i=1}^{n} \ln(x_i - \hat{\theta}) \right]^2
\]

(38)

It can be seen from equation (37) that the scale parameter \( \mu \) is estimated by the sample mean of the logarithmic-transformed data, \( \ln(x - \hat{\theta}) \).

Sometimes this estimation technique does not yield a valid result for the threshold, because local maximums of the likelihood function do not always exist, especially in small samples (Cohen et al. 1985). If this is the case, the parameters of the distribution can be estimated from the sample by an alternative technique, based on modified moment estimators (MME). This methodology is given in Cohen et al. 1985.

In practical applications, once the threshold \( \hat{\theta} \) is calculated for a kiln and a set of conditions, it will be assumed as constant until further proof that it has changed. The other two parameters are estimated from each sample using equations (37) and (38).
ARITHMETIC MEAN VS. GEOMETRIC MEAN

It is known that the arithmetic mean quantifies the central tendency of Normal variables. When the observations are not distributed Normal, but their natural logarithms are, then the geometric mean should be used. This is true for the two-parameter Lognormal distribution. For the Lognormal variable \( x_i \ (i = 1, 2, \ldots, n) \), \( y_i = \ln(x_i) \) is a Normal variable. With a few algebraic operations, the average of \( y_i \) (called \( \bar{y} \)) becomes:

\[
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i) = \frac{1}{n} \ln\left( \prod_{i=1}^{n} x_i \right) = \ln\left( \left( \prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} \right) \tag{39}
\]

It is known that

\[
\left( \prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} \tag{40}
\]

is the geometric mean of \( x_i \) (further called "geomean(x_i)").

Equation (39) becomes:

\[
\bar{y} = \ln(\text{geomean}(x_i)) \tag{41}
\]

If we calculate the average of \( y_i \) and we want to express it in the original units of \( x_i \), then from \( x_i = \exp(y_i) \) we have:

\[
\text{anti log}(\bar{y}) = \exp(\bar{y}) = \exp[\ln(\text{geomean}(x_i))] = \text{geomean}(x_i) \tag{42}
\]

In other words, the geometric mean of a Lognormal variable corresponds to the arithmetic mean of the log of that variable (which is a Normal variable). Hence, for the two-
parameter Lognormal the geometric mean should be used as a measure of central tendency, rather than the average (arithmetic mean).

For the three-parameter case, \( y_i = \ln(x_i - \theta) \) is a Normal variable, and the average \( \bar{y} \) is given by:

\[
\bar{y} = \ln(\text{geomean}(x_i - \theta))
\]  

(43)

Which is equivalent to:

\[
\exp(\bar{y}) = \text{geomean}(x_i - \hat{\theta})
\]  

(44)

This means that the geometric mean of the moisture content less the threshold should be actually monitored on a chart in the original scale of measurements.

Once the parameters of the underlying distribution are estimated for the in-control process, the control limits and center lines of the control charts can be determined.

**CONTROL CHART FOR SCALE PARAMETER (\( \bar{Y} \) CHART)**

In quality control literature the customary chart for monitoring sample averages is usually called "\( \bar{x} \) chart". To adopt this nomenclature for our case, we will call "\( \bar{y} \) chart" the chart that monitors sample averages of the transformed data – the \( Y \) variable – which is distributed Normal. It was shown earlier that these sample averages are equivalent to sample scale parameter of the Lognormal variable.
If \( m \) samples each of size \( n \) are used as preliminary data, and \( \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_m \) are the
averages of each sample – identical with the scale parameter of each sample, calculated with
equation (37) – then the scale parameter of the Lognormal distribution is given by

\[
\bar{y} = \frac{\bar{y}_1 + \bar{y}_2 + \ldots + \bar{y}_m}{m} \tag{45}
\]

Similarly, if \( S_i \) is the standard deviation of the \( i^{th} \) sample, then

\[
\bar{S} = \frac{S_1 + S_2 + \ldots + S_m}{m} \tag{46}
\]

The relationship between the standard deviation (of the Normal variable \( Y \)) and the shape
parameter (of the Lognormal variable \( X \)) follows from the (biased) maximum likelihood
estimator of the variance \( \sigma^2 \) for the Normal variable \( Y \) (from Johnson et al. 1994, and
Montgomery 1997):

\[
S^2 = \frac{\sum_i^n (y_i - \bar{y})^2}{n} = \frac{\sum_i^n y_i^2 - \left( \sum_i^n y_i \right)^2}{n} \tag{47}
\]

If we substitute \( \ln(x_i - \hat{\theta}) = y \), in \( (47) \) the equation becomes:

\[
S^2 = \frac{1}{n} \sum_i^n \ln^2(x_i - \hat{\theta}) - \left[ \frac{1}{n} \sum_i^n \ln(x_i - \hat{\theta}) \right]^2 \tag{48}
\]

which is identical in expression with equation (38). However, the standard deviation of
the sample, \( S \), is not an unbiased estimator of \( \sigma \). To obtain an unbiased estimator of \( \sigma \), \( S \) must
be divided by \( c_4 \), a constant that depends on the sample size. Therefore, \( S/c_4 \) is to be used for estimating the standard deviation \( \sigma \) of the Normal population, which is equivalent to estimating the shape parameter \( \sigma \) of the Lognormal population.

