12-2007

Parametric and Non-Parametric Regression Tree Models of the Strength Properties of Engineered Wood Panels Using Real-time Industrial Data

Timothy Mark Young
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Timothy Mark Young entitled "Parametric and Non-Parametric Regression Tree Models of the Strength Properties of Engineered Wood Panels Using Real-time Industrial Data." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Natural Resources.

Timothy G. Rials, Frank M. Guess, Major Professor

We have read this dissertation and recommend its acceptance:

Donald G. Hodges, J. Larry Wilson

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Timothy Mark Young entitled "Parametric and non-parametric regression tree models of the strength properties of engineered wood panels using real-time industrial data." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degrees of Doctor of Philosophy, with a major in Natural Resources.

_____
Timothy G. Rials, Co-Major Professor

_____
Frank M. Guess, Co-Major Professor

We have read this dissertation
and recommend its acceptance

_____
Donald G. Hodges

_____
J. Larry Wilson

Acceptance for the Council:

Carolyn R. Hodges, Vice Provost and
Dean of the Graduate School

_____

(Original signatures are on file with official student records)

PARAMETRIC AND NON-PARAMETRIC REGRESSION TREE MODELS OF THE STRENGTH PROPERTIES OF ENGINEERED WOOD PANELS USING REAL-TIME INDUSTRIAL DATA

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Timothy Mark Young
December 2007

Dedication

This dissertation would not have been possible without the love and support of my family. They have given me encouragement in their words, actions and prayers. To my wife Anne, thanks for putting up with my many moods and long hours away from home. You are the love of my life and your sacrifice for my collegiate endeavors will always be remembered. To my daughter Elizabeth, thanks for the instant messaging words of support late at night while I was working on this dissertation. Your kind words via instant messaging, love, prayers and encouragement helped me endure some of those very late nights. To my son Steven, your love, prayers and sense of humor made me laugh at times when I thought I could no longer smile during this process.

To my mother, thank you so much for all that you have given me. Your love, kind words and prayers are very special to me. Thank you for that strong work ethic and sense of discipline you instilled in me at a very young age (I still remember and I'm thankful for driving me on my early morning paper route before school on days when the winter weather in Wisconsin was very cold). To my sister, thanks for your love and support and always reminding me of my father's words and his encouragement. I dedicate this dissertation in memory of my father, James R. Young. He was the best father a son could ever wish for and I believe he would be proud of this accomplishment. I will always remember one of his last statements of encouragement to me, "Son, give to the world your best and you will receive the best in return."

I am very thankful for my loving family and all that the dear Lord has given me. This dissertation would not have been possible without my spiritual beliefs. Just think what the world would be like if everyone tried to act like Jesus for one day.

Professional Acknowledgements

I have so many people to thank for their mentoring during this dissertation process. First, I would like to thank my committee members. Dr. Timothy G. Rials (co-chair), thank you for your continuous support, strong mentoring and sense of humor during this process. Dr. Frank M. Guess (co-chair), thank you for always providing that positive encouragement and trying to make me laugh when laughs did not seem possible. Dr. J. Larry Wilson, thank you for your mentoring and professional guidance, your career advice is always greatly appreciated. Dr. Donald G. Hodges, thank you for your guidance and I will always remember your first two economic questions during the oral defense of my written comprehensive.

Other special colleagues and friends that encouraged me during this process that I would like to thank are: Dr. George Hopper (former Department Head); Dr. Dave Ostermeier, Dr. Paul Winistorfer; Dr. Keith Belli (current Department Head) and Mr. Jim Perdue (U.S. Forest Service). I also would like to thank the Administration of the Institute of Agriculture, Agricultural Experiment Station for their support. I would also like to thank Dr. Loh at the University of Wisconsin for responding to my e-mails and questions concerning regression trees and GUIDE software.

A good teacher learns well from his or her students. I would like to thank three of my former M.S. GRAs, Ms. Leslie Shaffer, Ms. Diane Perhac and Dr. Yang Wang whom each in their own way provided encouragement to me during this process. I will enjoy following their progress in their professional careers. I would like to thank Ms. Amanda Silk, Administrative Assistant at the Forest Products Center, for her kind words on some of the long enduring days. For those I have forgotten to thank, please excuse my lapse.

Abstract

The forest products industry is undergoing unprecedented change from international competition, increasing fiber costs, rising energy prices and falling product prices. Competitive businesses have the key ability to adapt quickly to change through improved knowledge. Among adaptations to change are better product development, improved process efficiency and superior product quality. This dissertation is directly related to improving the knowledge of forest products manufacturers by investigating data mining (DM) methods that improve the ability to quantify causality of sources of variation. A contemporary DM method related to decision theory is decision trees (DTs). DTs are designed for heterogeneous data and are highly resistant to irrelevant regressors. The tree structures of DTs are also easy to interpret.

The research hypothesis of this dissertation is that there is no significant difference in the explanatory or predictive capabilities of multiple linear regression (MLR) models, parametric regression trees (RTs) and non-parametric quantile RTs. To test this hypothesis 1,335 statistical models are developed. Box Cox transforms of Y are considered. Models are developed for the internal bond (IB) of medium density fiberboard (MDF) and the IB (and Parallel EI) of oriented strand board (OSB) from automatically fused data of destructive test data and real-time production line sensor data.

Models with good predictability of the validation data set are possible for MDF IB when using traditional MLR methods with short record lengths without Box Cox transforms. Significant regressors ($\alpha < 0.01$) for MDF MLR models are related to overall pressing time and press pre-position time settings.

Parametric and non-parametric RT models without Box Cox transforms outperform the predictability of MLR models. For MDF IB, process variables related to overall pressing time, press position times and core fiber moisture are significant ($\alpha < 0.01$). RT models with Box Cox transforms of OSB IB improve predictability for record lengths less than 100. Significant regressors ($\alpha < 0.01$) of OSB IB are related to pressing times and core layer moisture. Significant regressors ($\alpha < 0.01$) of OSB Parallel EI are related to forming speed and pressing times. There is evidence from the extensive investigation of 1,335 models to support the alternative research hypothesis.

Table of Contents

Table of Contents

(continued)

Table of Contents

(continued)

List of Tables

Page

Appendix to Chapter III

Appendix to Chapter IV

List of Tables

(continued)

Page

Appendix to Chapter V

List of Tables

(continued)

List of Figures

Page

Appendix to Chapter II

Appendix to Chapter III

List of Figures
(continued)

Page

Appendix to Chapter IV

List of Figures
(continued)

List of Figures
(continued)

Page

List of Figures
(continued)

Page

Appendix to Chapter V

List of Figures
(continued)

Page

List of Figures
(continued)

Page

List of Figures
(continued)

Page

List of Figures
(continued)

Page

Appendix to Chapter VI

List of Figures
(continued)

Page

CHAPTER I. INTRODUCTION

An underlying basis of statistical methods is the study of variance (σ^2). Variance in the context of manufacturing is defined as estimated process variance ($\hat{\sigma}^2$). In forest products manufacturing, process variance results in inferior product quality, poor product safety and noncompetitive costs. Process variance is masked by higher than necessary operational targets (e.g., weight, thickness, density, resin, etc.) which require higher than necessary energy use which in combination are not competitive or sustainable in a highly competitive market place.

A common goal for statistical research is to investigate and quantify causality between independent variables (*X*) and response variables (*Y*) with a high level of scientific inference. As Friedman (2001) notes, given a set of measured values of attributes, characteristics or properties on a object (observation) $\mathbf{X} = (X_1, X_2, \ldots X_n)$, which are often called "variables," the goal is to predict (estimate) the unknown value of another attribute *Y*. In quantifying causality, de Mast and Trip (2007) note the important distinction between exploratory and confirmatory data analysis which they attribute to Tukey's (1977) work. As Tukey (1977) pointed out, confirmatory data analysis is concerned with testing a pre-specified hypothesis. The purpose of exploratory data analysis is hypothesis generation (de Mast and Trip 2007). This dissertation is undertaken in the spirit of exploratory data analysis and hypothesis generation. The dissertation is aligned with Gleser's (1996) "First Law of Applied Statistics," i.e., two individuals using the same statistical method on the same data should arrive at the same conclusion.

The goal of the dissertation is to improve the understanding of causality for the strength properties of MDF and OSB from industrial derived data. The dissertation is focused on exploratory analysis in the context of data mining, i.e., quantifying unknown causality from large volumes of electronically collected data which are fused with destructive data of strength properties.

Data mining (DM), also called Knowledge-Discovery in Databases (KDD) or Knowledge-Discovery and Data Mining, is the process of automatically searching large volumes of data for patterns (http://en.wikipedia.org/wiki/Data_mining. referenced 10/4/07). DM is the contemporary edge of the sciences of Artificial Intelligence, Machine Learning, Pattern Recognition and Data Visualization. DM evolved from advancements in database management systems (DBMS) and on-line (real-time) transaction processing (OLTP). From a statistical perspective it can be viewed as computer automated exploratory data analysis of large complex data sets (Friedman and Wall 2005).

Exponential growth of data mining applications has occurred globally for many industries (Harding et al. 2006). Rapid growth in data mining applications in the forest products industry is imminent and is desperately needed for the industry's economic survival.

The New Millennium for the Forest Products Industry

Forests sustain an important forest products economy in the U.S. and state of Tennessee. The forest products industry contributed more than $240 billion to the U.S. economy and employed more than 1,000,000 Americans in 2002 (U.S. Census Bureau 2004). Over 180,000 Tennesseans were employed by the forest products industry in 2000 accounting for 6.6 percent of Tennessee's economy by generating $21.7 billion in value in

that same year (Young et al. 2007). Threats to this economic sector have arisen in the form of unprecedented levels of international competition, constrained credit markets, increasing fiber costs, increasing energy costs and substitution from renewable wood products to non-renewable oil- and cement-derived products.[1]

Wood costs represent the largest single component of total manufacturing costs for most forest products manufacturers. Some U.S. manufacturers must contend with wood costs as high as 60% of total manufacturing costs.[2] Demand/capacity ratios for the engineered wood panel sector are falling below the critical threshold of 85 percent, a level that results in declining real prices. A renewed emphasis on reducing costs is desperately needed by this important economic sector.

Rationale and Justification

Poor production efficiencies in the engineered wood panel sector occur from unacceptably high levels of wood waste due to low strength and high wood-density targets. Poor production efficiencies lead to high wood use, high energy usage, and an overall lack of business competitiveness. Wood waste is a significant contributor to costs. In 2003, the engineered wood panel sector produced 64.3 billion square feet of panels and wood waste ranged from three percent to nine percent (Composite Panel Association 2004, TECO 2004). Reducing wood waste by one percent could translate into annual savings of as much as $700,000 per producer and promote wiser use of the forest resource.[3]

---

[1]U.S. structural wood panel mills lost 5.8 percent of the North American market in 2004, primarily from Europe and South America. Imports over the long term are forecast to increase (Engineered Wood Association 2004).
[2]Personal communications 2005 and 2006: Georgia-Pacific, J.M. Huber Corporation, Louisiana-Pacific Corporation, Norbord Corporation and Weyerhaeuser Corporations.
[3]Personal communications 2005 and 2006: Georgia-Pacific, J.M. Huber Corporation, Louisiana-Pacific Corporation, Norbord Corporation and Weyerhaeuser Corporations.

Improved production efficiencies, reduced wood waste, and lower costs are possible from data mining by improving the knowledge of causality of sources of unknown process variation.

Many organizations can be labeled as "data-rich" and "knowledge-poor" (Chen 2005). Modeling of wood composite manufacturing processes enables complex processes to be better understood by examining the patterns in data related to the previous behavior of a manufacturing process (Young and Guess 2002). The benefits of first-order models in engineered wood panel production are well documented (Young 1997, Gruebel 1999, Bernardy and Scherff 1998, 1999, Erilsson et al. 2000, Young and Guess 2002, Guess et al. 2003, and Kim et al. 2007). Erilsson et al. (2000) discussed the potential of stochastic models for engineered wood manufacture, while Gruebel (1999) documented medium density fiberboard (MDF) manufacturing cost savings of five percent to ten percent from the use of "off-line" first-order statistical models. Dawson et al. (2006) developed a genetic algorithm/neural network (GANN) real-time predictive model of MDF and oriented strand board (OSB) manufacturing processes that resulted in cost annual savings at two test sites ranging from $700,000 to $1.2 million.

This dissertation investigates the use of Regression Tree (RT) models to identify unknown causality between process variables and strength properties of wood composites. RT models are known for their high explanatory value and RT models are at least as predictive as black box deterministic methods (Loh 2002).

Problem Statement and Scope

The problem statement of this dissertation is to explore the explanatory and predictive capabilities of parametric and non-parametric (quantile) regression tree models

of the strength properties of MDF and OSB using real-time sensor data amassed in manufacturers' data warehouses. Specifically, this research will investigate the explanatory and predictive capabilities of three modeling methods: first-, second- and third-order statistical multiple linear regression models with interaction terms, parametric regression trees, and non-parametric (quantile) regression trees. The models are developed for one MDF and one OSB mill both located in the southeastern United States.

Dissertation Hypothesis

The null research hypothesis of this dissertation is that there is no significant difference in the explanatory or predictive capabilities of three modeling methods: first-, second- and third-order statistical multiple linear regression models with interaction terms; parametric regression trees; and non-parametric (quantile) regression trees. The test of this research hypothesis hopefully will incrementally advance the statistical and industrial engineering sciences as applied to wood composites manufacture.

Dissertation Objectives

1. Investigate first-, second- and third-order MLR models

First-, second- and third-order MLR models with interaction terms are investigated for MDF and OSB wood composite strength properties. The database to support this work is the real-time relation database developed by Young and Guess (2002), enhanced by Dawson et al. (2006). MLR models are developed for three predominately manufactured products (0.500", 0.625" and 0.750" industrial grades) for MDF and one OSB product (7/16" roof sheathing) for each test site.

2. Investigate regression tree (RT) models

Parametric and non-parametric (quantile) regression trees (RT) are investigated for each MDF and OSB manufacturing test site. The same database of the first objective is used. RT models are developed for the same nominally produced MDF and OSB products defined in the first objective.

3. Compare the explanatory and predictive capabilities of the MLR and RT models developed in the first and second objectives.

The explanatory and predictive capabilities of the MLR and RT models developed in the first and second objectives are compared. Model prediction capabilities are analyzed for appropriate validation data sets from each MDF and OSB mill.

There is no documentation in the literature of the investigation and application of decision tree theory to manufactured wood composites strength properties. It is hoped that this research will at least incrementally expand the sciences of wood composites manufacture and decision tree theory.

The dissertation is organized in seven chapters. In Chapter II the relevant literature for the dissertation is reviewed. The methods used in the research are presented in Chapter III, with a description of the data and an assessment of the data quality (descriptive statistics and distribution fits) for the important response variables. The results of the first objective are given in Chapter IV. Chapter V is the core chapter of the dissertation where results are presented on regression trees. Chapter VI compares the results of the first two objectives. Chapter VII summarizes the dissertation with conclusions and a discussion of future research. The Bibliography and General Appendices follow Chapter VII.

CHAPTER II. LITERATURE REVIEW

Information technology is the largest single capital investment for many enterprises (Thorpe 1998). However, many companies struggle with making use of the vast amount of data that is acquired at increasingly faster rates. Thorpe (1998) called this phenomenon the "Information Paradox" where companies invest increasing amounts of money on information acquisition but cannot demonstrate a connection between the money spent and business results. This paradox is caused from the lack of useful "real-time relational databases" that are of sufficient design and organization where parametric and non-parametric statistical methods can be used to investigate unknown causality and develop scientific knowledge. The data warehouse in many ways is the nucleus for process knowledge of a manufacturing enterprise. As Harding et al. (2006) noted, "Knowledge is the most valuable asset of a manufacturing enterprise, as it enables a business to differentiate itself from competitors and to compete efficiently and effectively to the best of its ability." The Harding et al. (2006) statement is very appropriate for the wood composites industry in the present era of unprecedented competition, increasing raw material costs, increasing energy costs and declining product prices.

Data Warehouse

As Inmom and Hackathorn (1994) noted, a data warehouse is the main repository of the organization's historical data, its corporate memory.[4] The central concept of a data

[4] The origin of the data warehouse can be traced to studies at MIT in the 1970s which were targeted at developing an optimal technical architecture. At the time, the craft of data processing was evolving into the profession of information management. The MIT work led to the modern

warehouse is that it is a collection of records. Data warehouses usually consist of one or more databases of volumes of records. The structure of a database is known as a schema. The schema describes the objects that are represented in the database and the table relationships among them. Multiple related tables each consisting of rows and columns is the most common form of schema (White 2002). Schema design is a critical factor in ensuring optimal storage and data compression, and also ensures the overall usefulness of the data during retrieval for analysis.

Real-time Data Warehouse

Real-time data warehousing originated and evolved with the computer industry. Real-time data captures manufacturing activity as it occurs. Real-time data usually are stored in a data warehouse either at the occurrence of an event or as a function of time. Most real-time data warehouse platforms can efficiently store multiple gigabytes of process data. Real-time data warehousing has become affordable in the last decade and it is hard to find a modern forest products manufacturer that does not have some type of real-time data warehousing platform. However, most forest products manufacturers use real-time data for simple trending analysis and rudimentary process knowledge. They struggle with using real-time databases for advanced analytics and scientific knowledge of the process. This dissertation directly addresses improved scientific knowledge of processes from the use of parametric and non-parametric regression tree methods using real-time process data.

Real-time databases have inherent data storage characteristics that need to be understood before advanced analytics can occur. Data quality is a key obstacle in the use

---

concept of the Information Center (http://www.damanconsulting.com/company/articles/dwrealtime.htm referenced 10/5/07).

of real-time data storage. Real-time data quality problems such as null fields, repeated records, correct time stamps, bi-modality/multi-modality and data leverage are significant issues which confound advanced analytics and data mining efforts. Some research has addressed data quality issues during real-time data retrieval using first-order statistical models and deterministic algorithms (e.g., genetic algorithms and neural networks) to model the wood composite process (Gruebel 1999; Bernardy and Scherff 1998, 1999; Young et al. 2004 and Dawson et al. 2006). The key first step in the use of real-time data is the development of the real-time relational database.

Real-time Relational Database

A real-time relational database is defined as the alignment of real-time process sensor data from the production line with product quality data, e.g., destructive test data of strength quality developed from the mill testing laboratory. The real-time relational database used in this dissertation is considered to be distributed data fusion (also called track-to-track fusion) where data from multiple diverse sensors are combined in order to make inferences about a physical event, activity or situation, e.g., internal bond tensile strength, modulus of elasticity flexure strength, etc. (Hall 1992).[5] Intellectual latency is the most significant issue in real-time relational databases. Some intellectual latency results from improper time alignment of process sensor data with product quality data. Young and Guess (2002) developed an automated relational database that addressed some of the issues of intellectual latency. Clapp et al. (2007) use the Eigenvalues from principal

[5] Data fusion or information fusion are names that have been given to a variety of interrelated expert system problems which have arisen primarily in military applications (Goodman et al. 1997). Other applications of data fusion include remote sensing, medical diagnostics and robotics (Blackman and Broida 1990, Hovanessian 1980).

component analysis to identify improper time alignment of real-time process sensors with destructive test data for MDF.

A data warehouse or real-time data warehouse contains just data. The key to scientific inference and improvement is the conversion of information or data into knowledge. This dissertation attempts to advance the scientific understanding of MDF and OSB manufacture. Many forest products manufacturers are unsuccessful in the information-to-knowledge transformation because they lack the key foundation of an automated real-time relational database.