Using the usual notations for control charts, the parameters of the control chart for sample means – or scale parameters – \( \bar{y} \) (called “\( \bar{y} \) chart”) become:

\[
\begin{align*}
\text{UCL}_N &= \bar{y} + 3\frac{\bar{S}}{c_4 \sqrt{n}} \\
\text{Center Line}_N &= \bar{y} \\
\text{LCL}_N &= \bar{y} - 3\frac{\bar{S}}{c_4 \sqrt{n}}
\end{align*}
\]

(49)

where:

\( n \) = the size of the samples that will be used for subsequent plotting;

\( \bar{y} \) = estimate of the Normal population mean, obtained from preliminary data, given by equation (45); it is also an estimate of the Lognormal population scale parameter,

\( \bar{S} \) = an estimate of the standard deviation of the Normal population, given by equation (46); it is also an estimate of the Lognormal population shape parameter \( \sigma \);

\(^1\) For a detailed discussion about \( c_4 \) and its values see Johnson et al. 1994, or Montgomery 1997.
\[ c_4 = \text{a constant that depends on the sample size; this constant is introduced because } \bar{S}/c_4 \text{ is an unbiased estimator of the population standard deviation; for } n = 5, c_4 = 0.940 \ (\text{Johnson et al. 1994, or Montgomery 1997}). \]

This chart could be used as it is. However, it will not provide meaningful information, for the values plotted are dimensionless, because of the logarithmic transformation. To solve this problem, a chart can be constructed which plots the data in the original scale. The control limits and center line for the original scale are obtained by adding the threshold and then taking the antilog (by exponentiation). From \( Y = \ln(X - \theta) \) we have \( X = \theta + \exp(Y) \), and we will use this relationship to transform all the results back to the original scale of measurements.

\[
\begin{align*}
\text{UCL}_{LN} &= \hat{\theta} + \exp(UCL_N) \\
\text{Center Line}_{LN} &= \hat{\theta} + \exp\left(\text{Center Line}_N\right) \\
\text{LCL}_{LN} &= \hat{\theta} + \exp(LCL_N)
\end{align*}
\tag{50}
\]

The subscript "LN" in equations (50) refers to the fact that the control limits are developed for the Lognormal distributed data, which is the moisture content.

By substituting (49) in (50) the limits of the original-scale control chart (further called "antilog \( \bar{y} \) chart") become:

\[
\begin{align*}
\text{UCL}_{LN} &= \hat{\theta} + \exp\left(\bar{y} + 3*\frac{\bar{S}}{c_4\sqrt{n}}\right) \\
\text{Center Line}_{LN} &= \hat{\theta} + \exp(\bar{y}) \\
\text{LCL}_{LN} &= \hat{\theta} + \exp\left(\bar{y} - 3*\frac{\bar{S}}{c_4\sqrt{n}}\right)
\end{align*}
\tag{51}
\]
Once these control limits are established, they can be used to monitor subsequent kiln charges that are obtained under the same drying conditions. To plot a point on the $\bar{y}$ chart, a sample $x_i (i = 1, 2, \ldots, n)$ of MC measurements is taken, then $x_i$ is transformed to a normal variable $y_i = \ln(x_i - \hat{\theta})$. The sample average of $y_i$, $\bar{y}$, is then plotted on the $\bar{y}$ chart. To plot the corresponding point on the antilog $\bar{y}$ chart, $\bar{y}$ needs to be transformed to the original scale of measurements, with the equation:

$$\text{anti log}(\bar{y}) = \hat{\theta} + \exp(\bar{y}) \quad (52)$$

At the first sight, plotting $\text{anti log}(\bar{y})$ on the antilog $\bar{y}$ chart seems to be just an algebraic manipulation. However, it was demonstrated earlier that $\exp(\bar{y})$ is in fact the geometric mean of $(x_i - \theta)$.

From (44) it follows that the point plotted on the antilog $\bar{y}$ chart is:

$$\text{anti log}(\bar{y}) = \hat{\theta} + \text{geomean}(x_i - \hat{\theta}) \quad (53)$$

This means that the geometric mean of the MC data less the threshold is actually monitored on the antilog $\bar{y}$ chart. The threshold is added back in just to give the right scale to the chart.

The following is a procedure for constructing the charts for sample means (scale parameters) and the corresponding original-scale control charts:
Procedure 10: Construction of charts for sample means – or scale parameters – (\( \bar{y} \) chart) and the corresponding Lognormal control charts in original scale of measurements (antilog \( \bar{y} \) chart)

1. For each kiln and set of conditions (species, type of lumber, stacking method, drying schedule), collect \( m \) random samples of \( n \) moisture content measurements, from when the process is thought to be in-control (it is recommended that \( m \geq 20 \); \( n \) can be as low as 5 – Montgomery 1997);

2. Determine if the moisture content is distributed Lognormal, using the methods described in ‘Graphical Analysis’ and ‘Goodness Of Fit Tests’ sections. If the Lognormal is not a good model for the data, then other appropriate methods should be used. This procedure assumes hereafter that the MC is distributed Lognormal;

3. From the preliminary data estimate the threshold parameter using equation (36). The threshold should be estimated from the entirety of preliminary data, instead as average of sample estimates, because of estimation problems explained above. With \( \hat{\theta} \) known, the other two parameters are estimated from the transformed samples, using (45) and (46);

4. Establish a sample size \( n \) to be used in subsequent plotting, then calculate the center lines and the control limits for both charts using relations (49) and (51). Equations from (49) define the chart for the sample averages (or scale parameter), the “\( \bar{y} \) chart”, which is used to monitor the lognormal-transformed data. Equations from (51) describe the corresponding chart that monitors the moisture content in the original scale of measurements, the “antilog \( \bar{y} \) chart”;
5. To plot the $\bar{y}$ data for a subsequent sample of size $n$, first normalize $x_i$ by subtracting $\hat{\theta}$ from each value, and then taking the natural logarithm, $\ln(x_i - \hat{\theta})$. Calculate the mean of the transformed sample, $\bar{y}$, which is plotted on the "$\bar{y}$ chart";

6. To plot the corresponding value on the "antilog $\bar{y}$ chart", the result needs to be brought back to the original scale of measurements, with the expression: $\text{anti log}(\bar{y}) = \hat{\theta} + \exp(\bar{y})$.

**CONTROL CHART FOR SHAPE PARAMETER ($S$ CHART)**

To monitor the Lognormal shape parameter (or the sample standard deviation of the transformed data), the "3-sigma" limits cannot be used because the sample standard deviation is not Normally distributed, not is the sample variance. Therefore, probability limits are proposed to construct control charts for sample standard deviation. The construction of these charts is based on the fact that the variance of the Normal variable $Y$ is ‘chi-square’ distributed.

$$\text{If } Y \sim N(\mu, \sigma^2), \text{then } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$  \hspace{1cm} (54)

where $Y$ is the normal variable with mean $\mu$ and variance $\sigma^2$, $S^2$ is the sample variance, and $\chi^2_{n-1}$ is the chi-square distribution with $n-1$ degrees of freedom. A $100(1-\alpha)\%$ two sided confidence interval on the variance is:

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$  \hspace{1cm} (55)
where \( \chi^2_{\alpha/2,n-1} \) denotes the percentage point of the chi-square distribution such that \( P\{\chi^2_{n-1} \geq \chi^2_{\alpha/2,n-1}\} = \alpha/2 \). The control limits for this chart, called the “S chart”, are calculated as follows:

\[
\begin{align*}
\text{UCL}_N &= \frac{\overline{S}}{c_4} \sqrt{\frac{\chi^2_{\alpha/2,n-1}}{n-1}} \\
\text{Center Line}_N &= \frac{\overline{S}}{c_4} \\
\text{LCL}_N &= \frac{\overline{S}}{c_4} \sqrt{\frac{\chi^2_{1-\alpha/2,n-1}}{n-1}}
\end{align*}
\]

(56)

where \( \overline{S} \) is an estimate of the standard deviation of the Normal population, given by equation (46), and \( c_4 \) is the constant that depends on the sample size, described earlier.

To construct this chart in the original scale of measurements, the above control limits just need to be exponentiated, without adding the threshold parameter. As it was explained earlier, the Lognormal shape parameter (the standard deviation of the transformed data) does not change when the threshold value changes. The parameters of the original-scale control chart (called the “antilog S chart”) become:

\[
\begin{align*}
\text{UCL}_{LN} &= \exp\left(\frac{\overline{S}}{c_4} \sqrt{\frac{\chi^2_{\alpha/2,n-1}}{n-1}}\right) \\
\text{Center Line}_{LN} &= \exp\left(\frac{\overline{S}}{c_4}\right) \\
\text{LCL}_{LN} &= \exp\left(\frac{\overline{S}}{c_4} \sqrt{\frac{\chi^2_{1-\alpha/2,n-1}}{n-1}}\right)
\end{align*}
\]

(57)
Once these control limits are established, they can be used to monitor subsequent kiln charges that are obtained under the same drying conditions. The sample collected earlier for the \( \bar{y} \) chart is used, and the standard deviation \( S \) of the transformed sample is plotted on the \( S \) chart. To plot the corresponding point on the antilog \( S \) chart, \( S \) needs to be transformed to the original scale of measurements by exponentiation.

The following is a procedure for constructing the charts for Lognormal shape parameter (or Normal sample standard deviation) and the corresponding original-scale control charts:

**Procedure 11**: Construction of charts for shape parameter – sample standard deviations - \( (S \) chart) and the corresponding Lognormal control charts in original scale of measurements (antilog \( S \) chart)

1. Using the results from steps 1 to 4 of Procedure 10, calculate the center lines and the control limits using relations (56) and (57). Equations from (56) define the chart for the shape parameter (or sample standard deviation), called the “\( S \) chart”, which is used to monitor the variation of lognormal-transformed data. Equations from (57) describe the corresponding chart in the original scale of measurements, called the “antilog \( S \) chart”;

2. To plot a subsequent sample of size \( n \), calculate the standard deviation \( S \) of the transformed sample, \( \ln(x_i - \hat{\theta}) \), and then plot this value on the \( S \) chart;

3. To plot the corresponding value on the antilog \( S \) chart, the result needs to be brought back to the original scale of measurements, just by exponentiating: \( \text{antilog}(S) = \exp(S) \);
4 RESULTS AND DISCUSSION

4.1 OVERVIEW

The probabilistic distribution of the experimental moisture content data was analyzed both graphically and numerically. Results of two graphical methods – analysis of symmetry and probability plotting – are presented in the form of ECDF plots, symmetry tables and charts, and probability plots for the three distributions investigated: Normal, Lognormal, and Weibull. Furthermore, results of the hypothesis tests are showed for the numerical goodness-of-fit methods, by comparing test statistics with critical values. Finally, the true underlying distribution of moisture content is determined by consolidating the results of these methods.