Data Mining

Data mining (DM) is used to discover patterns and relationships in data, with an emphasis on large observational databases (Friedman and Wall 2005). DM is a large discipline and a plethora of literature exists on the subject. The literature review of DM in this chapter is not intended to be comprehensive, but instead a helpful precursor for the analytical methods used in this dissertation.

DM is the contemporary edge of the sciences of Artificial Intelligence, Machine Learning, Pattern Recognition and Data Visualization. DM evolved from advances in database management systems (DBMS) and on-line (real-time) transaction processing (OLTP). From a statistical perspective it can be viewed as computer automated exploratory data analysis of large complex data sets (Friedman 2001). DM is directly related to the field of Decision Theory. As Friedman (2001) notes, "It also affords enormous research opportunities for new methodological developments… ….Statistics can potentially have a major influence on Data Mining."

DM is closely related to machine learning and prediction. The predictive or machine learning problem is easy to state if difficult to solve in general (Friedman 2001). Given a set of measured values of attributes, characteristics or properties, on a object (observation) $\mathbf{X} = (X_1, X_2, \ldots X_n)$, which are often called "variables," the goal is to predict (estimate) the unknown value of another attribute *Y* (Friedman 2001). The quantity *Y* is called the "output," "dependent" or "response" variable, and $\mathbf{X} = (X_1, X_2, \ldots X_n)$ are referred to as the "input," "independent," "predictor" or "regressor" variables (Friedman 2001). The prediction takes the form of a function $\hat{Y} = F(x_1, x_2, ...x_n) = F(x)$ that maps a point *X* in the space of all joint values of the predictor variables, to a point $\hat{Y}$ in the space of response values (Friedman 2001). Most scientists agree that the goal is to produce a "good" predictive *F(x)*.

Decision trees are one of the most popular predictive learning methods used in data mining. Decision trees were developed largely in response to the limitations of kernel methods (Friedman 2001).[6] No matter how high the dimensionality of the predictor variable space, or how many variables are actually used for prediction (splits), the entire model can be represented by a two-dimensional graphic, which can be plotted and easily interpreted (Friedman 2001). Decision trees have an advantage of being very resistant to irrelevant predictor or regressor variables, i.e., since the recursive tree building algorithm estimates the optimal variable on which to split at each step, regressors unrelated to the response tend not to be chosen for splitting (Breiman et al. 1984). Friedman (2001) also

[6] Kernel Methods (KMs) are a class of algorithms for pattern analysis, whose best known element is the Support Vector Machine (SVM). Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. The general task of pattern analysis is to find and study general types of relations (for example clusters, rankings, principal components, correlations, classifications) in general types of data (http://en.wikipedia.org/wiki/Kernel_methods referenced 10/5/07).

notes a strength of decision trees is that regressors do not have to be tuned (standardized) which makes the method an "off-the-shelf" procedure. Fore example, NIR spectral data have been used with decision trees to enhance automated classifications of fruit and organic matter in soil (Ware et al. 2001, Shepherd and Walsh 2002, Shepherd et al. 2003).

This ease of interpretation from two-dimensional plots makes decision trees a powerful tool for the practitioner and an appropriate methodology for this dissertation for ease of use by practitioner. Fitting quotes supportive of this dissertation are by C. Dickens and J.H. Friedman (Friedman 1994), "Every time computing power increases by a factor of ten we should totally rethink how we compute." Friedman's (2001) corollary, "Every time the amount of data increases by a factor of ten, we should totally rethink how and what we compute." A more detailed literature review of decision trees is presented later in this chapter.

Multiple Linear Regression

The method of least squares and the precursor to regression analysis can be dated to 1805 by the publication of Legendre's work on the subject (Stigler 1986). Sir Francis Galton discovered regression around 1885 in studies of heredity (Stigler 1986). Galton's regression (as finally developed by Yule) was not simply an adaptation of least squares to a different set of problems; it was a new way of thinking about multivariate data (Stigler 1986).

Today regression analysis remains one of the most popular and globally used tools. Practitioners like regression analysis because of ease of interpretation in the coefficients that do not require standardization of the data of either the dependent and independent variables. Practitioners also like the visual interpretation of regression. Simple linear (SL)

and multiple linear regression (MLR) methods are widely available on business and statistical software, and MLR is a prerequisite for most undergraduate business and science degrees.

For situations where the data are drawn from reasonably homogeneous populations, traditional methods such as MLR can yield insightful analyses. The usefulness of MLR in data mining can breakdown quickly if the stringent assumptions associated with MLR are not met, e.g., normality assumption of the response (*Y*).

There is a plethora of literature on regression analysis and many tomes are available on the method. An extensive literature review of the heavily referenced MLR method is not presented in this dissertation given that it is not the primary method used for analysis.

Quantile Regression

As noted by Koenker (2005), Edgeworth's (1888) work on median methods is the genesis of the idea of quantile regression. Edgeworth emphasize that the assumed optimality of the sample mean as an estimator of location was crucially dependent on the assumption that the observations came from a common normal distribution. If the observations departed from normals with different variances, the median, Edgeworth argued, could easily be superior to the mean. Koenker (2005) notes that Edgeworth (1888) discards the Boscovich-Laplace constraint related to least squares that the residuals sum to zero and proposes to minimize the sum of the absolute residuals in both slope and intercept parameters. Unfortunately, the computational rigors associated with Edgeworth's (1888) work limited the application of the method until the development of linear programming which provides an efficient conceptual approach. Mosteller (1946)

discovered that quantile estimators are almost as efficient as the maximum likelihood estimators for most conventional parametric models.

Quantile regression as introduced by Koenker and Bassett (1978) seeks to extend these ideas to the estimation of conditional quantile function, i.e., models in which quantiles of the conditional distribution of the response variable are expressed as functions of observed covariates. The quantile regression literature in economics makes a persuasive case for the value of going beyond models for the conditional mean (Chamberlain 1994). Koenker and Billias (2001) explore quantile regression models for unemployment duration data and offer an introduction to quantile regression for demand analysis. There is also a growing literature database in empirical finance employing quantile regression methods. Bassett and Chen (2001) consider quantile regression index models to characterize mutual fund investment styles. Shaffer (2007) and Young et al. (2007c) explore the first uses of quantile regression in modeling the internal bond strength property of MDF. The method of quantile regression is described in more detail in the next chapter.

Decision Trees

The machine learning technique for inducing a decision tree from data is called decision tree learning, or (colloquially) "decision trees".[7] Decision tree (DT) models have grown into a powerful class of methods for examining complex relationships with many types of data (Kim et al. 2007). Researchers and practitioners find great explanatory value in DT models. DT models are more useful than MLR models when data are not homogeneous (Figure 2.1).

[7] http://en.wikipedia.org/wiki/Regression_Tree referenced 10/5/07.

A "regression tree" is a decision tree for numerical data. A "classification" tree is a decision tree for categorical data. Since this dissertation uses numerical data from industrial processes, the primary focus of the following literature review is for the regression tree (RT).

A regression tree is a piecewise constant or piecewise linear estimate of a regression function, constructed by recursively partitioning the data and sample space. Its name derives from the practice of displaying the partitions as a decision tree, from which the roles of the regressors are inferred (Figure 2.2).

Construction of a regression tree consists of the following general steps performed iteratively, ending with step four:

- Partition the data,
- Fit a model to the data in each partition,
- Stop when the residuals of the model are near zero or a small fraction of observations are left,
- Prune the tree if it over fits.

Most of the contemporary regression tree algorithms differ on steps one and two. Many popular graphical-user interface software packages that have DT algorithms do not have step four (e.g., JMP - http://www.jmp.com/, Statistica - http://www.statsoft.com/, etc., referenced 10/5/07).

The AID (*"Automatic Interaction Detection"*) algorithm by Morgan and Sunquist (1963), Kass (1975) and Fielding (1977) is the first implementations of the DT idea. AID searches over all axis-orthogonal partitions and yields a piecewise constant estimate (Loh 2002). At each stage, the partition that minimizes the total sums of squared errors (SSE) is selected. Splitting stops if the fractional decrease in total SSE is less than a pre-specified γ or if the sample size is too small. As noted by Loh (2002), a weakness of AID is that it is

hard to specify γ, i.e., too small or too large a γ leads to over- or under-fitting. Another weakness of AID (Doyle 1973) is that the "greedy search" approach induces a bias in variable selection, e.g., if X_1 and X_2 are ordered regressors with $n_1 > n_2$, X_1 will have a higher chance of being selected which leads to erroneous inferences from the final tree structure. "Greedy search" methods find the regressor that minimizes the total SSE in the regression models fitted to the data subsets defined by the split (e.g., JMP and Statistica). "Greedy search" methods are also computationally expensive (Loh 2002).

The CART© (*"Classification and Regression Trees",* http://www.salford-systems.com/ referenced 9/20/07) algorithm followed AID and is a popular DT method (Breiman et al. 1984). Unlike AID, it avoids choosing a γ by employing a backward-elimination strategy to determine the tree (Loh 2002). It grows an overly large tree and then prunes away some branches, using a test sample or v-fold cross-validation (CV) to estimate the total SSE. In step two of the four general steps of decision trees, the CART regression tree fits a mean function in each partition (also called a piecewise constant regression tree).

The MARS© (*"Multivariate Adaptive Regression Splines"*) by Friedman (1991, http://www.salford-systems.com/ referenced 9/20/07) method combines spline fitting with recursive partitioning to produce a continuous regression function estimate (Chaudhuri et al. 1995). Chaudhuri at al. (1995) note that the complexity of the estimate from MARS© makes interpretation difficult and theoretical analysis of the spline statistical properties extremely challenging.

Quinlan's (1992) M5 method constructs an ordinary regression tree with a stepwise linear regression model fitted to each node at every stage. As noted in Kim et al. (2007b), Chaudhuri et al. (1994) chose a residual-based approach from MLR models. This

approach selects the variable with the signs of the residuals which appear most non-random, as determined by the significance probabilities of two-sample t-tests.

The FIRM algorithm (*"Formal Inference-based Recursive Modeling"*) by Hawkins (1997) addresses the bias problem of AID by using Bonferroni-adjusted significance tests to select predictors for splitting. Unlike AID and CART, FIRM splits each node into as many as ten subnodes for an ordered regressor. As Loh (2002) noted from Hawkins (1997) work, the Bonferroni adjustment can over-correct, resulting in a bias toward regressors that have fewer splits.

As cited by Loh (2002), other methods have been proposed for determining the final tree: Ciampi et al. (1988, 1991) combine non-adjacent partitions; Chaudhuri et al. (1994) use a CV-based, look-ahead procedure; Marshall (1995) finds non-hierarchical partitions; Chipman et al. (1998) and Denison et al. (1998) employ Bayesian methods to search among trees; and Li et al. (2000) use a stopping rule based on statistical significance tests.

The GUIDE, ver. 5.2 (*"Generalized, Unbiased, Interaction Detection and Estimation"*) DT algorithm is used in this research. GUIDE (Loh 2002, Chaudhuri and Loh 2002) extend the idea of Chaudhuri et al. (1994) by means of "curvature tests" (www.stat.wisc.edu/~loh/ referenced 10/5/07). A curvature test is a chi-square test of association for a two-way contingency table where the rows are determined by the signs of the residuals (positive versus non-positive) from a fitted regression model. The idea is that if a model fits well, its residuals should have little or no association with the values of the regressor variable.

As Kim et al. (2007) note, GUIDE has five properties that make it desirable for the analysis and interpretation of large datasets: (1) negligible bias in split variable selection; (2) sensitivity to curvature and local pairwise interactions between predictor variables; (3) applicability to numerical (continuous) and categorical variables; (4) choice of simple linear, multiple, best, Poisson, or quantile regression models (and proportional hazard analysis); and (5) choice of three roles for each numerical predictor variable (split selection only, regression modeling only, or both). Another strength of GUIDE is the boot-strap adjustment of p-values, which is important consideration when dealing with small sample sizes often encountered with industrial data. Preliminary versions of the GUIDE algorithm are described in Chaudhuri et al. (1994) and Chaudhuri (2000). Additional documentation can be found in Loh (2006), Kim et al. (2007), Loh (2007a), Loh (2007b), Loh et al. (2007) and at the web-site www.stat.wisc.edu/~ loh/ referenced 10/5/07.

Since the main advantage of a regression tree over other models is the ease with which the model can be interpreted, it is important that the construction method be free of selection bias (Loh 2002). GUIDE achieves this goal by employing a lack-of-fit test followed by a bootstrap adjustment of the p-values which is critical because parametric p-values are data-size dependent (Loh 2007a).[8]

---

[8] Bootstrapping is the practice of estimating properties of an estimator (such as its variance) by measuring those properties when sampling from an approximating distribution. One standard choice for approximating a distribution is the empirical distribution of the observed data. The advantage of bootstrapping over analytical method is its great simplicity - it is straightforward to apply the bootstrap to derive estimates of standard errors and confidence intervals for complex estimators of complex parameters of the distribution, such as percentile points, proportions, odds ratio, and correlation coefficients (http://en.wikipedia.org/wiki/Bootstrapping_%28statistics%29 referenced 10/5/07). Bootstrapping is distinguished from the jackknife procedure used to detect outliers, and v-fold cross-validation used to make sure that results are repeatable. Bootstrapping is becoming popular because it does not require the normality assumption to be met, and because it

Decision trees represent a contemporary scientifically-based decision-making method for forest products practitioners interested in improving the understanding of industrial process. Decision trees represent an "off the shelf" technology and may be superior to MLR models (and kernel methods) when data are non-homogeneous.

Predictive Modeling of Engineered Wood Panels

Engineered wood panel manufacturing may have a large number of differing, but interdependent, process variables that may have complex functional forms which influence strength properties.[9] Wood passes through many processing stages that may influence final strength properties. Key process parameters may include mat-forming consistency, line speed, press temperature, press closing rates, wood chip dimensions, fiber dimension, fiber-resin formation, etc. At the time of production, the quality of engineered wood is unknown, i.e., samples are analyzed at a later time in a lab using destructive testing. The time span between destructive tests may vary from two to six hours depending on the type of product. Hours of unacceptable engineered wood production may go undetected between these tests. Many engineered wood panel producers create a hedge of higher than needed density targets to make up for the lack of product quality knowledge between destructive tests. As a consequence, high density targets as a hedge require higher than necessary resin, wood fiber and energy inputs. In an era of strong market competition, higher than necessary density targets are not sustainable or competitive in the long term.

---

can be effectively utilized with smaller sample sizes ($n < 20$), http://en.wikipedia.org/wiki/Bootstrapping_%28statistics%29 (referenced 10/5/07).

[9] Strength properties are usually determined from destructive testing, e.g., internal bond tensile strength, maximum load flexure strength, maximum deflection flexure strength, modulus of rupture flexure strength, modulus of elasticity flexure strength, etc.

Process variables may exert simple linear univariate effects on final product quality characteristics while others may produce non-linear multivariate effects. These effects may be dynamic and be dependent on wood furnish, temperature, line speed, tool sharpness, etc. Some work has been initiated in real-time data mining and predictive modeling of final product quality characteristics of forest products using statistical methods (Young 1997; Bernardy and Scherff 1998, 1999; Gruebel 1999; Erilsson et al. 2000; Young and Guess 2002; Young and Huber 2004; Clapp et al. 2007). Erilsson et al. (2000) discussed the potential of statistical models for engineered wood manufacture, while Gruebel (1999) documented MDF manufacturing cost savings of 5 to 10 percent from the use of "off-line" first-order statistical models (e.g., faster line speeds, reduced raw material inputs, reduced energy usage, etc.). There is no evidence from the literature of the use of decision tree methods for analyzing or predicting strength properties of engineered wood panels.

Other research has investigated non-statistical heuristic models (e.g., genetic algorithms and neural networks) to develop real-time predictions of product quality characteristics of forest products (Cook and Wolfe 1991, Cook and Chiu 1997, Estévez et al. 2003, Toivanen et al. 2003, Young et al. 2004 and Dawson et al. 2006). Much work is published on theoretical models that explain final product quality characteristics (Wu and Piao 1999; Xu 2000; Humphrey and Thoemen 2000; Barnes 2001; Shupe et al. 2001; Zombori et al. 2001). Applications of theoretical models for wood composites manufacture are not evident in the literature.

Medium Density Fiberboard (MDF)

Large-scale production of MDF began in the 1980s. MDF is an engineered wood product formed by breaking down softwood (*Pinus* spp.) into wood fibers, often in a

defibrator (i.e.,"refiner"), combining it with wax and resin, forming mats and applying high temperature and pressure to create panels. MDF is a popular wood composite material. MDF advantages are that it has uniform density and that it has a surface that is smooth and free of grain patterns and defects. Its smooth surface makes it an excellent base for laminates for counter tops and cabinets. MDF is a non-structural panel that has extensive use in furniture, shelving, laminate flooring, decorative molding and doors (Figure 2.3).

MDF's name comes from the distinction in densities. MDF typically greater than ½" in thickness has a density of 600-800 kg/m³ (38-50 lbs/ft^3). High-density fiberboard (less than ½" thickness) has a density of 500-1,400 kg/m³ (31-90 lbs/ft^3). An illustrative comparison of the densities of wood composites and natural solid wood is presented in (Figure 2.4).

Some people prefer using MDF over regular lumber because it has a lower impact on the environment. MDF is made from cellulosic waste products, which sometime are dumped in landfills. This attraction has helped MDF gain popularity among homeowners (www.wisegeek.com/what-is-mdf.htm referenced 10/5/07).

One contentious issue for MDF is the use of formaldehyde (HCHO) resins and the associated health risks. MDF contains a higher resin-to-wood ratio than any other urea-formaldehyde (UF) wood composites and is the highest formaldehyde-emitting wood composite. Under U.S. Department of Housing and Urban Development (HUD) rule 24 CFR, HCHO emissions are limited to 0.2 parts per million (ppm) for floor underlayment and manufactured home floor decking, and 0.3 ppm for other products (TECO 2007).

In June of 2004, the International Agency for Research into Cancer (part of World Health Organization), upgraded formaldehyde from category 2A (probably carcinogenic to

humans) to Category 1 (carcinogenic to humans), see Sharp (2004). Reclassification is based on evidence of increased incidence of the relatively rare, nasopharyngeal cancer among individuals exposed in the past to high levels of formaldehyde (Sharp 2004).

The litigation potential of formaldehyde poisoning from wood composites is illustrated in Spake (2007). An estimated 275,000 Americans are living in more than 102,000 mobile homes that FEMA purchased for $2.6 billion after hurricane Katrina (Spake 2007). A class-action lawsuit was filed against FEMA and some trailer manufacturers in Louisiana in June 2006 on behalf of residents suffering from respiratory and flu-like illnesses they attribute to formaldehyde inside their trailers (Spake 2007).

Oriented Strand Board (OSB)

Oriented strand board (OSB) is a structural engineered wood composite panel consisting of mats formed from resinated wood strands of approximately 0.030" (inches) in thickness, 2" in width and 4" in length. The mats are pressed into panels under heat and pressure in multi-opening or continuous presses. OSB is a structural product used in residential and non-residential construction for sheathing in walls, floors and roofs (Figure 2.5). OSB is the most commonly used structural engineered wood panel in new residential housing construction in North America (http://www.osbguide.com/faqs/faq1.html referenced 10/5/07).

The OSB industry is currently experiencing unprecedented growth in North America in new mill startups and mill capacity expansion. Since 1990, new startups of mills have increased by 85 percent to 65 mills, while production capacity has increased by more than 100%, to a record 28 billion square feet per year (Adair 2005). OSB is aggressively replacing plywood as the primary sheathing demanded in North America.