4.2 GRAPHICAL ANALYSIS

The symmetry of the moisture content distribution was first assessed by a plot of the empirical cumulative distribution function (ECDF). The moisture content measurements were ordered from smallest to largest, and then the empirical cumulative distribution function was calculated for each moisture content value. The ECDF plot, presented in Figure 2, shows that the moisture content distribution has a positive skewness, because the portion of the ECDF for i/n values greater than 0.9 is longer and flatter (almost parallel to the horizontal axis) than the portion of the ECDF for i/n values smaller than 0.1.
Figure 2. Empirical Cumulative Distribution of MC for Douglas-Fir, and the relation of 10th and 90th percentiles about the median (50th percentile).

The symmetry of the empirical distribution of MC was also analyzed by comparing the absolute distances of percentiles from the median. The following pairs of percentiles were calculated: 10th and 90th, 20th and 80th, 25th and 75th, 40th and 60th. The distances of the percentiles from the median (the 50th percentile) are presented in Table 1, for all the pairs of percentiles. These distances are not equal for each pair, proving that the distribution of MC is not symmetric. The table also shows that all the upper percentiles are “further away” from the median than their corresponding lower percentile. This supports the observation that the distribution is positively skewed.
Table 1. Absolute values of distances from sample median to percentiles

The symmetry was finally assessed by a plot of upper vs. lower observations (Figure 3).

Figure 3. Plot of upper vs. lower observations for investigation of symmetry.
The slope of the plotted observations had a value of \(-1.67\). Following the considerations presented in the ‘Methods’ section, a slope greater than 1 in absolute value also indicates positive skewness of the moisture content distribution.

The conclusion of the symmetry investigation of our experimental data is that the moisture content distribution is not symmetrical, and it is positively skewed. This supports the idea that the Normal distribution is not a good model for the moisture content, which might be better modeled by the Lognormal and/or Weibull. The next step of the graphical analysis was to construct probability plots to visually assess the correctness of these hypothesized distributions.

The Normal probability plot in Figure 4 was constructed using equations (4) and (8) – (10), by plotting the inverse standard normal cumulative against each value of MC. Visual examination of this plot shows that the moisture content data does not follow a straight line, indicating that the Normal distribution does not provide a good fit.

A Lognormal probability plot is shown in Figure 5. It is important to note that, in order to obtain a straight line, the threshold (location) value of 7.661 was subtracted from the data. The plot suggests that the Lognormal provides a good model for the distribution of the moisture content.

The Weibull probability plot shown in Figure 6 indicates that the hypothesized Weibull distribution does not appear to be the true underlying distribution for the moisture content data.
Figure 4. Normal probability plot for the moisture content data.

Figure 5. Lognormal probability plot for moisture content data
4.000
-10.000
-12.000
\ln[x(i) - \text{location}]

Figure 6. Weibull probability plot for moisture content data

The symmetry analysis and the probability plots indicated that the distribution of the moisture content appears to be better modeled by the Lognormal distribution. To substantiate these findings, the graphical analysis was followed by more formal numerical tests for goodness of fit.

4.3 GOODNESS OF FIT TESTS

Three hypothesis tests were conducted for each of the distributions under investigation. The null hypothesis for the goodness of fit tests was that the sample of 675 moisture content values comes respectively from Normal, Lognormal, or Weibull. Test statistics based on the
empirical distribution function were calculated for each hypothesized distribution, and then compared with the critical values.

To test whether our sample came from a Normal distribution or not, the values for $D$, $W^2$, and $A^2$ were calculated using equations (20), (22) and (24). Because the parameters of the Normal distribution had to be estimated from the sample, the modified statistics $\tilde{D}$, $\tilde{W}^2$, $\tilde{A}^2$ were computed using equations (25) through (27). The critical values of these statistics were obtained from Stephens (1974), for significance levels ($\alpha$) of 0.01 and 0.05. The parameters for the Normal distribution, estimated from the sample, were: mean = 14.58 and standard deviation = 1.645. The computed and critical values for each statistic are given in Table 2. For all three goodness of fit statistics, the observed statistic was greater than the critical value, therefore the null hypothesis was rejected, and it was concluded that MC data does not come from a normal distribution. This result is significant at both 1% and 5% levels. This conclusion is consistent with the findings from the graphical analysis.

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Normal Test</th>
<th>Lognormal Test</th>
<th>Weibull Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>$D$</td>
<td>1.566</td>
<td>1.035</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>$\tilde{W}^2$</td>
<td>0.629</td>
<td>0.178</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>$\tilde{A}^2$</td>
<td>3.996</td>
<td>1.092</td>
</tr>
</tbody>
</table>

Table 2. Observed statistics and critical values for ECDF statistics for hypothesis testing.
To conduct the lognormal test, the parameters of the lognormal distribution were estimated from the sample: threshold = 7.661, scale = 1.907, and shape = 0.236. The modified tests statistics given by equations (25) through (27) were used because of the parameter estimation. Table 2 shows the computed values for $\tilde{D}$, $\tilde{W}^2$, $\tilde{A}^2$ and the critical values - from Stephens (1974) - for significance levels ($\alpha$) of 0.01 and 0.05. For all three goodness of fit statistics, the observed statistic was smaller than the critical value, and therefore the null hypothesis is not rejected. Taking into account the considerations detailed in the Methods section, the null hypothesis was accepted. At $\alpha = 0.01$ and 0.05 significance levels, all three tests support the conclusion that the three-parameter Lognormal distribution with threshold parameter $\theta = 7.661$, scale parameter $\mu = 1.907$, and shape parameter $\sigma = 0.236$ provides a good model for the distribution of moisture content. This is consistent with the conclusion from the graphical analysis.

For the Weibull goodness of fit test, the Cramer-von Mises $W^2$, and the Anderson-Darling $A^2$ test statistics were calculated using equations (22) and (24). The critical values were found by interpolation from Lockhart and Stephens (1994), for significance levels ($\alpha$) of 0.01 and 0.05. These values are given in Table 2. The parameters used for Weibull distribution, estimated from the sample, were: threshold = 10.867, scale = 4.197, and shape = 2.401. For both $W^2$ and $A^2$ goodness of fit statistics, the observed statistic was greater than the critical value. Therefore, the null hypothesis was rejected, and it was concluded that MC data does not come from a Weibull distribution. This result is significant at both 1% and 5% levels. The conclusion is consistent with the findings from the graphical analysis.
All the above hypothesis tests were replicated two more times, for two other samples of 675 measurements, to confirm the validity of the results. The results of both tests, presented in Appendix D, confirmed the conclusions of the tests performed on the first sample of 675 measurements.

The numerical goodness of fit tests confirmed that the Lognormal provides a good model for the distribution of moisture content data, whereas Normal and Weibull do not.

4.4 CONTROL CHARTS FOR LOGNORMAL VARIABLES

The charts proposed in this paper are based on the assumption that the moisture content has a Lognormal distribution. On the other hand, in many practical applications the moisture content is assumed to have a Normal distribution. To demonstrate the difference between these two approaches, the control charts are constructed first by assuming a Normal distribution, and then by assuming the Lognormal.

The preliminary data, consisting of 20 samples of 5 measurements each, was used to estimate the parameters of the distribution. These parameters are given in Table 3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>$\bar{x}$</td>
<td>14.264</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$s$</td>
<td>1.626</td>
</tr>
</tbody>
</table>

Table 3. Parameters of preliminary data for construction of control charts based on the Normal distribution

If the underlying distribution were assumed to be Normal, then the customary control charts based on this assumption would have the control limits given in Table 4.
Table 4. Control limits and center lines for control charts based on Normal assumption

In the construction of these charts, 3-sigma limits were employed for the $\bar{x}$ chart NORMAL, and probability limits were used for the $S$ chart NORMAL. The charts based on the Normality assumption are shown in Figure 7 and Figure 8. The points plotted on these charts came from 20 samples randomly selected from a subsequent kiln charge with the same species and drying conditions.

<table>
<thead>
<tr>
<th></th>
<th>$\bar{x}$ chart NORMAL</th>
<th>$S$ chart NORMAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCL</td>
<td>16.585</td>
<td>3.717</td>
</tr>
<tr>
<td>Center Line</td>
<td>14.264</td>
<td>1.7298</td>
</tr>
<tr>
<td>LCL</td>
<td>11.943</td>
<td>0.261</td>
</tr>
</tbody>
</table>

Figure 7. Control chart for sample averages based on the Normality assumption.

Figure 8. Control chart for sample standard deviations based on Normality assumption.
The $\bar{x}$ chart in Figure 7 shows no out-of-control situations. However, at least two points are observed around or above the "2-sigma" region. The "S chart NORMAL" in Figure 8 exhibits no out-of-control situation. The analysis of the above charts suggests that the drying process performs acceptable. This conclusion is based on the assumption that the moisture content is Normally distributed.

As it was found earlier, the distribution of experimental data is better modeled by a Lognormal distribution with threshold $\hat{\theta} = 7.661$, scale $\hat{\mu} = 1.907$, and shape $\hat{\sigma} = 0.236$. These results are summarized in Table 5. It is interesting to note that the estimated threshold value ($=7.661$) is very close to the lowest equilibrium moisture content controlled by the drying schedule ($=7.7$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal variable $X$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threshold</td>
<td>$\hat{\theta}$</td>
<td>7.661</td>
</tr>
<tr>
<td>Scale</td>
<td>$\hat{\mu}$</td>
<td>1.907</td>
</tr>
<tr>
<td>Shape</td>
<td>$\hat{\sigma}$</td>
<td>0.236</td>
</tr>
<tr>
<td>Number of Samples</td>
<td>$m$</td>
<td>20</td>
</tr>
<tr>
<td>Sample size</td>
<td>$n$</td>
<td>5</td>
</tr>
<tr>
<td>Normal variable $Y$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean of the Normal Population, defined by equation (45)</td>
<td>$\bar{y}$</td>
<td>1.907</td>
</tr>
<tr>
<td>Standard deviation of the Normal population, defined by equation (46)</td>
<td>$\bar{s}$</td>
<td>0.236</td>
</tr>
</tbody>
</table>

Table 5. Parameters of preliminary data for construction of control charts based on the Lognormal distribution
Using the procedures described in the ‘Control Charts for Lognormal Data’ section, the control limits were calculated with equations (49), (51), (56) and (57), and the results are presented in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>( \bar{y} ) chart</th>
<th>( \text{antilog } \bar{y} ) chart</th>
<th>( S ) chart</th>
<th>( \text{antilog } S ) chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{UCL} )</td>
<td>2.233</td>
<td>16.813</td>
<td>0.562</td>
<td>1.754</td>
</tr>
<tr>
<td>( \text{Center Line} )</td>
<td>1.882</td>
<td>14.052</td>
<td>0.262</td>
<td>1.299</td>
</tr>
<tr>
<td>( \text{LCL} )</td>
<td>1.532</td>
<td>12.108</td>
<td>0.039</td>
<td>1.040</td>
</tr>
</tbody>
</table>

Table 6. Control limits and center lines for control charts based on Lognormal assumption.