Approximately 65 percent of the 43 billion square feet of construction sheathing used in 2005 consisted of OSB, while the remaining 35 percent consisted of plywood sheathing (Adair 2005). Note that 73 percent of all OSB sheathing is used in residential housing construction. Residential housing construction in the U.S. is predicted to decline from a record of almost 2.0 million annual new housing starts in 2005 to approximately 1.8 million housing starts by 2010 (Adair 2005). The decline in housing starts, in conjunction with recent OSB capacity expansion, will put downward pressure on OSB market prices and producers will be forced to improve efficiency. These market pressures will require OSB manufacturers to maintain a strong focus on reliability, quality and cost (Wang et al. 2007). The methods and results from this dissertation, if adopted, are directly beneficial to practitioners in the wood composites industry.

Appendix to Chapter II

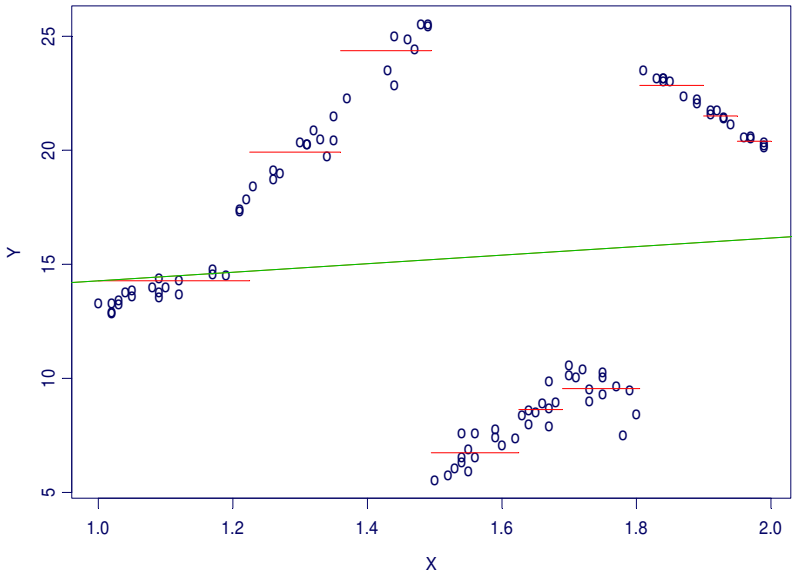


Figure 2.1. Illustration of MLR fit and DT piecewise linear fit to non homogeneous data (Kim et al. 2007).

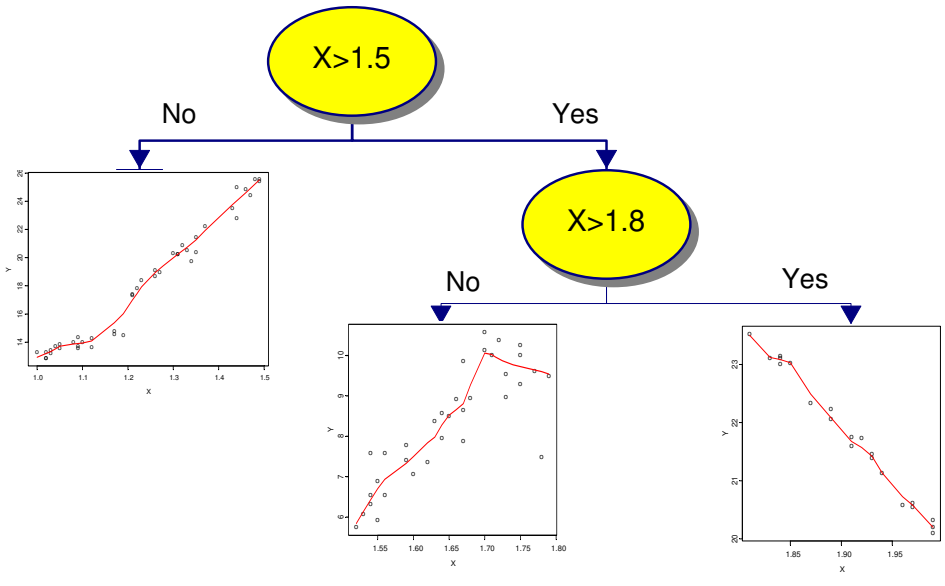


Figure 2.2. Illustration of a decision tree (Kim et al. 2007).

Figure 2.3. Illustration of applications of MDF (http://images.google.com/images?hl=en&q=mdf&gbv=2 referenced 10/5/07).

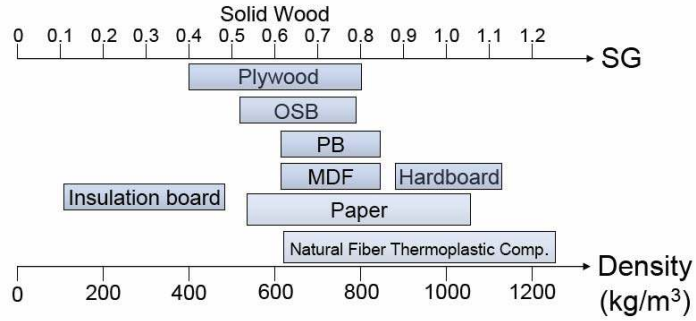


Figure 2.4. Illustration of densities for solid wood, MDF and WPC (http://pas.ce.wsu.edu/CE546/Lectures/Lecture1-Aug2006.pdf referenced 10/5/07).

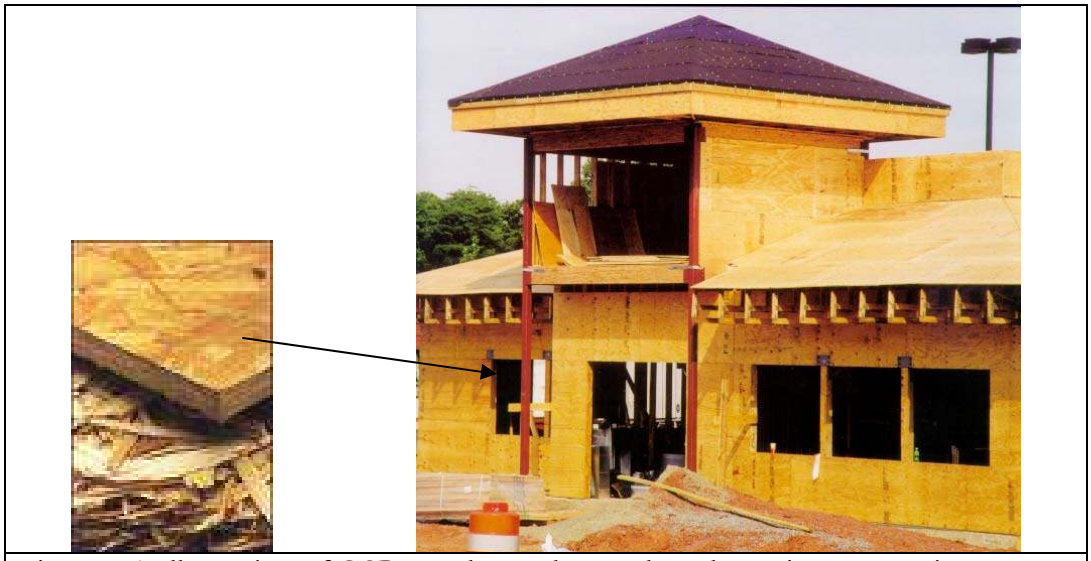


Figure 2.5. Illustration of OSB wood strands, panels and uses in construction.

CHAPTER III. METHODS

Multiple Linear Regression

One of the most traditional and popular statistical methods is multiple linear regression (MLR). For situations where the data are drawn from reasonably homogeneous populations and the response (Y) is normally distributed, traditional methods such as MLR can yield insightful analyses. The usefulness of MLR can breakdown quickly if these stringent assumptions are not met.

Most practitioners use first-order multiple linear regression (MLR) models of the form:

$$Y = \alpha + \beta X + \varepsilon \qquad [3.1]$$

where, $\mathbf{Y}$ is an ($n \times 1$) vector of dependent observations, $\mathbf{X}$ is an ($n \times p$) matrix of independent variables of known form, $\boldsymbol{\beta}$ is an ($p \times 1$) vector of parameters, $\boldsymbol{\varepsilon}$ is an ($n \times 1$) vector of errors. From a practical perspective second- and third-order MLR models of the form:

$$Y = \alpha + \beta X + \gamma X^2 + \varepsilon, \qquad [3.2]$$

and

$$Y = \alpha + \beta X + \gamma X^2 + \delta X^3 + \varepsilon \qquad [3.3]$$

may be more helpful. In this research, MLR models are investigated up to third-order models with interaction terms.

The least squares method is fundamental to MLR and is used to find an affine function that best fits a given set of data.[10] The least squares method is very defendable by minimizing the sum of the *n* squared errors (SSE) of the predicted values on the fitted line.

A key step in using MLR for data mining is to develop a model building or best model criteria. A popular model building method for MLR is stepwise regression. Stepwise regression was introduced by Efroymson (1960) and was intended to be an automated procedure for selecting the most statistically significant variables from a large pool of explanatory variables. The mixed selection procedure is the most defendable stepwise procedure and it is a combination of the forward selection and backward elimination procedures.[11] In stepwise regression the user specifies the probabilities (α) for an independent variable "to stay" and also the probabilities "to leave" the model.

Stepwise regression is typically used in conjunction with a "best model criteria." Young and Guess (2002) found multicollinearity and heteroscedasticity (i.e., unequal variances of the residuals) to be significant problems when modeling MDF product quality from real-time data. Young and Guess (2002) used the following "best model criteria" which is also used in this research:

---

[10]An affine (from the Latin, *affinis*, "connected with") subspace of a vector space (sometimes called a linear manifold) is a coset of a linear subspace. A linear subspace of a vector space is a subset that is closed under linear combinations, e.g., linear regression equation of a linear subspace (http://mathworld.wolfram.com/AffineFunction.html. referenced 10/4/07).

[11] The forward selection procedure attempts to insert variables until the regression is satisfactory (Draper and Smith 1981). The order of insertion is determined by using the partial correlation coefficient as a measure of the importance of variables not yet in the equation (Neter et al. 1996). This starts by finding the most correlated independent variable (*x*) with Y, and so forth. The overall regression is checked for significance, the improvements in R^2 value and the partial F-values are noted. The partial F-values are compared with an appropriate F percentage point and the corresponding independent variables are retained or rejected from the model according to significance. This continues until a suitable the first-order linear regression equation is developed (Kutner et al. 2004, Neter et. al. 1996, and Myers 1990).

- maximum Adjusted R^2_a,
- minimum Akaike's Information Criterion (AIC), Akaike (1974),
- Variance Inflation Factor (VIF) < 10,
- significance of independent variables p-value < 0.10,
- absence of heteroscedasticity in residuals, $E(\varepsilon_i) = 0$,[12]
- examination or residual plots,
- root mean square error of prediction (RMSEP),
- XY scatter plot of predicted and observed values for the validation data set.

Adjusted R^2_a is a better measure for building models with the potential of a large number of independent variables than is the coefficient of determination (R^2). R^2 will always increase as an additional independent variable is added to the model, where R^2_a will only increase if the residual sum of squares decreases (Draper and Smith 1981). R^2_a minimizes the risk of "over-fitting" and penalizes for it. AIC protects against model bias and VIFs less than ten protect against multicollinearity and the development of ill-founded models.

An important procedure for variable reduction is called "all possible subsets" or "all possible regression procedure" (Neter et al. 1996). The purpose of all possible subsets is to identify a small group of regression models that are "good" according to a specified criterion, e.g.., criteria specified by Young and Guess (2002). The limited number of regressors might consist of three to six good subsets according to specified criteria. The all possible subsets approach assumes that the number of observations exceeds the

---

[12] $R^2_a = 1 - \left(1 - R^2\right)\left(\frac{n-1}{n-p}\right)$, $0 \le R^2_a \le 1$; $R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{SSE}{SSTO}$, $0 \le R^2 \le 1$

$C_p = \left(\frac{SSE_p}{s^2}\right) - (N - 2p)$; $AIC = n \ln\left(\frac{SSE}{n}\right) + 2p$.

maximum number of parameters ($n > p$), and that n should be six to ten times larger than p (Kutner et al. 2004).

Plotting residuals $(e = \hat{y} - y)$ is an important diagnostic for checking model quality (Draper and Smith 1981). Departures from normal errors can detect if:

- the regression function is not linear,
- the error terms do not have a constant variance,
- the error terms are not independent,
- the model fits all but one or a few outlier observations,
- the error terms are not normally distributed,
- one or several important predictor variables have been omitted from the model.

A key method for assessing the quality of model predictions of the validation data set is cross-validation (Kutner et al. 2004). A validation sample is simply a sample that is withheld from the estimation of a regression model. The model developed is then used to predict the true values of the records withheld. Statistics such as $R^2_{validation}$ (coefficient of determination for the validation sample) and root mean square error of the predicted (RMSEP) are calculated for the validation data set to compare the performance of the training models. The RMSEP statistic is:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}} \qquad [3.4]$$

where, n is the number of observations, Y_i is the i-th actual value and $\hat{Y}_i$ is the i-th predicted value.

Box Cox Transforms of *Y*

Sometimes a transformation on the response Y fits the model better than the original response. A commonly used transformation raises the response Y to some power.

Models are investigated in this research using the Box Cox transforms of the response Y (Box and Cox 1964). Box and Cox (1964) formalized a family of power transformations. The formula for the transformation is constructed so that it provides a continuous definition and the error sums of squares are comparable:

$$Y^{(\lambda)} = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda \dot{Y}^{\lambda - 1}} & \text{if } \lambda \neq 0, \\ \dot{Y} \ln(Y) & \text{if } \lambda = 0 \end{cases} \qquad [3.5]$$

where $\dot{Y}$ is the geometric mean.[13] The plot of Y^{λ} and λ (Figure 3.1) illustrates the effect of this family of power transformations on Y.

Quantile Regression Trees

For $0 < \alpha < 1$, quantile regression analysis focuses on the conditional α-th quantile of the response Y given a covariate vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)$. Unlike usual regression analysis, which focuses only on the mean of Y given $\mathbf{X}$, quantile regression is capable of providing insight into the median as well as the lower and upper tails of the conditional distribution of the response with varying choices of α (Chaudhuri and Loh 2002). As a

---

[13] The geometric mean of a collection of positive data is defined as the nth root of the product of all the members of the data set, where n is the number of members. The geometric mean of a data set $[a_1, a_2, \ldots, a_n]$ is given by:

$$\left(\prod_{i=1}^{n} a_i \right)^{1/n} = \sqrt[n]{a_1 \cdot a_2 \cdot \ldots \cdot a_n} .$$

The geometric mean of a data set is smaller than or equal to the data set's arithmetic mean (the two means are equal if and only if all members of the data set are equal).

result, quantile regression is quite effective as a tool for exploring and modeling the nature of dependence of a response on the covariates when the covariates have different effects on different parts of the conditional distribution of the response (Chaudhuri and Loh 2002).

Examining causality between process variables and product quality characteristics beyond the mean of the distribution is an important issue for forest products manufacturers. Most forest products manufacturers (especially wood composites manufacturers) have a strong interest in understanding the lower percentiles (failures, safety risk, claims, etc.) of the distribution of manufactured product quality.

Traditionally, MLR is used to study causality between independent variables and the average of a response variable, with an important goal of making useful predictions of the response variable. However, there are three important assumptions of the MLR approach: 1) assumption of linearity; 2) the assumption of a normal or Gaussian distribution for the response variable; and 3) models the mean of the distribution of the response variable.

Quantile regression (QR) is intended to offer a comprehensive strategy for completing the regression picture (Koenker 2005). As Mosteller and Tukey (1977) note in their influential text, as cited by Koenker (2005): "…the regression curve gives a grand summary for the averages of the distributions corresponding to the set of *X*s…and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions."

Quantile Regression (QR) is an approach that allows us to examine the behavior of the target variable (*Y*) beyond its average of the Gaussian distribution (e.g., median or 50th percentile, 10th percentile, 80th percentile, 95th percentile, etc.). Examining causality of the median and average tendencies of the distribution of a product quality characteristic may yield different conclusions. Shaffer et al. (2007) and Young et al. (2007b) note that independent variables influencing the response variable of the IB of MDF varied dramatically by quantile. In some cases the sign and strength of the coefficient of similar independent variables explaining IB was reversed by quantile.

The QR model does not require the product quality characteristics to be normally distributed and does not have the other rigid assumptions associated with MLR. The first-order QR model has the form:

$$Q_{y_i}\left\langle \tau \middle| x \right\rangle = \beta_0 + x_i\beta_1 + F_u^{-1}(\tau) \qquad [3.6]$$

where, Q_{y_i} is the conditional value of the response variable given τ in the i[th] trial, β_o is the intercept, β_1 are parameters, τ denotes the quantile, x_i is the value of the independent variable in the i[th] trial, F_u is the common distribution function (e.g., normal, Weibull, lognormal, other, etc.) of the error given τ, $E(F_u^{-1}(\tau)) = 0$, for i = 1 ,…, n (Koenker 2005).

Just as we can define the sample mean as the solution to the problem of minimizing a sum of squared residuals, we can define the median as the solution to the problem of minimizing a sum of absolute residuals (Koenker and Hallock 2001). The symmetry of the piecewise linear absolute value function implies that the minimization of the sum of absolute residuals must equate the number of positive and negative residuals, thus assuring that there are the same number of observations above and below the median

(Koenker and Hallock 2001). Minimizing the sum of asymmetrically-weighted absolute residuals yields the quantiles (Koenker and Hallock 2001). Solving

$$\min \sum_{i=1}^{n} \rho_{\tau}(y_i - \xi), \qquad [3.7]$$

where the function $\rho_{\tau}(|)$ is a tilted absolute value function that yields the τth sample quantile as its solution (Koenker and Hallock 2001), Figure 3.2. To obtain an estimate of the conditional median function in quantile regression, we simply replace the scalar ξ in equation [3.7] by the parametric function $\xi(x_i, \beta)$ and set τ to ½.[14] To obtain estimates of the other conditional quantile functions, replace absolute values by $\rho_{\tau}(|)$ and solve,

$$\hat{\beta}(\tau) = \min \sum_{i=1}^{n} \rho_{\tau}(y_i - \xi(x_i, \beta)) \qquad [3.8]$$

For any quantile $\tau \in (0,1)$. The quantity $\hat{\beta}(\tau)$ is called the τth regression quantile.

QR is an important non-parametric statistical method for forest products practitioners interested in exploring causality beyond the mean of the distribution. A strength of GUIDE decision trees is that it allows for quantile regression fits in the terminal nodes of the trees. Use of quantile regression trees for examining causality of engineered wood panel strength properties is not documented in the literature.

Decision Trees and the GUIDE Method

As noted earlier, a regression tree is a piecewise constant or piecewise intrinsically linear estimate of a regression function, constructed by recursively partitioning the data and sample space (Loh 2002). A decision tree partitions the data space of all joint

[14] Variants of this idea were proposed in the mid-eighteenth century by Boscovich and subsequently investigated by Laplace and Edgeworth (Koenker and Hallock 2001).

regressor values **X** into *J*-disjoint regions $\{R_j\}_1^J$ (Friedman 2001). For a given set of joint regressor values **X**, the tree prediction $\hat{Y} = T_j(\boldsymbol{X})$ assigns as the response estimate, the value assigned to each region containing **X:**

$$X \in R_j \Rightarrow T_j(\boldsymbol{X}) = \hat{y}_j \ . \qquad [3.9]$$

Given a set of regions, the optimal response values associated with each region are easily obtained, namely the value that minimizes prediction error (risk) in that region:

$$\hat{Y}_j = \arg\min_{y'} E_y[L(y, y') \mid \boldsymbol{X} \in R_j]. \qquad [3.10]$$

As noted by Friedman (2001), the difficult problem is to find a good set of regions $\{R_j\}_1^J$. Unlike kernel methods (e.g., rudimentary kernel method such as multiple linear regression and assumption of homogeneity), decision trees attempt to use the data to estimate a good partition instead of a user defined model.