Using these parameters, the Lognormal charts were constructed. Figure 9 shows the \( \bar{y} \) chart, and Figure 10 shows the \( \text{antilog } \bar{y} \) chart. The \( S \) chart is shown in Figure 11, and the \( \text{antilog } S \) chart is shown in Figure 12.

As expected, the control limits for the \( \bar{y} \) chart are symmetric, while for the other three charts they are asymmetric. The limits of the \( \bar{y} \) chart as given by (49), are symmetric about \( \bar{y} \) with \( \pm \) "3-sigma". For the \( S \) chart the limits are not symmetric because they include percentage points of the chi-square distribution, which is not a symmetrical distribution. For the other two charts, "antilog \( \bar{y} \)" and "antilog \( S \)", the limits are not symmetric because they also involve exponentiation.

The points plotted on the Lognormal charts came from the same samples chosen for the Normality assumption. From the second kiln charge, a first sample of 5 random moisture content measurements was selected. After the log transformation to normality, the average of the
transformed sample was plotted on the $\bar{y}$ chart in Figure 9. This point was also plotted on the antilog $\bar{y}$ chart (Figure 10), after transforming it to the original scale of measurements. The standard deviation of the transformed sample was plotted on the $s$ chart (Figure 11), and its corresponding value in the original scale was also plotted on the antilog $s$ chart in Figure 12. A total of 20 successive samples were selected and plotted in this manner in charts from Figure 9 through Figure 12.

Figure 9. Control chart for sample averages ($\bar{y}$ chart) based on Lognormal assumption.

Figure 10. Antilog $\bar{y}$ chart (in original scale) based on Lognormal assumption.
Figure 11. Control chart for sample standard deviations (S chart) based on Lognormal assumption.

Figure 12. Antilog S chart (in original scale) based on Lognormal assumption.

It can be seen that the pair of charts “\( \bar{y} \)” and “antilog \( \bar{y} \)” have practically identical appearances and the plots look the same, although the scale of measurements is different. The same is true for the other pair of charts, “S” and “antilog S”. This means that the analysis of patterns and the sensitizing rules, which are valid for the \( \bar{y} \) chart, can also be applied to the corresponding “antilog \( \bar{y} \) chart”. Therefore, only one chart from each pair is sufficient for practical applications. It is recommended that the pair of charts “antilog \( \bar{y} \)” and “antilog S” be used because they show the control limits and plotted data in the original scale of measurements – percentages of moisture content.
The Lognormal charts in Figure 9 and Figure 10 show no out-of-control conditions. This is in accordance with the conclusions drawn previously from the "$\bar{x}$ chart NORMAL". It can be seen that the "$\bar{x}$ chart NORMAL" is somewhat similar in appearance with the $\bar{y}$ and antilog $\bar{y}$ charts, but not identical. Also, the Lognormal charts in Figure 11 and Figure 12 show no points out-of-control. This in accordance with results from the chart in Figure 8, which was based on the Normality assumption.

It can be seen that the charts based on the Lognormal assumption are somewhat similar to the charts based on the Normality. However, when these charts are compared more thoroughly it is obvious that the Normal charts tend to overestimate the plotted parameters, and in the long run are more prone to falsely signal out-of-control conditions. The two types of charts are shown again in pairs to better visualize and compare them, in Figure 13, Figure 14, Figure 15, and Figure 16.

From the comparative analysis of charts in Figure 13 and Figure 14 it can be easily seen that the Normality assumption tends to overestimate the points plotted on the "$\bar{x}$ chart NORMAL". This is more evident for the points corresponding to samples 3 and 6, which are closer to the upper limit of the "$\bar{x}$ chart NORMAL". It is expected that in the long run, this overestimation will create many false alarms (points show out-of-control conditions when in fact the process is in-control). Comparing the control limits of the two charts, it can be seen that the centerline of the "antilog $\bar{y}$" chart is smaller in value (14.052 vs. 14.264). As it was demonstrated earlier, this is because the parameter monitored on the "antilog $\bar{y}$" is the geometric mean of MC less the threshold, and the "$\bar{x}$ chart NORMAL" monitors the arithmetic mean of MC.
It is known that the geometric mean is always less than or equal to the arithmetic mean for a set of data. The asymmetrical control limits of the “antilog $\bar{y}$” chart will allow for larger variations of MC, which are due to the positive skewness, and not to outlier data. As it was discussed earlier, the fact that there may be MC observations in the upper tail of the distribution is a typical occurrence, and a properly chosen control chart should consider it this way. On the other hand, the “$\bar{x}$ chart NORMAL” will consider these points as “outliers” and will falsely signal them as out-of-control conditions.
Figure 15. Control chart for sample standard deviations based on the Normality assumption.

Figure 16. Antilog S chart (in original scale) based on Lognormal assumption.

The analysis of charts in Figure 15 and Figure 16 also reveals that the Normality assumption in the “S chart NORMAL” tends to overestimate the dispersion of MC. This is more evident in samples 3 and 6, but it can also be observed in sample 9. The use of the “S chart NORMAL” to monitor the experimental data would lead to incorrect conclusions about the performance of the process, especially about its dispersion.
4.5 LIMITATIONS OF THE STUDY

The specific results of the study can be applied only to Douglas-Fir lumber from the interior of British Columbia, Canada, for comparable kiln types, drying conditions, and lumber sizes. However, the methodology of determining the moisture content distribution and the methods for constructing control charts can be applied to many practical situations. For each kiln and set of conditions (species, lumber type and size, drying conditions), the distribution of moisture content could be assessed with the methods presented here. If the distribution of moisture content is proved to be well modeled by the Lognormal, control limits could then be established using the calculated distribution parameters, and the proposed charts should be used to monitor the process average and dispersion.
5 CONCLUSIONS

5.1 OVERVIEW

In this section the main findings of the research are summarized by drawing conclusions and recommending areas of interest for further investigations.