GUIDE can recursively partition a dataset and fit a constant, best, multiple linear or quantile regression model to the observations in each partition. GUIDE first constructs a nested sequence of tree-structured models and then uses v-fold cross-validation to select the smallest tree-structure which has an estimated mean prediction deviance that lies within a minimum variance (e.g., standard error of training data set or standard error of both training and validation set ~ "global range"). GUIDE employs the Pearson chi-square (χ^2) test to detect lack-of-fit of the residuals in choosing a regressor variable to partition at each stage. As Loh (2006b) notes, GUIDE does not have the selection bias of CART (Breiman et al. 1984) and other tree algorithms that rely solely on greedy search optimization.

The GUIDE algorithms (Loh 2002b) for fitting piecewise constant and piecewise linear models are:

Algorithm 1. Chi-square tests for constant fit:

1. Obtain the residuals from a constant model fitted to the *Y* data.

2. For each numerical-valued variable, divide the data into four groups at the sample quartiles; construct a 2 x 4 contingency table with the signs of the residuals (positive versus non-positive) as rows and the groups as columns; count the number of observations in each cell and compute the χ^2 statistic and its theoretical p-value from a χ^2_3 distribution. We refer to this as a curvature test.

3. Do the same for each categorical variable, using the categories of the variable to form the columns of the contingency table and omitting columns with zero column totals.

4. To detect interactions between a pair of numerical-valued variables (X_i, X_j), divide the (X_i X_j)-space into four quadrants by splitting the range of each variable into two halves at the sample median; construct a 2 x 4 contingency table using the residual signs as rows and the quadrants as columns; compute the χ^2 statistic and p-value. Again, columns with zero column totals are omitted.

5. Do the same for each pair of categorical variables, using their value pairs to divide the sample space. For example, if X_i and X_j take c_i and c_j values, respectively, the χ^2 statistic and p-value are computed from a table with two rows and number of columns equal to c_ic_j less the number of columns with zero totals.

6. For each pair of variables (X_i, X_j) where X_i is numerical-valued and X_j is categorical, divide the X_i-space into two at the sample median and the X_j space into as many sets as the number of categories in its range (if X_j has *c* categories, this splits the (X_i, X_j) space into 2*c* subsets);
construct a 2 x 2*c* contingency table with the signs of the residuals as rows and the subsets as columns; compute a χ^2 statistic and p-value for the table after omitting columns with zero totals.

If the smallest *p*-value is from a curvature test, it is natural to select the associated *X* variable to split the node. If the smallest *p*-value is from an interaction test, one of the two interacting variables is selected. In order to fit a constant model in each node, the choice of variable is based on reduction in SSE.

Algorithm 2. Choose between interacting pair of X variables (Loh 2002):

Suppose that a pair of variables is selected because their interaction test is the most significant among all the curvature and interaction tests.

1. If both variables are numerical-valued, the node is split in turn along the sample mean of each variable; for each split, the SSE for a constant model is obtained for each subnode; the variable yielding the split with the smaller total SSE is selected.

2. Otherwise if at least one variable is categorical, the one with the smaller curvature p-value is selected.

If a variable from a significant interaction is selected to spilt a node, the other variable is not automatically required in the pair to split in the children nodes. Instead, it competes for splits at every node with all of the other variables that remain.

To complete the tree construction algorithm the split points are selected as well as the size of the tree.

The GUIDE algorithms (Loh 2006) for fitting piecewise linear models via stepwise regression are:

Algorithm 1. (Tree construction). These steps are applied recursively to each node of the tree, starting with the root node that holds the whole dataset.

1. Let t denote the current node. Fit a simple linear regression to each predictor variable in the data in t. Choose the regressor yielding the smallest residual mean squared error and record its model R^2.

2. Stop if $R^2 > 0.99$ or if the number of observations is less than $2n_0$, where n_0 is a small user-specified constant. Otherwise, go to the next step.

3. For each observation associated with a positive residual, define the class variable $Z = 1$; else define $Z = 0$.

4. Use Algorithm 2 to find a variable X' to split t into left and right subnodes t_L and t_R.

    a) If X' is ordered, search for a split of the form $X' \le x$. For every x such that t_L and t_R contain at least n_0 observations each, find

S, the smallest total sum of squared residuals obtainable by fitting a simple linear model to the data in t_L and t_R separately. Select the smallest value of x that minimizes S.

b) If X′ is categorical, search for a split of the form X′ ∈ C, where C is a subset of the values taken by X′. For every C such that t_L and t_R have at least n_0 observations each, calculate the sample variances of Z in t_L and t_R. Select the set C for which the weighted sum of the variances is minimum, with weights proportional to sample sizes in t_L and t_R .

5. Apply step 1 to t_L and t_R separately.

Algorithm 2. (Split variable selection):

1. Use Algorithms 3 and 4 (next page) to find the smallest curvature and interaction p-values $p^{(c)}$ and $p^{(i)}$ and their associated variables $X^{(c)}$ and $\{X_1^{(i)}, X_2^{(j)}\}$

2. If $p^{(c)} \le p^{(i)}$, define $X' = X^{(c)}$ to be the variable to split t.

3. Otherwise, if $p^{(c)} > p^{(i)}$, then:

a) If either $X_1^{(i)}$ or $X_2^{(j)}$ is categorical, define $X' = X_1^{(i)}$ if it has the smaller curvature p-value; otherwise, define $X' = X_2^{(i)}$

b) Otherwise, if $X_1^{(i)}$ and $X_2^{(j)}$ are both ordered variables, search over all splits of t along $X_1^{(i)}$. For each split into subnodes t_L and t_R, fit a simple linear model on $X_1^{(i)}$ to the data in t_L and t_R separately and record the total sum of squared residuals. Let S_1 denote the smallest total sum of squared residuals over all possible splits of t on $X_1^{(i)}$. Repeat the process with $X_2^{(i)}$ and obtain the corresponding smallest total sum of squared residuals S_2. If $S_1 \le S_2$, define $X' = X_1^{(i)}$; otherwise, define $X' = X_2^{(i)}$.

Algorithm 3. (Curvature tests):

1. For each predictor variable X:

a) Construct a $2 \times m$ cross-classification table. The rows of the table are formed by the values of Z. If X is a categorical variable, its values define the columns, i.e., m is the number of distinct values of X. If X is quantitative, its values are grouped

into four intervals at the sample quartiles and the groups constitute the columns, i.e., $m = 4$.

b) Compute the significance probability of the chi-squared test of association between the rows and columns of the table.

2. Let $p^{(c)}$ denote the smallest significance probability and let $X^{(c)}$ denote the associated X variable.

Algorithm 4. (Interaction tests):

1. For each pair of variables X_i and X_j , carry out the following interaction test:

a) If X_i and X_j are both ordered variables, divide the (X_i,X_j)-space into four quadrants by splitting the range of each variable into two halves at the sample median; construct a 2×4 contingency table using the Z values as rows and the quadrants as columns. After dropping any columns with zero column totals, compute the chi-squared statistic and its p-value.

b) If X_i and X_j are both categorical variables, use their value-pairs to divide the sample space. For example, if X_i and X_j take c_i and c_j values, respectively, the chi-squared statistic and p-value are computed from a table with two rows and number of columns equal to c_ic_j less the number of columns with zero totals.

c) If X_i is ordered and X_j is categorical, divide the X_i-space into two at the sample median and the X_j-space into as many sets as the number of categories in its range—if X_j has c categories, this splits the (X_i, X_j)-space into $2c$ subsets. Construct a $2\times2c$ contingency table with the signs of the residuals as rows and the $2c$ subsets as columns. Compute the chi-squared statistic and its p-value, after dropping any columns with zero totals.

2. Let $p^{(i)}$ denote the smallest p-value and let $X_1^{(i)}$ and $X_2^{(i)}$ denote the pair of variables associated with $p^{(i)}$.

After tree building terminates, the tree is pruned using v-fold cross-validation. Let E_0 be the smallest v-fold cross-validation estimate of prediction mean squared error and let α be a positive number. The smallest subtree is selected whose v-fold cross-validation estimate of mean square error is within α times the standard error of E_0 (Loh 2006).

The split selection approach is different from that of CART, which constructs piecewise constant models only and which searches for the best variable to split and the best split point simultaneously at each node. CART's variable selection is inherently biased toward choosing variables that permit more splits (Loh 2006). GUIDE does not have such bias because it uses p-values for variable selection (Loh 2006).

v-fold cross-validation

This type of v-fold cross-validation is useful when no test sample is available and the learning sample is too small to have the test sample taken from it. The user-specified v value for v-fold cross-validation (its default value is 10) determines the number of random subsamples, as equal in size as possible, that are formed from the training (learning) sample. A tree of the specified size is computed v times, each time leaving out one of the subsamples from the computations, and using that subsample as a test sample for v-fold cross-validation, so that each subsample is used (v - 1) times in the learning sample and just once as the test sample. The v-fold cross-validation costs are computed for each of the v test samples and then averaged to give the v-fold estimate of the v-fold cross-validation costs Breiman et al. (1984).

The total number of cases is divided into v subsamples Z_1, Z_2, ..., Z_v of almost equal sizes. The subsample Z - Z_v is used to construct the predictor d. Then the v-fold cross-validation estimate is computed from the subsample Z_v in the following way:

$$R^{CV}(d) = \frac{1}{N_v} \sum_{v} \sum_{(x_n, y_n) \in Z_v} (y_i - d^{(v)}(x_n))^2 \qquad [3.11]$$

where $d^{(v)}(x)$ is computed from the sub sample Z - Z_v.

GUIDE Importance Ranking

The GUIDE importance ranking (Loh 2007) is used in this dissertation to reduce the number of regressors and increase the dimensionality in the data set for the use of quantile regression in each terminal node of the regression tree. The GUIDE importance ranking is estimated by:

1. Fit a constant model to the data in the node and obtain residuals,

2. Cross-tabulate each categorical variable with the signs of the residuals,

3. Discretize each numerical variable X into four groups at the quartiles and cross-tabulate with the signs of the residuals,

4. Compute each χ^2_v-value and convert it to a χ^2_1-value using two applications of the Wilson-Hilferty (1931) approximation:

$$\chi^2_1 = \max\left(0, \left[\frac{7}{9} + \sqrt{v}\left\{\left(\frac{\chi^2_v}{v}\right)^{1/3} - 1 + \frac{2}{9v}\right\}\right]^3\right) \qquad [3.12]$$

6. Select the variable with largest value χ^2_1 to split the node.

Importance score of X_i is:

$$IMP(i) = \sum_t \sqrt{n(t)}\chi^2_1(t,i) \qquad [3.13]$$

where, summation is over all intermediate nodes t,
n(t) is the sample size at node t,
$\chi^2_1(t,i)$ is the Wilson-Hilferty chi-squared value of X_i at t.

Data Set Description

Medium Density Fiberboard (MDF)

The MDF data set is from a southeastern U.S. manufacturer. The plant has an annual capacity of 100 million sq. ft., 5/8 inch basis and a 12-platen multi-opening ("day-light") press. The plant started production in early 1993. The primary mechanical property defining strength quality for MDF producers and their customers is internal bond (IB). Destructive tests of IB are taken at approximate time intervals of two hours from one press platen (#8) while the production line is running.

The real-time data fusion database of Young and Guess (2002) as refined by Dawson et al. (2006) is used for this dissertation research (Appendix A). Three nominal products of the producer are used in this research. The products are 0.750 inches ("), 0.625" and 0.500", all industrial ("IND") grade. The manufacturer requested that the densities of these "IND" products not be reported in the public domain. The 0.500" product has 209 records dating from 11/25/05 to 09/30/06; 0.625" product has 517 records dating from 11/15/05 to 10/18/06; and the 0.750" product has 245 records dating from 11/16/05 to 10/14/06. Five records are eliminated from the 0.750" product set given an excessive level of null fields in the records (#639 - 1/4/06; #786 - 1/19/06; #1526 - 4/7/06; #1526 – 4/7/06; a duplicate record, and #1642 - 4/26/06). Two records are eliminated from the 0.625" product set given an excessive level of null fields and possible incorrect product classification (#1878 - 5/17/06; #2996 – 9/13/06). No records are eliminated from the 0.500" product set.

The MDF data set has 183 independent variables (regressors) that correspond to real-time sensors on the production line. Sensor data are time-lagged in the data to reflect

the location of the sensor relative to the location of the press where the MDF panel is created (Dawson et al. 2006).

Oriented Strand Board (OSB)

The OSB data set is from a southeastern U.S. manufacturer. The plant has an annual capacity of 500 million sq. ft., 3/8-inch roof sheathing (RS) basis and has a 16-platen multiple opening press. The plant started production in December 2004. The primary mechanical properties defining product quality for this OSB manufacturer are IB tensile strength and Parallel Elasticity Index (EI) flexure strength. Destructive tests are taken two or three times per 12-hour shift to determine product quality of manufactured product.

The real-time data fusion database created by Young and Guess (2002) as refined by Dawson et al. (2006) is also used in this research (Appendix A). The fused database consists of destructive test data with 234 process sensor regressor variables. The data set contains 238 records from 7/27/2005 to 11/20/2006. After communication with the Technical Director, four data records are removed from the data set given that the records are defined as experimental for either a "new product trial" or resin experiment. The records are: #10106 (04/22/2006 2:58:41 PM); #9250612 (09/26/2006 8:10:44 PM); #92806316 (10/05/2006 12:41:53 PM); and #935 (11/04/2006 1:10:48 PM). Similar to the MDF data, sensor data are time-lagged in the data to reflect the location of the sensor relative to the location of the press where the OSB panel is created (Dawson et al. 2006).

In model building a general rule of thumb is to use 80 percent of the entire data set for the training (or learning) data set and the remaining 20 percent for the validation (or calibration) data set (Kutner et al. 2004). Validation records when time-ordered, as is the

case of this data, are the most current 20 percent of the records (Kutner et al. 2004). A challenge in this research is low dimensionality of both MDF and OSB data sets. Many authors note that ideal record length should be six to ten times the number of independent variables (Draper and Smith 1981, Myers 1990, Neter et al. 1996, and Kutner et al, 2004). Unfortunately, the data fusion records developed by Dawson et al. (2006) given the long sampling intervals between destructive tests did not allow for this ideal to be met. I envision future research with these companies that may facilitate improved data set dimensionality.

Data Quality and Descriptive Statistics

Medium Density Fiberboard (MDF)

0.500" Thickness. -- Descriptive statistics of IB that characterize the location, variability and shape of the 0.500" thickness data are presented. The mean (141.6 p.s.i.) and median (141 p.s.i.) for this product are similar (Table 3.1). The coefficient of variation (CV) characterizes variability of the data and a CV of 10.4 percent (standard deviation 14.8 p.s.i.) for 0.500" illustrates that the standard deviation comprises almost 10 percent of the scale of the mean. The shape of the data is characterized by skewness and kurtosis. The low kurtosis value of 0.07 indicates a distribution that has a rounded peak with wide shoulders. Skewness measures the direction and degree of asymmetry. The skewness value of 0.15 indicates some mild positive skewness (tail to the right) of the data. The histogram of 0.500" IB (Figure 3.3) indicates mild asymmetry of the distribution. The box plot, a useful visualization tool indicates that this product may have four IB values that are possible outliers, i.e., any point outside the whisker and the box are possible outliers.

Probability plots are a common graphical technique to demonstrate how a particular data set fits a specific candidate probability distribution. The data are ordered and plotted against the theoretical order statistics of the desired distribution. There is evidence that a data set conforms to a specific distribution if the data fall along the straight calibration line between the two ordered data sets.

Simultaneous confidence bands, along with pointwise confidence intervals provide objective assessments of deviation from the line (Meeker and Escobar 1998). Data points outside the confidence bands are shown to deviate from the candidate probability distribution under investigation. S-Plus and SPLIDA (www.insightful.com/splus referenced 10/5/07) are used to investigate the Normal, Logistic, Log Logistic, Log Normal, Weibull (two parameter), Largest Extreme Value, Smallest Extreme Value, Frechet and Exponential probability distributions. The log likelihood and AIC scores are also developed in this research to provide quantitative evidence of the distribution that best fits the data (Akaike 1974, Bozdogan 2000). The AIC scores and probability plots indicate that the Normal or possibly the Log Logistic distributions are reasonable fits for the IB of 0.500" MDF (Table 3.2, Figures 3.4 and 3.5).

0.625" Thickness. -- The mean (139.1 p.s.i.) and median (139 p.s.i.) for 0.625" MDF are similar and smaller than for 0.500" MDF (Table 3.3). The CV of 10.7 percent indicates a standard deviation (14.9 p.s.i.) that comprises approximately 10 percent of the scale of the mean. The low kurtosis value of 0.19 indicates a distribution that has a rounded peak with wide shoulders. The skewness value of 0.26 indicates some mild positive (tail to the right) skewness of the data which is more skewed than 0.500" MDF. The histogram of 0.625"

IB (Figure 3.6) illustrates asymmetry of the distribution. The box plot indicates that this product may have as many as 10 outliers.

The AIC scores and probability plots indicate that the Normal or possibly the Log Normal distributions are reasonable fits for the IB of 0.625" MDF (Table 3.4, Figures 3.7 and 3.8). This Normal distribution fit is similar to that of 0.500" MDF. This distribution fit is similar to the findings of Edwards (2004) for the same product but different mill.

0.750" Thickness. -- The mean (138.6 p.s.i.) and median (138 p.s.i.) for this product are very similar (Table 3.5). The location statistics for the 0.750" MDF product relative to the aforementioned products suggest that IB decreases as thickness increases for this data set. The CV of 10.8 percent for this product indicates that that standard deviation (14.9 p.s.i.) comprises more than 10 percent of the scale of the mean. The low kurtosis value of 0.15 for this product indicates a distribution that has a rounded peak with wide shoulders. The skewness value of 0.008 for this product indicates minimal skewness of the data. The histogram of 0.750" IB (Figure 3.9) indicates that the distribution of IB is approximately symmetrical. The box plot indicates that this product may have five IB values that are outliers.

There is statistical evidence that the Normal or Logistic distributions are reasonable fits for the IB of 0.750" MDF (Table 3.6). The AIC scores and probability plots are used to support this conclusion (Figures 3.10 and 3.11).

Oriented Strand Board

Internal Bond. -- The mean (46.2 p.s.i.) and median (45.4 p.s.i.) for this product are dissimilar (Table 3.7). The CV of 20.3 percent (standard deviation 9.4 p.s.i.) encompasses 20 percent of the scale of the mean. The low kurtosis value of 0.18 indicates a distribution

with positive kurtosis that is mildly leptokurtic, i.e., it has a more acute "peak" around the mean. The skewness value of 0.62 indicates a positive skewness (tail to the right) of the data. The histogram of IB (Figure 3.12) indicates an asymmetric distribution with a tail to the right. The box plot indicates seven IB values that are possible outliers, all to the right-side of the distribution.

The AIC scores and probability plots indicate that the Log Normal or Largest Extreme Values distributions are reasonable fits for the IB of 0.500" MDF (Table 3.8, Figures 3.13 and 3.14). Note the distinct difference in distribution fits relative to the Normal distribution fits of MDF products.

Parallel EI. -- The mean (59,666 lb-in^2/ft) and median (58,963 lb-in^2/ft) for 7/16" RS OSB are dissimilar (Table 3.9). A CV of 7.5 percent indicates a standard deviation (4480 lb-in^2/ft) that encompasses about seven percent of the scale of the mean. The high kurtosis value of 0.99 indicates a leptokurtic distribution with sharp peaks and wider tails. A skewness value of 0.87 indicates positive skewness (tail to the right) of the data. The histogram of 7/16" RS Parallel EI (Figure 3.15) is asymmetric and is a good illustration tool to support the kurtosis and skewness descriptive statistics. The box plot indicates that this strength property may have as many as ten outliers.

The AIC scores and probability plots indicate that the Largest Extreme Value or Log Logistic are reasonable fits for the Parallel EI of 7/16" RS OSB (Table 3.10, Figures 3.16 and 3.17). Again, note the departure in distribution fits relative to the Normal distribution fits for MDF products.

Appendix to Chapter III

Table 3.1. Descriptive statistics for the IB of 0.500" MDF.

Statistic	Value
Minimum	102
Maximum	184
Range	82
Median	141
Mean	141.63
Standard Deviation	14.79
Variance	218.62
Coefficient of Variation	10.44
Skewness	0.1462
Kurtosis	0.0728
N	209

Table 3.2. Selected model scores for the IB of 0.500" MDF.

Model Fit	Log Likelihood	AIC
Normal	-859.0	1722.0
Log Logistic	-860.3	1724.6
Logistic	-860.4	1724.8
Weibull	-870.2	1744.4
Largest Extreme Value	-871.2	1746.4
Smallest Extreme Value	-880.3	1764.6
Frechet	-882.4	1768.8
Log Normal	-895.2	1794.4
Exponential	-1244.0	2492.0

Table 3.3. Descriptive statistics for the IB of 0.625" MDF.

Statistic	Value
Minimum	97
Maximum	186
Range	89
Median	139
Mean	139.13
Standard Deviation	14.93
Variance	222.89
Coefficient of Variation	10.73
Skewness	0.2589
Kurtosis	0.1909
N	517

Table 3.4. Selected model scores for the IB of 0.625" MDF.

Model Fit	Log Likelihood	AIC
Normal	-2097	4198
Log Normal	-2098	4200
Logistic	-2101	4206
Log Logistic	-2102	4208
Weibull	-2126	4256
Largest Extreme Value	-2128	4260
Smallest Extreme Value	-2153	4310
Frechet	-2155	4314
Exponential	-3062	6128

Table 3.5. Descriptive statistics for the IB of 0.750" MDF.

Statistic	Value
Minimum	100
Maximum	177
Range	77
Median	138
Mean	138.57
Standard Deviation	14.94
Variance	223.15
Coefficient of Variation	10.78
Skewness	0.0078
Kurtosis	0.1491
N	245

Table 3.6. Selected model scores for the IB of 0.750" MDF.

Model Fit	Log Likelihood	AIC
Normal	-1010	2024
Logistic	-1010	2024
Log Logistic	-1011	2026
Log Normal	-1012	2028
Weibull	-1019	2042
Largest Extreme Value	-1030	2064
Smallest Extreme Value	-1030	2064
Frechet	-1045	2094
Exponential	-1453	2910

Table 3.7. Descriptive statistics for the IB of OSB.

Statistic	Value
Minimum	27.90
Maximum	74.82
Range	46.92
Median	45.44
Mean	46.18
Standard Deviation	9.39
Variance	88.21
Coefficient of Variation	20.34
Skewness	0.6216
Kurtosis	0.1784
N	234

Table 3.8. Selected model scores for the IB of OSB.

Model Fit	Log Likelihood	AIC
Log Normal	-1423	2850
Largest Extreme Value	-1424	2852
Log Logistic	-1429	2862
Normal	-1437	2878
Logistic	-1439	2882
Frechet	-1440	2884
Weibull	-1456	2916
Smallest Extreme Value	-1501	3006
Exponential	-1899	3802

Table 3.9. Descriptive statistics for the Parallel EI of OSB.

Statistic	Value
Minimum	47245.98
Maximum	76444.00
Range	29198.03
Median	58962.93
Mean	59665.90
Standard Deviation	4479.57
Variance	20066524.40
Coefficient of Variation	7.51
Skewness	0.8733
Kurtosis	0.9999
N	234

Table 3.10. Selected model scores for the Parallel EI of OSB.

Model Fit	Log Likelihood	AIC
Largest Extreme Value	-3841	7686
Log Logistic	-3848	7700
Frechet	-3848	7700
Log Normal	-3850	7704
Logistic	-3855	7714
Normal	-3861	7726
Weibull	-3918	7840
Smallest Extreme Value	-3942	7888
Exponential	-4715	9434

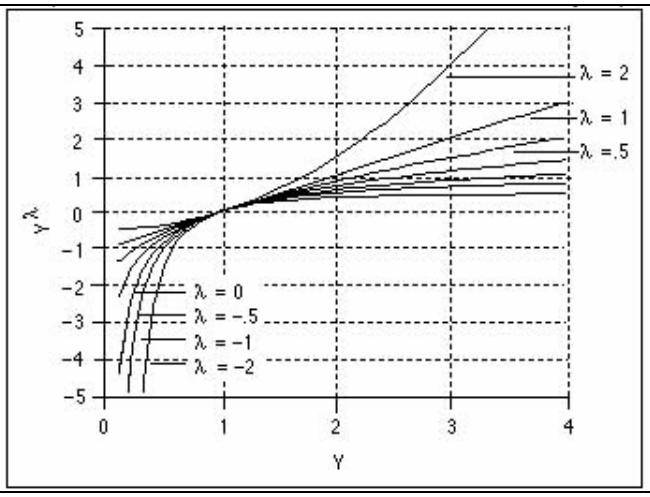


Figure 3.1. Plot of Y^{λ} and λ illustrating the effect of this family of power transformations on Y (SAS Institute, Inc. 2007).

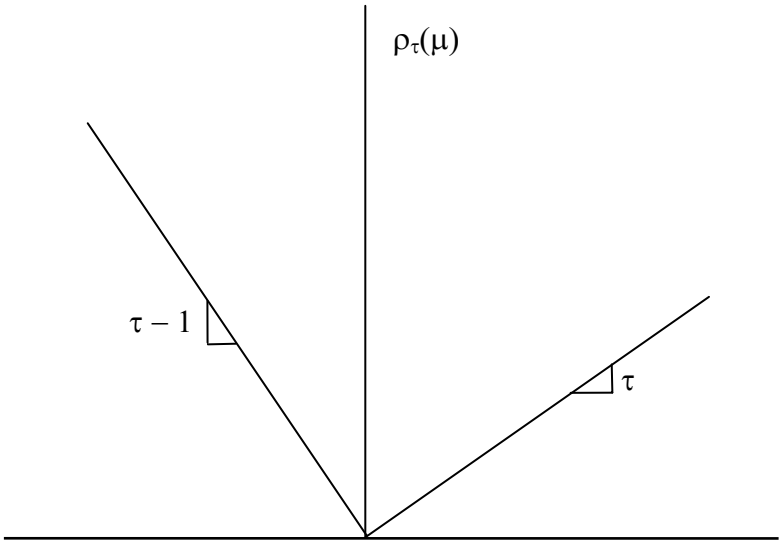


Figure 3.2. Quantile regression ρ function.

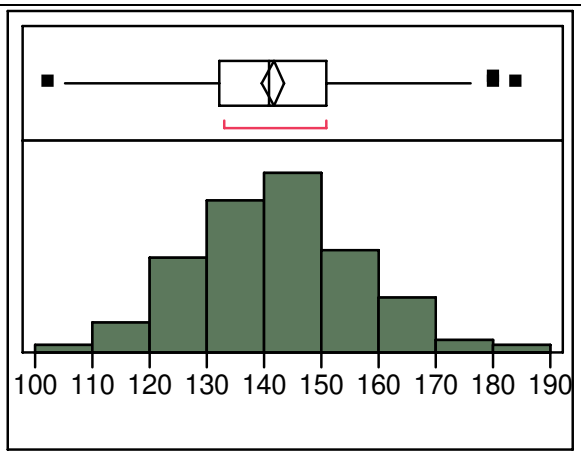


Figure 3.3. Box plot and histogram of the IB of 0.500" MDF.

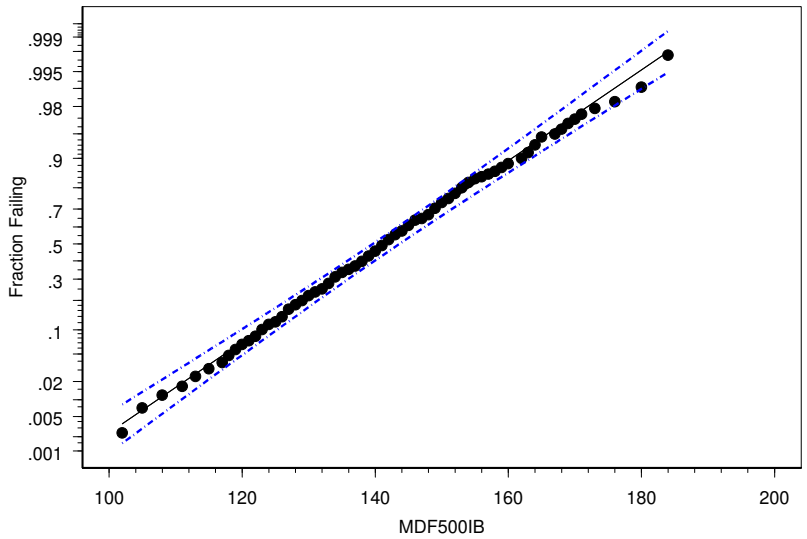


Figure 3.4. Normal probability plot of the IB of 0.500" MDF.

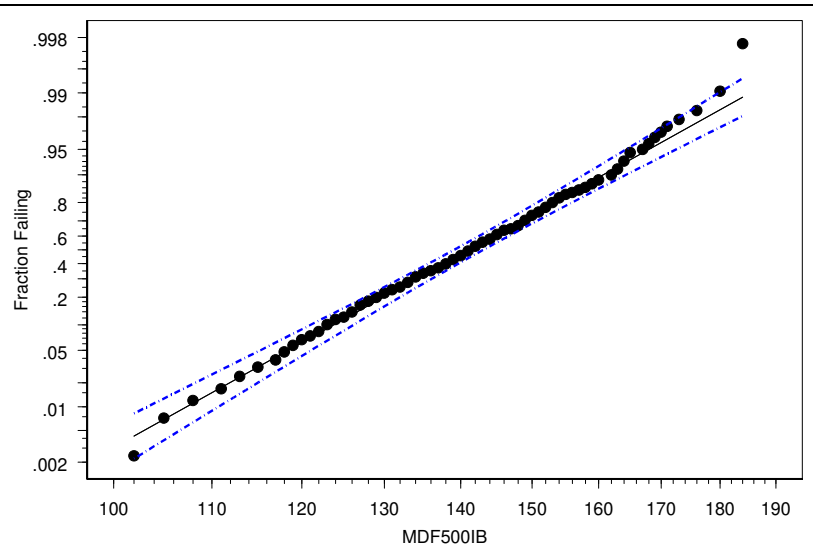


Figure 3.5. Log Logistic probability plot of the IB of 0.500" MDF.

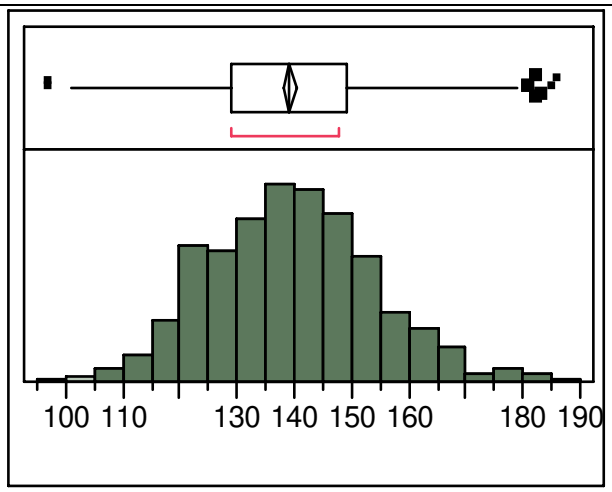


Figure 3.6. Box plot and histogram of the IB of 0.625" MDF.

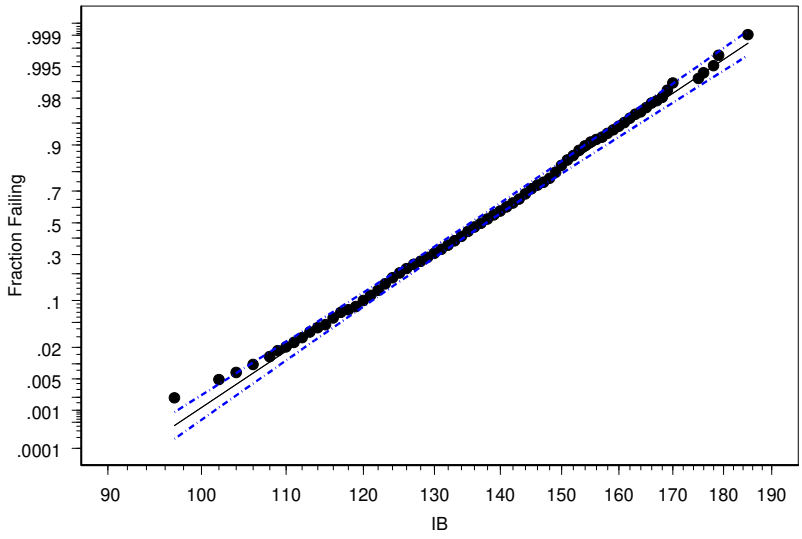


Figure 3.7. Normal probability plot of the IB of 0.625" MDF.

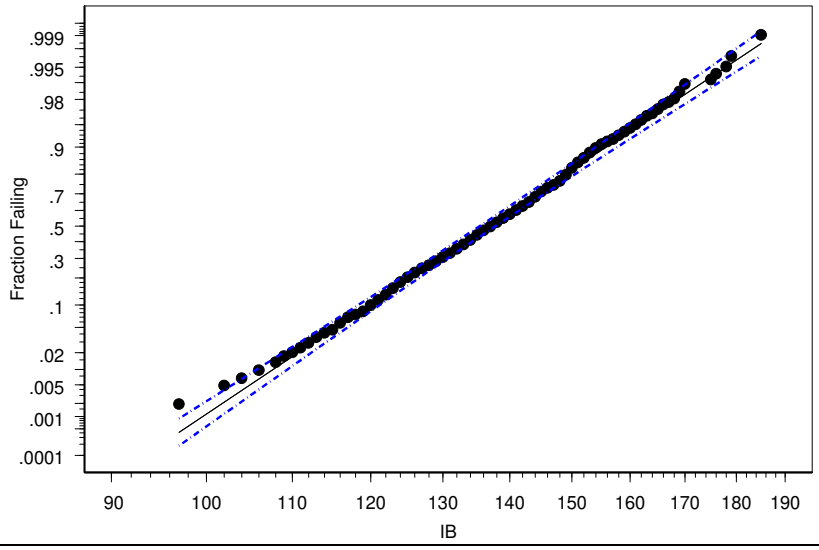


Figure 3.8. Log Normal probability plot of the IB of 0.625" MDF.

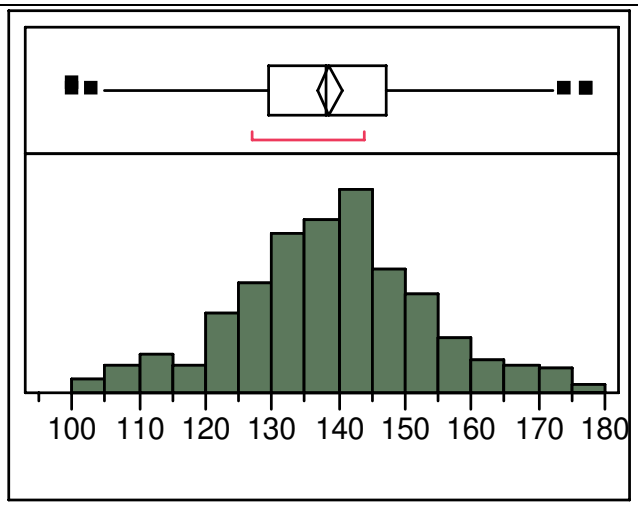


Figure 3.9. Box plot and histogram of the IB of 0.750" MDF.

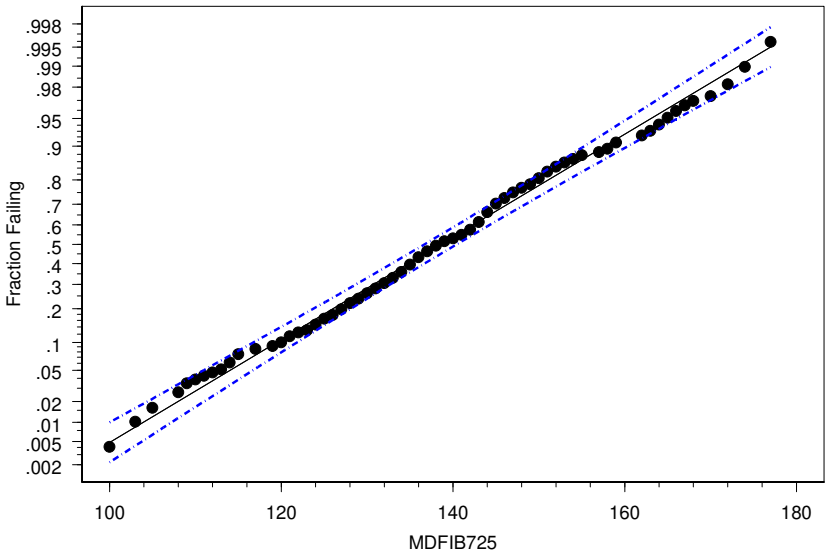


Figure 3.10. Normal probability plot of the IB of 0.750" MDF.

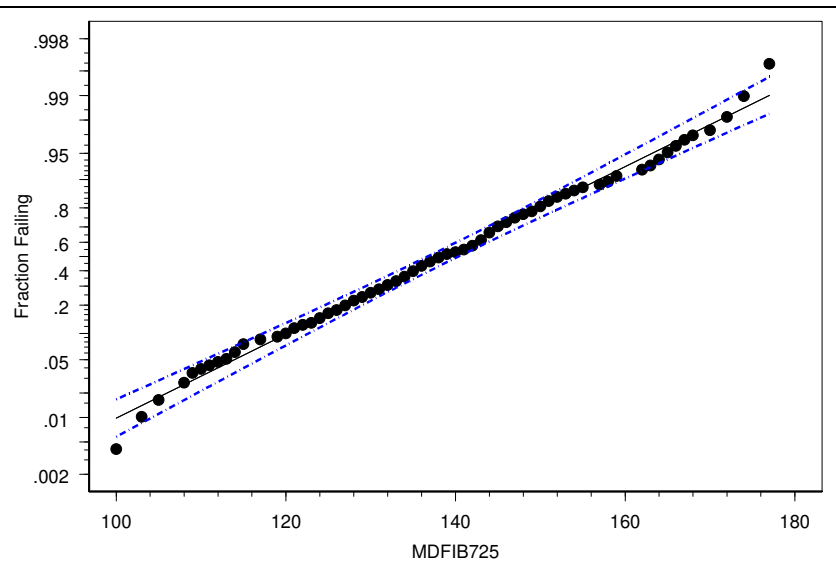


Figure 3.11. Logistic probability plot of the IB of 0.750" MDF.