5.2 CONCLUSIONS

This study presented the methodology of using graphical and numerical methods for assessing the goodness of fit of an experimental distribution against the Normal, Lognormal, and Weibull distributions. Graphical methods rely on various approximations and therefore they were used for informal and preliminary judgments only. It was found that the threshold (location) parameter must be estimated and subtracted from the data for correct Lognormal and Weibull probability plotting. The threshold parameter of the experimental data ($\theta = 7.661\%$) was very close to the value of EMC (equilibrium moisture content) given by the drying schedule (EMC = 7.700\%). When applied to Douglas-Fir kiln-dried lumber data, the symmetry plots suggested that the empirical distribution is not symmetric and is positively skewed, and the probability plots indicated that Lognormal might be the true distribution. These findings were confirmed by formal numerical goodness-of-fit tests, and the overall conclusion of the goodness of fit tests was that the moisture content of the experimental data is well modeled by the three-parameter Lognormal distribution.
Control charts which monitor the parameters of a Lognormally distributed variable were proposed. When these Lognormal charts were compared to the customary charts based on the Normality assumption, it was found that the latter may wrongfully imply that the process is out of control, even when the skewness of the distribution is a natural result of the drying process. Control charts that are based on the Lognormal assumption are more appropriate to use in cases when the moisture content has such a skewed distribution.

A practical implementation of the proposed charts is recommended to begin with the determination of the underlying distribution of moisture content. Graphical methods as well as formal goodness-of-fit tests will determine whether the moisture content is better modeled by a Normal, Lognormal, or Weibull distribution. The analysis should start with the investigation of symmetry. If the empirical distribution is asymmetric then the Lognormal or Weibull may provide a better fit, and this should be first checked with probability plots. Carefulness is required in constructing these plots because they can be easily misused by failing to recognize that the moisture content distribution is left-bounded not by zero, but by another positive value – threshold or location. To obtain a correct probability plot, the threshold must be subtracted beforehand from the data. To generate rigorous conclusions about the underlying moisture content distribution, it is recommended that graphical methods be always accompanied by formal numerical tests, which should corroborate the findings. If it is concluded that a three-parameter Lognormal is a proper fit, then the central tendency and the variability of moisture content can be monitored with the charts proposed in this paper. The central tendency would be better modeled by the geometric mean, instead of the usual arithmetic mean, and this should be monitored with the \( \text{anti} \log \bar{y} \) chart proposed in this paper. The variability between different samples of moisture content can be monitored with the proposed \( \text{anti} \log S \) chart.
5.3 RECOMMENDATIONS FOR FURTHER RESEARCH

Considering the topics discussed and the findings of this research, several areas are recommended for further investigation.

One possible area of research is to determine how the distribution of moisture content changes during different stages of drying. Using the goodness of fit methods presented here, the initial distribution of MC (of the "green" wood) could be determined. Furthermore, the parameters of the distribution could be assessed during successive stages of the drying process, including the final MC. This research could lead to development of proper parameter-testing methods for each stage.

Another area of investigation is to ascertain precisely the relationship between the equilibrium moisture content controlled by the drying schedule, and the threshold (location) parameter of MC distribution, for each drying stage.

An additional research topic could be to further develop the Lognormal control charts proposed here, by determining their associated operating-characteristic and average-length curves, and process capability analysis.

Further research could determine the precise effect of non-normality on the design of control charts, specifically for a variable that has a three-parameter Lognormal distribution.
6 LITERATURE CITED


7 APPENDIX A - KILN CHARACTERISTICS

KILN

Double track conventional
Steam heated / steam injected
6 cross shaft fans
145,000 MBF capacity

KILN CAPACITY

160 units (loads) of 6’
Total of 36,480 pieces of 6’ in a single charge

UNITS

2”x4” – 12 pieces wide x 19 rows high (4’ high x 4’ wide)
228 pieces / unit
¾” stickers every 2’

KILN LOADING

60’ long tracks, 4 units high, 2 units wide
## 8 APPENDIX B – DRYING SCHEDULE

<table>
<thead>
<tr>
<th>Step No.</th>
<th>Time [h]</th>
<th>Temperature [°C]</th>
<th>EMC [%]</th>
<th>RH [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dry-bulb</td>
<td>Wet-bulb</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 to 8</td>
<td>26.7</td>
<td>26.7</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>8 to 16</td>
<td>65.6</td>
<td>62.8</td>
<td>15.4</td>
</tr>
<tr>
<td>3</td>
<td>16 to 24</td>
<td>71.1</td>
<td>67.2</td>
<td>13.4</td>
</tr>
<tr>
<td>4</td>
<td>24 to 32</td>
<td>76.7</td>
<td>71.7</td>
<td>11.8</td>
</tr>
<tr>
<td>5</td>
<td>32 to 40</td>
<td>82.2</td>
<td>75.6</td>
<td>10.1</td>
</tr>
<tr>
<td>6</td>
<td>40 to 48</td>
<td>87.8</td>
<td>79.4</td>
<td>8.9</td>
</tr>
<tr>
<td>7</td>
<td>48 to 56</td>
<td>93.3</td>
<td>83.3</td>
<td>7.7</td>
</tr>
<tr>
<td>8</td>
<td>56 to 80</td>
<td>93.3</td>
<td>90.6</td>
<td>14</td>
</tr>
</tbody>
</table>
### 9 APPENDIX C – DATA PLOTTED ON CONTROL CHARTS