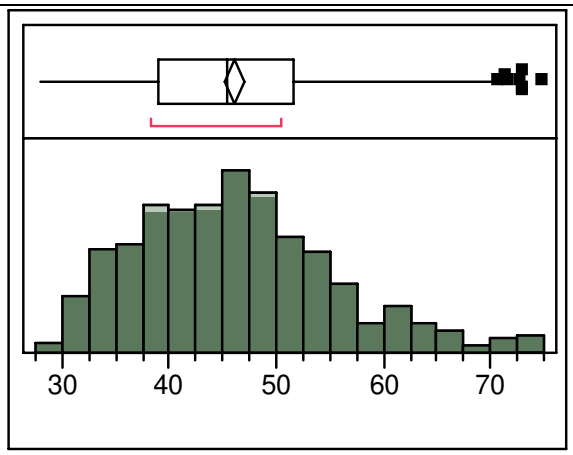


Figure 3.12. Box plot and histogram of the IB of OSB.

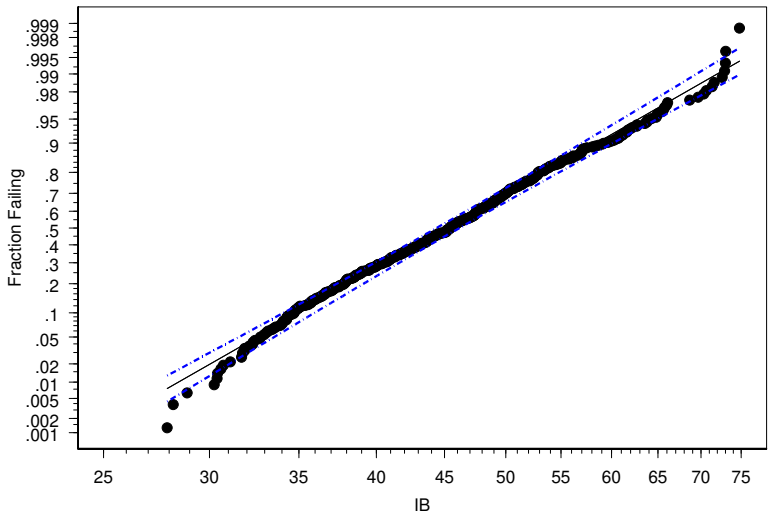


Figure 3.13. Log Normal probability plot of the IB of OSB.

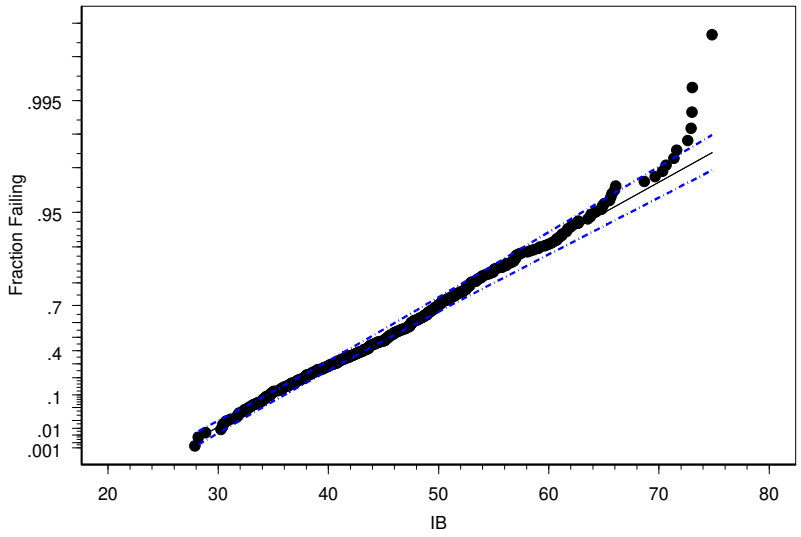


Figure 3.14. Largest Extreme Value probability plot of the IB of OSB.

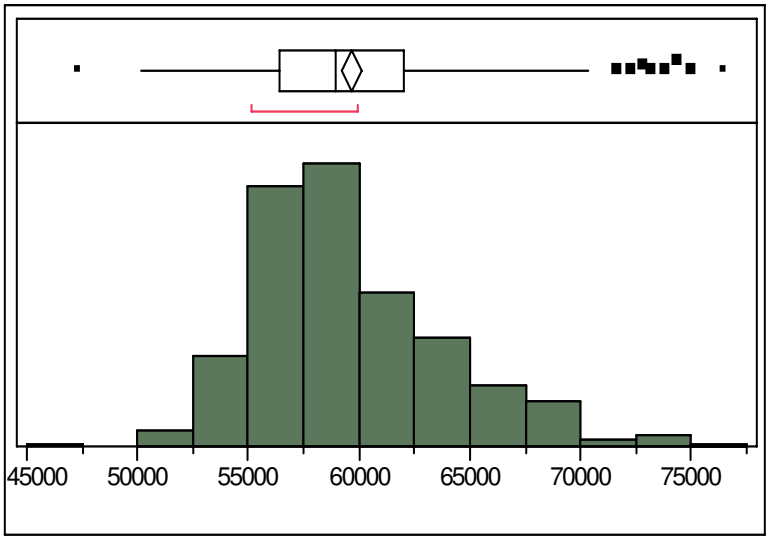


Figure 3.15. Box plot and histogram of the Parallel EI of OSB.

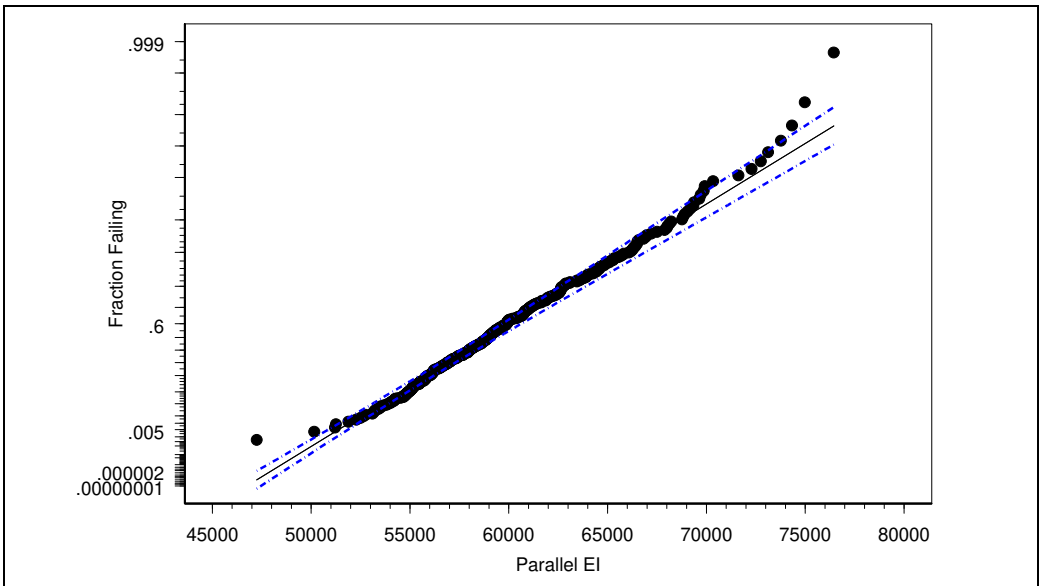


Figure 3.16. Largest Extreme Value probability plot of the Parallel EI of OSB.

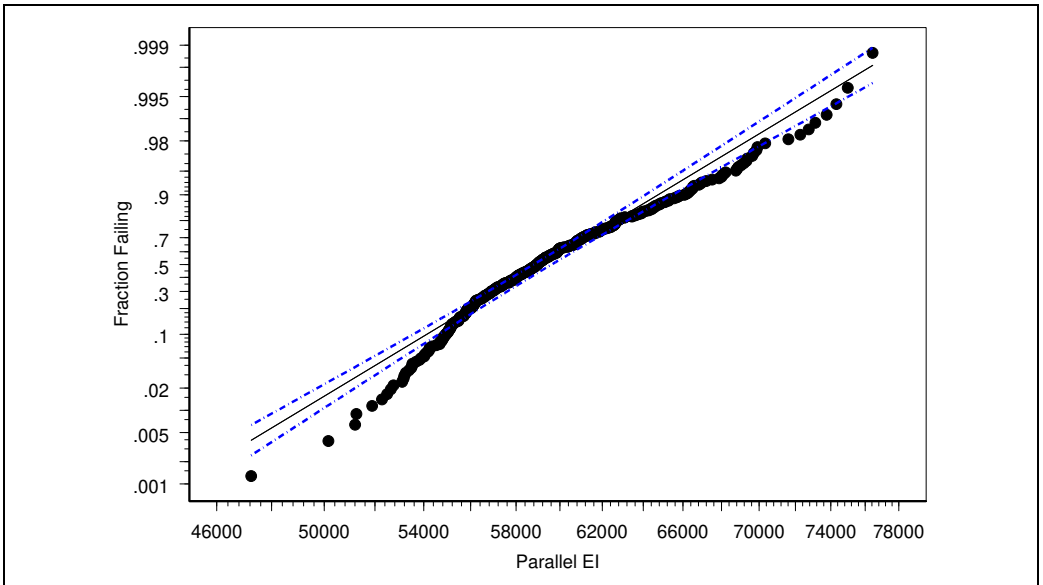


Figure 3.17. Log Logistic probability plot of the Parallel EI of OSB.

CHAPTER IV.

FIRST-, SECOND- AND THIRD-ORDER MULTIPLE LINEAR REGRESSION MODELS WITH INTERACTIONS OF MDF AND OSB STRENGTH PROPERTIES

Results are presented in this chapter for the first objective of the dissertation. Models of MDF and OSB strength properties using rudimentary regression methods are explored. The regression methods used in this chapter, even though fundamental, establish the foundation for the chapters to follow. The results of this chapter are presented in the spirit of notable statistical scholars as related to model building, i.e., less complex or rudimentary statistical model building methods should be investigated before proceeding with more complex statistical methods (Box and Cox 1964, Box 1979, Draper and Smith 1981, Deming 1986). This first objective of the dissertation will hopefully advance the philosophies of "All Possible Subsets" and "Best Model Criteria" for multiple linear regression modeling of industrial processes (Akaike 1974, Box 1979, Draper and Smith 1981, Myers 1990, Neter et al. 1996, Young and Guess 2002, and Kutner et al. 2004).

In the spirit of exploring less complex models initially; first-, second- and third-order multiple linear regression models with interaction terms are developed for the IB of MDF for three nominal products types (0.500", 0.625" and 0.750"). Models are also explored for the IB and Parallel EI of OSB for the product 7/16" RS. The Box Cox transforms of Y are further investigated for all regression models (Box and Cox 1964)

Forty regression models for MDF and 19 regression models for OSB are investigated in detail. Mixed stepwise regression with an all possible subsets criteria (discussed in Chapter II Methods) is used for initial model selection. Least squares methods are used for final model development (Plackett 1960). Models are developed for the entire record length for each product type of MDF and OSB. In addition, models are investigated for shorter record lengths. Investigating shorter record lengths builds upon the ideas presented by Bernardy and Scherff (1998, 1999), Erilsson et al. (2000), Young and Guess (2002), Young and Huber (2004) and Shaffer (2007) when investigating industrial data, i.e., the literature indicates that acceptable regression models of final product strength properties for engineered wood panels are possible for shorter record lengths. Modeling industrial processes using shorter record lengths addresses influences due to raw material change, product/setup change and tool/machinery-wear on final strength properties where these influences are transcended across longer record length models. For example, in MDF manufacture the refiner or defibrator plates that convert wood chips into fiber wear continuously through their life cycle which varies from seven to ten days unless a catastrophic event occurs, e.g., metal contaminate passes through defibrator plates (Suchsland and Woodson 1986, Maloney 1993, Young and Huber 2004). In OSB manufacture, the flaker knives that convert a log into wood strands are changed several times per week based on the number of cycle counts of the flaker or if catastrophic knife failure occurs (Woestheinrich and Meier 2001, Young and Huber 2004). Another significant challenge in modeling industrial processes is addressing influences on final strength properties from periodic setup changes that occur from changes in the customer order file.

To explore models with shorter record lengths, regression models are developed sequentially starting at a record length of 50 and ending at the total record length for each product type, e.g., 350 regression models are developed for 0.625" MDF for record lengths 50, 51, 52,….., 398, 399, 400. Record lengths for all models are contiguous and start at the most current record and progress backward in time for each record. Initial models with the highest adjusted R^2, lowest AIC, highest degrees of freedom and lowest RMSE are screened before investigating such models in more detail, e.g., plotting model residuals, RMSEP, plot of predicted values in the validation data set, etc. (Figures 4.1, 4.2 and 4.3). The analysis indicates that better MLR models result (using the best model criteria) from shorter record lengths of 60 to 70 records (approximately seven days) for the three MDF product types, and from 58 to 59 records (approximately 14 days) for OSB. These results are not contrary to the findings of previous research (Bernardy and Scherff 1998, 1999, Erilsson et al. 2000, Young and Guess 2002, Young and Huber 2004, and Shaffer 2007).

Medium Density Fiberboard

0.500" Thickness

An acceptable model for MDF 0.500" thickness (n=60) is a second-order model with 13 regressors (Tables 4.1 and 4.2). The model has an adjusted R^2 of 0.72, RMSE of 6.02 p.s.i. and homogeneous residual pattern (Figure 4.4). An attractive feature of the model is relatively accurate predictions of observed IB for extreme values in the validation data set (Figures 4.5 and 4.6). The regressors: "eVaprTemp" (boiler temperature) has a negative linear scaled estimate of -10.9 p.s.i.; "hPrAlTimeS" (press overall time set-point) has a positive linear scaled estimate of 12.6 p.s.i.; and "hPrCls2Tim" (press close two time)

has a negative polynomial scaled estimate of -14.9 p.s.i. on IB (Table 4.3).[15] This second-order model reveals a possible optimal IB of 138 p.s.i. for specific levels of the statistically significant ($\alpha < 0.02$) regressors in the model (Figures 4.7 and 4.8).

The best possible model for entire record length (n=175) for MDF 0.500" thickness is a first-order model with 17 regressors (Tables 4.1 and 4.4). The model has an adjusted R^2 of 0.69, RMSE of 8.33 p.s.i. and homogeneous residual pattern (Figure 4.9). The model did not approximate predictions of observed IB in the validation data set even though predicted IB follows the general time-ordered trend of IB (Figures 4.10 and 4.11). The regressors "hPrAlTimeS" (press overall time set-point) and "dCoreRsnS" (core resin set-point) have positive linear scaled estimates of 26.2 p.s.i. and 26.9 p.s.i., respectively (Table 4.5). The regressor "bFaceH202W" (face fiber water to wood ratio at refiners) has a negative linear scaled estimate of -15.1 p.s.i. on IB. This first-order model for all records reveals a possible optimal IB of 142 p.s.i for specific levels of the statistically significant ($\alpha < 0.02$) regressors (Figure 4.12).

0.625" Thickness

An acceptable model for MDF 0.625" thickness (n=62) is a first-order model with 11 regressors (Tables 4.1 and 4.6). The model has an adjusted R^2 of 0.78, RMSE of 7.15 p.s.i. and homogeneous residual pattern (Figure 4.13). Predicted IB approximated the time-trend and scale of observed IB in the validation data set (Figures 4.14 and 4.15). "gPreBBSpd" (pre-compressor bottom belt speed) has the strongest effect on IB with a negative linear scaled estimate of -32.6 p.s.i. (Table 4.7). An optimal IB of approximately

[15] Scaled estimate is the change in the dependent variable ("IB") when the regressor is moved over one-half of its range. Scaled estimate is a method of examining the relative influence of regressors on the dependent when the regressors have differing units of measure (scale).

138 p.s.i. is attainable for specific levels of the statistically significant regressors, $\alpha < 0.03$ (Figure 4.16). The response surface for "pre-compressor bottom belt speed" greater than 90 feet per minute and "cCO0046" (refiner steam pressure) greater than 35 results in an abrupt decline in IB which approaches the lower specification limit of 90 p.s.i. of the manufacturer (Figure 4.17).

The best possible model for entire record length (n=400) for MDF 0.625" thickness is a second-order model with 34 regressors (Tables 4.1 and 4.8). Even though this model has a large number of regressors it is not contrary to the literature on modeling industrial processes (Bernardy and Scherff 1998, 1999, Erilsson et al. 2000, Young and Guess 2002, Young and Huber 2004, and Shaffer 2007). The model has an adjusted R^2 of 0.60, RMSE of 8.80 p.s.i. and homogeneous residual pattern (Figure 4.18). Predicted IB does not approximate the time-trend and scale of observed IB in the validation data set (Figures 4.19 and 4.20). The regressor "hPrPPMTimS" (press pre-position move time set-point) has a negative polynomial scaled estimate of -17.7 p.s.i. on IB (Table 4.9). The regressor "cSwngChpL" (swing refiner wood chip level) has a positive linear scaled estimate of 20.6 p.s.i. (Table 4.9). This second-order model reveals an optimal IB of approximately 152 p.s.i at specific levels of the statistically significant ($\alpha < 0.10$) regressors (Figure 4.21). A negatively conical-shaped response surface exists for this model for "press pre-position move time set-point" and "core fiber humidifier temperature." This negatively conical-shaped response surface reveals a unique maximum IB of approximately 160 p.s.i. with other significant regressors held constant (Figure 4.22).

0.750" Thickness

An acceptable model for MDF 0.750" thickness (n=70) is a second-order model with 18 regressors (Tables 4.1 and 4.10). The model has an adjusted R^2 of 0.82, RMSE of 7.05 p.s.i. and mild heteroscedasticity in the residuals (Figure 4.23). Predicted IB does not approximate the time-trend and scale of observed IB in the validation data set (Figures 4.24 and 4.25). The regressor "dCoreDRSP" (core dust ratio set point) has a strong linear negative scaled estimate on IB of -29.4 p.s.i. (Table 4.11). "Core dust ratio set point" reflects the quantity of internal recycled dust waste from the sander and forming line trim saws that is added to the clean wood fiber generated at the defibrators. The regressor "dCoreScvWR" (core fiber scavenger resin content) has a negative polynomial scaled estimate on IB of -26.6 p.s.i. (Table 4.11). The regressor "aChipAugSp" (chip exit auger speed at the refiners) also has a negative linear scaled estimate on IB of -20.6 p.s.i. (Table 4.11). An optimal IB of approximately 129 p.s.i. is attainable at specific levels of the statistically significant regressors, $\alpha < 0.05$ (Figure 4.26). A negatively conical-shaped response surface exists for the regressors "fFaceMstm" (face fiber mat moisture content) and "core fiber scavenger resin content" which reveals a unique maximum of approximately 130 p.s.i. when other significant regressors are constant. The response surface between "hPrOpnTime" (press full open time) and "core fiber scavenger resin content" indicates a steep descent towards failing IB (i.e., less than the lower specification limit of 90 p.s.i.) for "core fiber scavenger resin content" levels greater than 5.5 percent and fast "press full open times" (Figure 4.27).

An acceptable multiple linear regression model is not possible for the MDF 0.750" thickness for the full record length of 200 (Tables 4.1 and 4.12). The best possible model

has an adjusted R^2 of only 0.42, RMSE of 11.27 p.s.i. and non-homogeneous residual pattern (Figure 4.28). Predicted IB did not approximate observed IB in the validation data set (Figures 4.29 and 4.30). The challenges of modeling strength properties of wood composites from industrial data using MLR methods are exemplified for the MDF 0.750" thickness and full record length.

Oriented Strand Board

Internal Bond – 7/16" RS

The strength properties (IB and Parallel EI) of OSB are a more challenging engineered wood panel to model using MLR methods and the given data set. Finding acceptable MLR models using the entire record length of 300 are difficult (Table 4.13). Figure 4.31 illustrates the poor performance for IB models (i.e., adjusted R^2, RMSE and AIC) for record lengths greater than 60 and Figure 4.32 illustrates the overall difficulty for any record length when modeling the Parallel EI of OSB. Recall from Chapter III that the distributions of IB and Parallel EI of OSB are non-normal (i.e., also recall the assumption of normality in the response variable required for multiple linear regression analysis). Box Cox transforms of *Y* (IB) tend to produce better MLR models for the IB of OSB. Box Cox transforms of Parallel EI did not improve the ability to develop acceptable MLR models for the Parallel EI of OSB. Two regression models for IB are discussed.

An acceptable model for OSB IB for a small record length (n=59) is a second-order model with the Box Cox transform of *Y,* where

$$Y^{\lambda} = \frac{Y^{1.8} - 1}{1.8\dot{Y}^{1.8-1}} = \frac{Y^{1.8} - 1}{38.7754}, \qquad [4.1]$$

and $\dot{Y}$ is defined as the geometric mean of IB. The model has 16 regressors (Tables 4.13 and 4.14). The model has an adjusted R^2 of 0.82, RMSE of 3.95 p.s.i. with some heteroscedasticity in the transformed residuals (Figure 4.33). Predicted IB approximates the time-trend and scale of observed IB in the 12 record validation data set (Figures 4.34 and 4.35). The regressor "Dry2Out" (dryer #2 outlet temperature) has a strong positive linear scaled estimate on IB of 15.3 p.s.i. (Table 4.15). "BnkSpdTCL" (bunker speed for top core layer) and "MD4OutTem" (main dryer #4 outlet temperature) have negative linear scaled estimates on IB of -11.2 and -10.6 p.s.i., respectively. The regressor "MSBCLOFDSP" (main spreader bottom core layer density set-point) also has a negative polynomial scaled estimate on IB of -10.1 p.s.i. The negative scaled estimates for bunker speed and spreader density set-points for the core layers may indicate that line speed is too fast for the machine capabilities of the wood strand forming system at the plant.

An optimal IB of approximately 48 p.s.i. is attainable at specific levels of the statistically significant regressors, $\alpha < 0.03$ (Figure 4.36). The second-order model reveals some useful relationships between the statistically significant regressors and IB (Figure 4.37). The three-dimensional polynomial graphs in Figure 4.37 illustrate that fast bunker speeds and low moisture content result in IB less than 50 p.s.i. The results of this regression model may indicate that the operational procedures at the plant for mat forming speed and moisture level need reevaluation.

The model for OSB IB for the entire record length (n=300) is presented to illustrate the problems associated with modeling the IB of OSB for longer record lengths using MLR (recall Figure 4.31). A second-order model with the Box Cox transform of Y yielded the best possible model. The Box Cox transform is:

$$Y^{\lambda} = \frac{Y^{0.2} - 1}{0.2\dot{Y}^{0.2-1}} = \frac{Y^{0.2} - 1}{0.0095}, \qquad [4.2]$$

where $\dot{Y}$ is defined as the geometric mean of IB. The model has 31 regressors (Tables 4.13 and 4.16). The model has a low adjusted R^2 of 0.45, RMSE of 6.70 p.s.i. and non-homogeneous residual pattern (Figure 4.38). Predicted IB has some approximation of the time-trend and scale of lower observed IB values in the 60 record validation data set but is generally unacceptable (Figures 4.39 and 4.40). The regressor "BnkSpdTCL" (bunker speed for top core layer) has a strong negative polynomial scaled estimate on IB of -17.7 p.s.i. (Table 4.17). "MSTSLOFSpA" (main spreader speed top surface layer) has a negative linear scaled estimate effect on IB of -14.2 p.s.i. The negative scaled estimates for bunker top core layer and top surface spreader speeds for the larger record length is in general agreement with the results of the shorter record length MLR model for OSB IB. Both the shorter and longer record length MLR models indicate that line speeds may be too fast for high strength properties given the machine capabilities of the mat forming system. Fast bunker speeds may result in poor mat forming which the literature also suggests has a negative influence on OSB strength properties (Suchsland and Woodson 1986, Maloney 1993, Kruse et al. 2000, Nishimura and Ansell 2002).

Any discussion of an optimal IB and three-dimensional polynomial graphs are inappropriate given the poor predictive capabilities of this longer record length MLR model. The results of the MLR models (shorter and longer record lengths) may have some explanatory value to the manufacturer in that mat forming procedures may need reevaluation given their negative influence on strength properties using the given data set.

Parallel EI – 7/16" RS

Parallel EI for OSB is a more difficult strength property to model using MLR relative to IB. This is illustrated by the results for shorter and longer record length MLR models in Table 4.13 and Figure 4.32. Box Cox transforms of *Y* did not improve model quality. Given the poor results of multiple linear regression models, one model is discussed for the Parallel EI of OSB.

The best possible model for OSB Parallel EI for a shorter record length (n=58) is a first-order model with 11 regressors (Tables 4.13 and 4.18). The model has an adjusted R^2 of 0.59, RMSE of 2,233 p.s.i. and homogeneous residual pattern (Figure 4.41). Predicted Parallel EI approximates some observed Parallel EI values in the validation data set (Figures 4.42 and 4.43). The regressor "MHOil2OilT" (main hot oil temperature for press) has a strong negative linear scaled estimate on Parallel EI of -5,891 in-lb^2/ft, i.e., as "main hot oil temperature for press" increases Parallel EI decreases (Table 4.19).

Discussions of an optimal IB and three-dimensional graphics are inappropriate given the poor predictive capabilities of this model. The inability to model Parallel EI using MLR may be indicative of non-homogenous data typical of OSB processes. However, the results of the first-order regression model for Parallel EI for the shorter record length may have some explanatory value, e.g., "main hot oil temperature for press" requires further root-cause investigation. Given that higher than necessary press oil temperatures result in lower Parallel EI strength (poor product safety and quality) and high oil temperatures require more energy, it may be advantageous for the manufacturer to investigate the effect of press hot oil temperatures and Parallel EI using a designed

experiment, i.e., the exploratory data analysis facilitates hypothesis generation (de Mast and Trip 2007).

Chapter IV Summary

Forty regression models for MDF and 19 regression models for OSB are investigated from a possible subset of 625 regression models for MDF and 500 possible regression models for OSB. First-, second- and third-order models with interaction terms and Box Cox transforms of *Y* are explored. Mixed stepwise regression with all possible subsets and a best model criteria (Young and Guess 2002) is used to develop a set of final candidate regression models. Candidate models are compared for the entire record length and smaller record lengths of the training data set.

Acceptable regression models are more feasible for MDF relative to OSB when using MLR methods for this given data set. MDF models for IB are more acceptable as thickness increases. First- and second-order models for MDF are more acceptable than higher-ordered models with interaction terms. Box Cox transforms of *Y* (IB) for MDF did not improve model quality or predictive capability. A surprising outcome of the MLR research is the lack of significance of interaction terms in most MLR models, i.e., MLR was unable to detect significant interactions for this data set.

The most appropriate models for MDF are a second-order model for 0.500" thickness for a small record length (n=60) and a first-order model for 0.625" thickness for a small record length (n=62). The MDF 0.500" thickness model has an adjusted R^2 of 0.72, RMSE of 6.02 p.s.i. and homogeneous residual pattern. An attractive feature of the model is relatively accurate predictions of observed IB for extreme values in the validation data set. Highly statistical significant ($\alpha < 0.0001$) regressors for this model are

"eVaprTemp" (boiler temperature), "hPrAlTimeS" (press overall time set-point) and "hPrCls2Tim" (press close two time).

The best possible model for entire record length (n=400) for MDF 0.625" thickness is a second-order model with 34 regressors. The model has an adjusted R^2 of 0.60, RMSE of 8.80 p.s.i. and homogeneous residual pattern. Predicted IB does not approximate the time-trend and scale of observed IB in the validation data set. Highly statistical significant ($\alpha < 0.0001$) regressors for this model are "hPrPPMTimS" (press pre-position move time set-point) and "cSwngChpL" (swing refiner wood chip level).

One acceptable model is developed for the shorter record length of IB for OSB. The best possible model for OSB IB for a small record length (n=59) is a second-order model with the Box Cox transform of *Y*, equation [4.1]. Recall from Chapter II that the distribution of OSB IB is non-normal. The model has 16 regressors with an adjusted R^2 of 0.82, RMSE of 3.95 p.s.i. with mild heteroscedasticity in the residuals. Predicted IB approximates the time-trend and scale of observed IB in the time-ordered validation data set. Highly statistically significant ($\alpha < 0.0001$) regressors for this model are "Dry2Out" (dryer #2 outlet temperature), "BnkSpdTCL" (bunker speed for top core layer), "MD4OutTem" (main dryer #4 outlet temperature) and "MSBCLOFDSP" (main spreader bottom core layer density set-point). Common among all statistically significant regressors for all models of OSB IB are process variables related to mat forming speed, e.g., regressors related to fast mat forming are negatively correlated with IB. This may suggest that mat forming/line speed procedures need reevaluation by the manufacturer. In general, MLR models for the IB and Parallel EI OSB may have some explanatory value to the manufacturer.

Appendix to Chapter IV

Table 4.1. Comparison of optimal record length with full record length for first-, second- and third-order stepwise regression models with interactions for MDF IB (shaded records are discussed).

MLR Model	R^2	n training	n validation	p	RMSE	RMSEP	RMSEP*
MDF 0.500": *Shorter Record Length*[16]							
First-order	0.40	60	13	3	9.10	15.68	71.96
Second-order	0.78	60	13	13	6.02	11.93	31.53
Third-order	0.79	60	13	17	6.13	16.54	16.89
Third-order interaction	0.68	60	13	11	7.12	16.48	16.26
N=175							
First-order	0.72	175	33	17	8.33	16.90	16.85
Second-order	0.75	175	33	26	8.01	33.05	33.14
Third order	*ns***	--	--	--	--	--	--
Second-order interaction	0.75	175	33	27	8.13	26.30	25.20
MDF 0.625": *Shorter Record Length*							
First-order	0.82	62	13	11	7.15	15.51	15.06
Second-order	*ns***	--	--	--	--	--	--
Third order	*ns***	--	--	--	--	--	--
First-order interaction	0.86	62	13	13	6.42	14.95	35.35
N=400							
First-order	0.56	400	80	27	9.61	14.88	74.51
Second-order	0.60	400	80	34	8.80	17.66	11.62
Third-order	0.57	400	80	33	9.57	20.01	30.98
Third-order interaction	0.62	400	80	39	8.98	17.00	16.60
MDF 0.750": *Shorter Record Length*							
First-order	0.88	70	14	17	6.55	50.40	35.77
Second-order	0.87	70	14	18	7.04	25.34	45.34
Third-order	*ns***	--	--	--	--	--	--
Second-order interaction	0.93	70	14	20	5.35	36.78	36.35
N=200							
First-order	0.40	200	40	13	11.81	24.53	38.67
Second-order	0.48	200	40	19	11.20	21.41	22.32
Third-order	0.46	200	40	19	11.45	23.70	26.36
Third-order interaction	0.51	200	40	30	11.27	22.22	35.95

*Using Box Cox transform that minimizes SSE in training model.

**"ns" indicates that all higher-order terms or interaction are statistically insignificant, $\alpha \leq 0.05$

[16] "Shorter Record Length" indicates a algorithm that is developed for this dissertation in SAS where starting at record length of the 50 most current records of the training dataset, one record at a time is added to 50 going backward in time up to 400 (entire training data set). The record length that has the highest adjusted R^2, lowest Akaike's Information criteria, lowest RMSE and largest degrees of freedom for regressors with p-values ≤ 0.05 is defined as the "Shorter Record Length."

Table 4.2. Summary of fit for second-order model, MDF 0.500", n=60.

RSquare	0.782368
RSquare Adj	0.720863
RMSE	6.023386
Mean of Response	134.3
Observations (or Sum Wgts)	60

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	13	5999.6658	461.513	12.7204
Error	46	1668.9342	36.281	Prob > F
C. Total	59	7668.6000		<.0001

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
eVaprTemp	-0.177238	0.036078	-4.91	<.0001	1.5212056
hPrAlTimeS	1.2555327	0.308288	4.07	0.0002	3.7318
(hPrCls2Tim)^2	-14.87093	3.8799	-3.83	0.0004	1.4382511
aChipSloLv	-0.23975	0.067	-3.58	0.0008	1.1975309
eM2236Spd	1.8986316	0.534033	3.56	0.0009	1.5764171
(eBoilrStmP)^2	-0.009198	0.002654	-3.47	0.0012	1.3268543
eBoilrH20F	1.7599991	0.531146	3.31	0.0018	2.9295179
(aChipAugSp)^2	0.0928276	0.028699	3.23	0.0023	1.4450588
fFaceMstM	9.3605094	3.136594	2.98	0.0045	1.6468404
dCoreTemp	0.4680389	0.177172	2.64	0.0112	1.6400362
eBoilrStmP	-0.16841	0.065804	-2.56	0.0138	1.7141641
hPrCls2Tim	-2.537921	2.144555	-1.18	0.2427	2.1211698
aChipAugSp	0.2098292	0.187562	1.12	0.2691	1.6473552

Table 4.3. Scaled estimates for the second-order model, MDF 0.500", n=60.

Term	Scaled Estimate	Plot Estimate	Prob>|t|
Intercept	138.02053		<.0001
eVaprTemp	-10.93825		<.0001
hPrAlTimeS	12.555327		0.0002
(hPrCls2Tim)^2	-14.87093		0.0004
aChipSloLv	-7.900344		0.0008
eM2236Spd	11.111741		0.0009
(eBoilrStmP)^2	-10.56352		0.0012
eBoilrH20F	8.0502358		0.0018
(aChipAugSp)^2	10.502659		0.0023
fFaceMstM	7.1155784		0.0045
dCoreTemp	6.6639379		0.0112
eBoilrStmP	-5.707314		0.0138
hPrCls2Tim	-2.537921		0.2427
aChipAugSp	2.2319116		0.2691

Table 4.4. Summary of fit for the first-order model, MDF 0.500", n=175.

RSquare	0.718264
RSquare Adj	0.687757
RMSE	8.328322
Mean of Response	142.1314
Observations (or Sum Wgts)	175

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	17	27762.308	1633.08	23.5446
Error	157	10889.669	69.36	**Prob > F**
C. Total	174	38651.977		<.0001

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
hPrAlTimeS	1.3079995	0.133949	9.76	<.0001	3.7434208
bFaceH202W	-4.820736	0.541971	-8.89	<.0001	3.1612959
dCoreRsnS	17.32292	2.357255	7.35	<.0001	8.1238064
Intercept	-1098.295	149.7139	-7.34	<.0001	.
dCoreEFCur	9.0144848	1.265357	7.12	<.0001	3.788578
bFaceTempP	1.1762998	0.19568	6.01	<.0001	2.0295956
cSwngChpL	1.2455158	0.276299	4.51	<.0001	2.2207622
bFaceBlwPs	0.780404	0.17566	4.44	<.0001	1.591995
hPrCls3Tim	-1.657122	0.397479	-4.17	<.0001	1.4072354
cCI0023PT	0.6445722	0.162684	3.96	0.0001	3.7545353
cSwgTemp	-0.127347	0.032247	-3.95	0.0001	2.3223254
cSwgOutlet	-0.003666	0.001105	-3.32	0.0011	3.5325045
hPrTempS	-13.70927	4.202486	-3.26	0.0014	2.2348617
fShavOffT2	-11.09113	3.431005	-3.23	0.0015	1.16896
bSwgDigPrs	-1.07664	0.364255	-2.96	0.0036	2.0515018
hPrTempP	0.5310372	0.206305	2.57	0.0110	1.2991691
fCoreBtmSpd	-0.276621	0.109367	-2.53	0.0124	1.6958934
cSwFbrMst	2.2413875	0.915535	2.45	0.0155	2.1603165

Table 4.5. Scaled estimates for the first-order model, MDF 0.500", n=175.

Term	Scaled Estimate	Plot Estimate	Prob>|t|
Intercept	142.13143		<.0001
hPrAlTimeS	26.159989		<.0001
bFaceH202W	-15.05853		<.0001
dCoreRsnS	26.850526		<.0001
dCoreEFCur	14.545322		<.0001
bFaceTempP	12.263514		<.0001
cSwngChpL	9.5191039		<.0001
bFaceBlwPs	7.0052577		<.0001
hPrCls3Tim	-10.50891		<.0001
cCI0023PT	10.115271		0.0001
cSwgTemp	-8.633337		0.0001
cSwgOutlet	-8.829031		0.0011
hPrTempS	-5.483709		0.0014
fShavOffT2	-5.719322		0.0015
bSwgDigPrs	-5.52494		0.0036
hPrTempP	5.1266327		0.0110
fCoreBtmSpd	-6.638905		0.0124
cSwFbrMst	5.2163699		0.0155

Table 4.6. Summary of fit for the first-order model for MDF 0.625", n=62.

RSquare	0.820207
RSquare Adj	0.780653
RMSE	7.152845
Mean of Response	137.8387
Observations (or Sum Wgts)	62

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**	
Model	11	11670.227	1060.93	20.7362	
Error	50	2558.160	51.16	**Prob > F**	
C. Total	61	14228.387		<.0001	
Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
gPreBBSpd	-2.762799	0.309655	-8.92	<.0001	1.8988451
cCO0046	-1.15042	0.170876	-6.73	<.0001	2.0540133
fShavOffT4	21.492844	3.612926	5.95	<.0001	1.3207313
fFaceBtmSd	1.8526461	0.32647	5.67	<.0001	2.5182647
eDesp2KV	5.4013529	1.010135	5.35	<.0001	2.3358217
bFaceResnW	11.888464	2.391513	4.97	<.0001	1.3685181
fFaceHTemp	0.7713985	0.183562	4.20	0.0001	1.5121018
Intercept	-386.8612	95.4818	-4.05	0.0002	.
eBoilrStmP	0.2565715	0.070923	3.62	0.0007	1.2809381
aFaceBinLv	0.7548149	0.262673	2.87	0.0059	1.6346084
dCoreDgstP	2.1168635	0.886009	2.39	0.0207	2.1329888
eDesp3KV	-1.818336	0.795924	-2.28	0.0266	1.8925804

Table 4.7. Scaled estimates for the first-order model for MDF 0.625", n=62.

Term	**Scaled Estimate**	**Plot Estimate**	**Prob>|t|**
Intercept	137.83871		<.0001
gPreBBSpd	-32.56305		<.0001
cCO0046	-15.53067		<.0001
fShavOffT4	12.969104		<.0001
fFaceBtmSd	16.673815		<.0001
eDesp2KV	15.884569		<.0001
bFaceResnW	11.067209		<.0001
fFaceHTemp	10.015838		0.0001
eBoilrStmP	10.037077		0.0007
aFaceBinLv	12.389003		0.0059
dCoreDgstP	9.6865557		0.0207
eDesp3KV	-6.068786		0.0266

Table 4.8. Summary of fit for the second-order model for MDF 0.625", n=400.