<table>
<thead>
<tr>
<th>Sample #</th>
<th>$\bar{x}$ chart Normal [%]</th>
<th>$S$ chart Normal [%]</th>
<th>$\bar{y}$ chart [unitless]</th>
<th>antilog $\bar{y}$ chart [%]</th>
<th>$S$ chart [unitless]</th>
<th>antilog $S$ chart [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.767</td>
<td>0.934</td>
<td>1.829</td>
<td>13.711</td>
<td>0.151</td>
<td>1.1625</td>
</tr>
<tr>
<td>2</td>
<td>14.570</td>
<td>0.754</td>
<td>1.953</td>
<td>14.536</td>
<td>0.110</td>
<td>1.1161</td>
</tr>
<tr>
<td>3</td>
<td>15.866</td>
<td>2.565</td>
<td>2.089</td>
<td>15.563</td>
<td>0.301</td>
<td>1.3517</td>
</tr>
<tr>
<td>4</td>
<td>15.153</td>
<td>1.253</td>
<td>2.027</td>
<td>15.075</td>
<td>0.158</td>
<td>1.1713</td>
</tr>
<tr>
<td>5</td>
<td>14.218</td>
<td>1.806</td>
<td>1.879</td>
<td>14.033</td>
<td>0.261</td>
<td>1.2986</td>
</tr>
<tr>
<td>6</td>
<td>16.064</td>
<td>2.734</td>
<td>2.101</td>
<td>15.658</td>
<td>0.363</td>
<td>1.4382</td>
</tr>
<tr>
<td>7</td>
<td>13.778</td>
<td>0.900</td>
<td>1.831</td>
<td>13.725</td>
<td>0.147</td>
<td>1.1586</td>
</tr>
<tr>
<td>8</td>
<td>14.045</td>
<td>1.526</td>
<td>1.859</td>
<td>13.901</td>
<td>0.238</td>
<td>1.2684</td>
</tr>
<tr>
<td>9</td>
<td>14.957</td>
<td>1.865</td>
<td>1.990</td>
<td>14.796</td>
<td>0.225</td>
<td>1.2527</td>
</tr>
<tr>
<td>10</td>
<td>12.825</td>
<td>1.161</td>
<td>1.655</td>
<td>12.715</td>
<td>0.235</td>
<td>1.2643</td>
</tr>
<tr>
<td>11</td>
<td>14.577</td>
<td>2.313</td>
<td>1.916</td>
<td>14.280</td>
<td>0.328</td>
<td>1.3875</td>
</tr>
<tr>
<td>12</td>
<td>14.327</td>
<td>1.557</td>
<td>1.903</td>
<td>14.191</td>
<td>0.222</td>
<td>1.2491</td>
</tr>
<tr>
<td>13</td>
<td>13.968</td>
<td>0.810</td>
<td>1.863</td>
<td>13.926</td>
<td>0.129</td>
<td>1.1381</td>
</tr>
<tr>
<td>14</td>
<td>13.626</td>
<td>0.988</td>
<td>1.805</td>
<td>13.564</td>
<td>0.159</td>
<td>1.1727</td>
</tr>
<tr>
<td>15</td>
<td>14.228</td>
<td>2.068</td>
<td>1.869</td>
<td>13.965</td>
<td>0.322</td>
<td>1.3795</td>
</tr>
<tr>
<td>16</td>
<td>15.461</td>
<td>0.743</td>
<td>2.073</td>
<td>15.433</td>
<td>0.093</td>
<td>1.0977</td>
</tr>
<tr>
<td>17</td>
<td>14.016</td>
<td>0.767</td>
<td>1.871</td>
<td>13.981</td>
<td>0.115</td>
<td>1.1214</td>
</tr>
<tr>
<td>18</td>
<td>13.609</td>
<td>0.905</td>
<td>1.802</td>
<td>13.548</td>
<td>0.164</td>
<td>1.1782</td>
</tr>
<tr>
<td>19</td>
<td>13.989</td>
<td>1.389</td>
<td>1.857</td>
<td>13.884</td>
<td>0.196</td>
<td>1.2166</td>
</tr>
<tr>
<td>20</td>
<td>13.412</td>
<td>1.435</td>
<td>1.752</td>
<td>13.252</td>
<td>0.273</td>
<td>1.3133</td>
</tr>
</tbody>
</table>
APPENDIX D – TWO REPLICATIONS FOR GOODNESS OF FIT TESTS OF HYPOTHESIS

### Replicate #1

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Normal Test</th>
<th>Lognormal Test</th>
<th>Weibull Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>$\bar{D}$</td>
<td>1.5855</td>
<td>1.035</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>$\bar{w}^2$</td>
<td>0.5717</td>
<td>0.178</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>$\bar{A}^2$</td>
<td>3.5904</td>
<td>1.092</td>
</tr>
</tbody>
</table>

### Replicate #2

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Normal Test</th>
<th>Lognormal Test</th>
<th>Weibull Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>$\bar{D}$</td>
<td>1.6127</td>
<td>1.035</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>$\bar{w}^2$</td>
<td>0.5988</td>
<td>0.178</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>$\bar{A}^2$</td>
<td>3.6778</td>
<td>1.092</td>
</tr>
</tbody>
</table>