RSquare	0.633705
RSquare Adj	0.599584
Root Mean Square Error	8.804945
Mean of Response	138.74
Observations (or Sum Wgts)	400

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	34	48955.586	1439.87	18.5725
Error	365	28297.374	77.53	**Prob > F**
C. Total	399	77252.960		<.0001

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
(dCoreRsn2W) ²	-2.884208	0.745586	-3.87	0.0001	2.4949
(aShvRawWgt) ²	0.2360818	0.059078	4.00	<.0001	1.7308179
fFaceHTemp	0.5056718	0.110281	4.59	<.0001	2.8190149
cSwngGSF	0.005356	0.001134	4.72	<.0001	4.0698072
(dCoreRSSpd) ²	-0.007729	0.001633	-4.73	<.0001	1.47989
bFaceResnW	7.6020405	1.573282	4.83	<.0001	5.6197302
(fCoreHTmpT) ²	-0.043923	0.008833	-4.97	<.0001	1.2888891
fShaveOff1	-10.6519	2.058475	-5.17	<.0001	2.9677648
hPrPPHTimS	0.4146414	0.07813	5.31	<.0001	2.277471
fFaceHDP	-22.859	4.176386	-5.47	<.0001	3.810249
fFaceMstM	7.424371	1.282882	5.79	<.0001	2.0664754
cSwngChpL	0.9172788	0.152653	6.01	<.0001	3.0535694
Intercept	-192.8156	30.62812	-6.30	<.0001	.
(hPrPPMTimS) ²	-393.0008	49.07952	-8.01	<.0001	2.0281255
hPrPPMTimS	-73.05629	7.460664	-9.79	<.0001	2.9471743
(dCoreGrndP) ²	0.0051658	0.001367	3.78	0.0002	1.9470552
bFaceChpLv	-0.43902	0.118891	-3.69	0.0003	2.6179267
(bFacWxFlw) ²	-5.327495	1.501129	-3.55	0.0004	1.3707585
cSwOTemp	0.5184105	0.148662	3.49	0.0005	3.0764163
(dCoreTemp) ²	-0.035526	0.010608	-3.35	0.0009	1.3011849
(cCO0046) ²	-0.032509	0.009966	-3.26	0.0012	2.0437715
hPrCls3Tim	0.9748332	0.301569	3.23	0.0013	1.7977473
aShvRawWgt	-0.837664	0.265102	-3.16	0.0017	2.5084494
bFaceGrdSF	0.0026394	0.000834	3.16	0.0017	2.8795972
dCoreDgstP	0.3816478	0.133967	2.85	0.0046	2.3325764
(bFaceGrdSF) ²	-1.43e-6	5.379e-7	-2.66	0.0082	1.771709
(cSwngChpL)²	-0.019772	0.008023	-2.46	0.0142	1.4707526
gBlkDensty	2.5823849	1.212478	2.13	0.0339	1.7632221
cCO0046	-0.142804	0.08415	-1.70	0.0905	2.4181741
bFacWxFlw	1.8077344	1.322547	1.37	0.1725	2.0244267
dCoreTemp	-0.180852	0.135594	-1.33	0.1831	2.7771113
dCoreRsn2W	1.6574445	1.258765	1.32	0.1888	7.1452341
fCoreHTmpT	-0.101971	0.111246	-0.92	0.3599	2.4620296
dCoreGrndP	-0.052095	0.057937	-0.90	0.3692	4.4041632
dCoreRSSpd	0.0402315	0.04679	0.86	0.3904	2.8193236

Table 4.9. Scaled estimates for the second-order model for MDF 0.625", n=400.

Term	Scaled Estimate	Plot Estimate	Prob>|t|
Intercept	152.0052		<.0001
hPrPPMTimS	-15.52446		<.0001
(hPrPPMTimS)[2]	-17.74644		<.0001
cSwngChpL	20.587038		<.0001
fFaceMstM	9.8533654		<.0001
fFaceHDP	-16.00591		<.0001
hPrPPHTimS	8.6158952		<.0001
fShaveOff1	-12.1144		<.0001
(fCoreHTmpT)[2]	-17.41053		<.0001
bFaceResnW	11.48064		<.0001
(dCoreRSSpd)[2]	-10.80214		<.0001
cSwngGSF	9.4727133		<.0001
fFaceHTemp	8.2768612		<.0001
(aShvRawWgt)[2]	12.786978		<.0001
(dCoreRsn2W)[2]	-11.99363		0.0001
(dCoreGrndP)[2]	10.970469		0.0002
bFaceChpLv	-6.591924		0.0003
(bFacWxFlw)[2]	-11.60046		0.0004
cSwOTemp	9.065704		0.0005
(dCoreTemp)[2]	-9.112319		0.0009
(cCO0046)[2]	-6.835026		0.0012
hPrCls3Tim	6.3364159		0.0013
bFaceGrdSF	6.1360352		0.0017
aShvRawWgt	-6.164849		0.0017
dCoreDgstP	4.7175293		0.0046
(bFaceGrdSF)[2]	-7.726765		0.0082
(cSwngChpL)[2]	-9.959648		0.0142
gBlkDensty	3.9983969		0.0339
cCO0046	-2.070655		0.0905
bFacWxFlw	2.6675381		0.1725
dCoreTemp	-2.896438		0.1831
dCoreRsn2W	3.3798773		0.1888
fCoreHTmpT	-2.030186		0.3599
dCoreGrndP	-2.400699		0.3692
dCoreRSSpd	1.5040443		0.3904

Table 4.10. Summary of fit for the second-order model for MDF 0.750", n=70.

RSquare	0.866206
RSquare Adj	0.818985
RMSE	7.04955
Mean of Response	137.5429
Observations (or Sum Wgts)	70

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	18	16408.868	911.604	18.3435
Error	51	2534.504	49.696	**Prob > F**
C. Total	69	18943.371		<.0001

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
dCoreDRSP	-4.193256	0.488731	-8.58	<.0001	3.120832
aChipAugSp	-1.358512	0.194398	-6.99	<.0001	1.9185558
hPrClsTimP	1.0052124	0.165454	6.08	<.0001	4.4569595
eEI1071FT	444.82459	74.08204	6.00	<.0001	6.9689426
(cSwgWxFbrW)^2	0.0046517	0.000839	5.55	<.0001	1.8011744
(dCoreScvWR)^2	-331761.1	61638.22	-5.38	<.0001	2.8948707
eM2236Spd	2.1244118	0.442149	4.80	<.0001	2.4077811
eDespMamp	-0.377945	0.085829	-4.40	<.0001	1.8428637
dCoreGrndS	0.279431	0.068211	4.10	0.0002	5.5394067
eDFld3Mamp	0.0957255	0.025881	3.70	0.0005	6.772589
hPrOpnTime	-0.508435	0.149654	-3.40	0.0013	1.4694515
Intercept	-12247.58	3848.063	-3.18	0.0025	.
(eDFld3Mamp)^2	0.0010352	0.000326	3.17	0.0026	4.1411091
dCoreScvWR	1996.0091	643.7091	3.10	0.0031	2.6738128
cSwgWxFbrW	-0.106905	0.040723	-2.63	0.0114	2.2176688
eM2241Spd	0.210673	0.081191	2.59	0.0123	1.6239658
(fFaceMstM)^2	-17.40546	7.323693	-2.38	0.0213	1.5580044
bFacResnWS	8.5379183	4.081285	2.09	0.0414	4.0369594
fFaceMstM	5.2222213	3.820829	1.37	0.1777	2.0559436

Table 4.11. Scaled estimates for the second-order model for MDF 0.750", n=70.

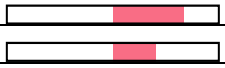
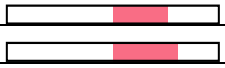
Term	Scaled Estimate	Plot Estimate	Prob>|t|
Intercept	128.95612		<.0001
dCoreDRSP	-29.35279		<.0001
aChipAugSp	-20.58037		<.0001
hPrClsTimP	22.805757		<.0001
eEI1071FT	22.381794		<.0001
$(cSwgWxFbrW)^2$	26.990087		<.0001
$(dCoreScvWR)^2$	-26.63432		<.0001
eM2236Spd	13.226269		<.0001
eDespMamp	-12.54792		<.0001
dCoreGrndS	19.898003		0.0002
eDFld3Mamp	11.715886		0.0005
hPrOpnTime	-7.781167		0.0013
$(eDFld3Mamp)^2$	15.506291		0.0026
dCoreScvWR	17.884241		0.0031
cSwgWxFbrW	-8.143196		0.0114
eM2241Spd	7.4345014		0.0123
$(fFaceMstM)^2$	-11.70343		0.0213
bFacResnWS	7.6841264		0.0414
fFaceMstM	4.2822215		0.1777

Table 4.12. Summary of fit for the second-order model with interaction terms, MDF 0.750", n=200.

RSquare	0.505768
RSquare Adj	0.418035
RMSE	11.27371
Mean of Response	139
Observations (or Sum Wgts)	200

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	30	21980.695	732.690	5.7648
Error	169	21479.305	127.096	**Prob > F**
C. Total	199	43460.000		<.0001

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
dCoreDRSP	-2.679268	0.541799	-4.95	<.0001	3.6151547
(hPosTime)^2	-0.0868	0.01899	-4.57	<.0001	2.7062869
(cSwDgstrL)*(fFaceHumid)	-0.473084	0.10574	-4.47	<.0001	1.425036
eM2237Spd	1.4618715	0.34692	4.21	<.0001	2.3819806
hPosTime	1.2161295	0.290269	4.19	<.0001	4.5833167
(aChipAugSp)^3	-0.007036	0.00188	-3.74	0.0002	4.3746127
dCoreTemp	0.7787551	0.21668	3.59	0.0004	2.3038131
fFaceMstM	-8.102769	2.312433	-3.50	0.0006	1.887095
cCI0023PT	0.441632	0.137226	3.22	0.0015	2.7396871
eBoilrStmF	-0.000725	0.000228	-3.18	0.0018	1.6058179
(hPrAlTimeS)^2	0.0158995	0.005161	3.08	0.0024	2.4529553
(dCoreH202W)^2	-18064.02	6498.356	-2.78	0.0061	1.8294963
cSwDgstrL	0.3675405	0.1342	2.74	0.0068	2.2961775
bFaceChpLv	0.7709831	0.296203	2.60	0.0101	2.7347898
dCoreWx	6.5576589	2.633773	2.49	0.0137	1.9611066
(dCoreDRSP)^2	0.2024948	0.100035	2.02	0.0445	2.6518824
dCoreH202W	-278.5607	179.3951	-1.55	0.1223	1.8647891
Intercept	12367.099	7999.038	1.55	0.1240	.
fFaceHumid	1.9617084	1.305062	1.50	0.1347	1.555024
(aCoreBinLv)*(fFaceHumid)	-0.244344	0.16361	-1.49	0.1372	1.3924855
(cSwgScv2W)*(fFaceHumid)	-917.4004	642.7628	-1.43	0.1553	1.6063435
bFaceGrdSF	0.0016839	0.001257	1.34	0.1821	1.4667581
eDespMamp	-0.093732	0.074288	-1.26	0.2088	2.0984209
aChipAugSp	0.2359341	0.276238	0.85	0.3943	4.9187954
(aCoreBinLv)*(cSwgTemp)	0.0035755	0.005454	0.66	0.5130	1.2229218
hPrAlTimeS	0.0643628	0.117366	0.55	0.5841	3.8740363
cSwgScv2W	-226.6042	439.4665	-0.52	0.6068	1.2430773
cSwgTemp	-0.014433	0.041891	-0.34	0.7309	1.6033888
(aChipAugSp)^2	-0.005469	0.018031	-0.30	0.7620	1.5456402
(eDespMamp)^2	-0.000426	0.002938	-0.15	0.8848	1.4204288
aCoreBinLv	0.0102356	0.1213	0.08	0.9329	1.620821

Table 4.13. Comparison of optimal record length with full record length for first, second- and third-order stepwise regression models with interactions for OSB IB and Parallel EI (shaded records are discussed).

MLR Model	R^2	n training	n validation	p	RMSE	RMSEP	RMSEP*
Internal Bond: *Shorter Record Length*[17]							
First-order	0.81	59	12	12	4.57	7.60	7.53
Second-order	0.87	59	12	16	3.95	5.05	4.67
Third-order	*ns***	--	--	--	--	--	--
Second-order interaction	*ns***	--	--	--	--	--	--
N=300							
First-order	0.42	300	60	23	7.14	9.64	8.95
Second-order	0.51	300	60	31	6.70	12.74	9.67
Third-order	0.41	300	60	25	7.26	12.67	12.26
Third-order interaction	*ns***	--	--	--	--	--	--
Parallel EI: *Shorter Record Length*							
First-order	0.67	58	16	11	2233	5901	--***
Second-order	0.32	58	16	6	3026	4371	--***
Third-order	0.77	58	16	14	1934	6724	--***
Third-order interaction	0.83	58	16	14	1632	4446	4345
N=300							
First-order	0.22	300	60	8	4141	4082	--***
Second-order	0.42	300	60	27	3708	4706	--***
Third-order	0.47	300	60	33	3583	5500	--***
Third-order interaction	0.42	300	60	27	3700	5860	--***

*Using Box Cox transform that minimizes SSE in training model.

**"ns" indicates that all higher-order or interaction terms are statistically insignificant, $\alpha \leq 0.05$

***No regressors statistically significant ($\alpha < .10$) with Box Cox transform.

[17] "Shorter Record Length" indicates a algorithm that is developed for this dissertation in SAS where starting at record length of the 50 most current records of the training dataset, one record at a time is added to 50 going backward in time up to 400 (entire training data set). The record length that has the highest adjusted R^2, lowest Akaike's Information criteria and largest degrees of freedom for regressors with p-values ≤ 0.05 is defined as the "Shorter Record Length."

Table 4.14. Summary of fit for the second-order model with Box Cox transform, OSB IB, n=59.

RSquare	0.868944
RSquare Adj	0.819018
Root Mean Square Error	3.9463
Mean of Response	47.273
Observations (or Sum Wgts)	59

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	16	4336.7443	271.047	17.4046
Error	42	654.0779	15.573	**Prob > F**
C. Total	58	4990.8222		<.0001

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
Dry2Out	0.2696975	0.06007	4.49	<.0001	3.8977687
PKI700QOTi	-1.085833	0.230843	-4.70	<.0001	1.1644498
MTCLMoiLev	4.1187788	0.866295	4.75	<.0001	1.4902689
Dr3OutMois	10.618419	2.172934	4.89	<.0001	2.9474968
(MSBCLOFDSP)^2	-7.958358	1.628046	-4.89	<.0001	1.6827701
MSBCLOFDSP	7.793946	1.465713	5.32	<.0001	2.5681587
MD4OutTem	-0.136677	0.023966	-5.70	<.0001	1.5609997
BnkSpdTCL	-1.11911	0.195796	-5.72	<.0001	1.9450153
Intercept	-344.7253	54.49197	-6.33	<.0001	.
PZI701Ste5	0.2558758	0.032004	8.00	<.0001	1.5380048
(Dry2Out)^2	0.002579	0.000849	3.04	0.0041	2.8050309
MHOil2OilT	0.2185511	0.0744	2.94	0.0054	1.9082734
(DyBiLBoTSL)^2	-0.003644	0.001255	-2.90	0.0058	1.930353
(MD1OutTem)^2	0.0020606	0.000794	2.60	0.0130	3.012923
(MTCLMoiLev)^2	1.4676874	0.639984	2.29	0.0269	1.1929714
MD1OutTem	0.0640726	0.047577	1.35	0.1853	3.008842
DyBiLBoTSL	0.0121548	0.054203	0.22	0.8237	2.2017116

Table 4.15. Scaled estimates for the second-order model with Box Cox transform, OSB IB, n=59.

Term	Scaled Estimate	Plot Estimate	Prob>|t|
(MSBCLOFDSP)[2]	-10.09921		<.0001
BnkSpdTCL	-11.1911		<.0001
Dr3OutMois	8.683211		<.0001
Dry2Out	15.305335		<.0001
Intercept	48.277364		<.0001
MD4OutTem	-10.59246		<.0001
MSBCLOFDSP	8.7798951		<.0001
MTCLMoiLev	8.3410961		<.0001
PKI700QOTi	-10.31541		<.0001
PZI701Ste5	13.817295		<.0001
(Dry2Out)[2]	8.3059136		0.0041
MHOil2OilT	6.7750838		0.0054
(DyBiLBoTSL)[2]	-8.22131		0.0058
(MD1OutTem)[2]	8.3742336		0.0130
(MTCLMoiLev)[2]	6.0192566		0.0269
MD1OutTem	4.0846307		0.1853
DyBiLBoTSL	0.5773526		0.8237

Table 4.16. Summary of fit for the second-order model with Box Cox transform, OSB IB, n=300.

RSquare	0.507502
RSquare Adj	0.450534
RMSE	6.700853
Mean of Response	46.10673
Observations (or Sum Wgts)	300

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	31	12400.187	400.006	8.9085
Error	268	12033.583	44.901	**Prob > F**
C. Total	300	24433.770		<.0001

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
PZI701Ste4	0.1454053	0.023306	6.24	<.0001	1.9433827
MBCRFRFB	29.266959	5.105569	5.73	<.0001	1.9484868
(BnkSpdTCL)^2	-0.026281	0.005682	-4.62	<.0001	4.4203533
MDBCLev	0.1332334	0.029632	4.50	<.0001	1.7135019
PLI795	-1.200244	0.269747	-4.45	<.0001	1.6807351
PKI700JPTi	0.1635721	0.043072	3.80	0.0002	4.6007777
DryWeBin5	0.9390069	0.24899	3.77	0.0002	7.6965839
(Dry3In) ^{2}	-2.131e-5	6.02e-6	-3.54	0.0005	8.0090149
(PLI795) ^{2}	-0.211566	0.061922	-3.42	0.0007	1.3924604
MSTSLOFSpA	-0.372706	0.110389	-3.38	0.0008	7.9630801
Dry4Out	-0.040535	0.01223	-3.31	0.0010	1.819225
(PHK71DP3Ti) ^{2}	-0.005221	0.001607	-3.25	0.0013	9.4723208
(DryWeBin5) ^{2}	0.0977245	0.03019	3.24	0.0014	8.6916278
PHK71DP3Ti	0.2581961	0.080939	3.19	0.0016	8.4073128
DryWeBin2	-0.75231	0.24009	-3.13	0.0019	8.5310591
Dry3In	-0.009475	0.003034	-3.12	0.0020	6.1050943
MFLTerWegt	-2.235229	0.72958	-3.06	0.0024	2.2760185
PTI770	-0.064417	0.021777	-2.96	0.0034	2.1184321
BnkSpdTCL	-0.544102	0.184877	-2.94	0.0035	9.0886904
MDBBSLLev	-0.079148	0.027698	-2.86	0.0046	2.0619137
MPTemp	0.2659322	0.104745	2.54	0.0117	2.0972742
(PrHK700C3R) ^{2}	-0.145795	0.05817	-2.51	0.0128	1.7591526
PZI701St13	-0.067193	0.029735	-2.26	0.0246	4.2815783
WeBi5Tot24	0.0045613	0.00202	2.26	0.0247	1.8583371
MD3OutTem	-0.045353	0.020082	-2.26	0.0247	2.6560252
Intercept	-138.1027	61.76715	-2.24	0.0262	.
MD5OutTem	0.0464253	0.021471	2.16	0.0315	1.7696192
(DryWeBin2) ^{2}	-0.056669	0.027626	-2.05	0.0412	9.0713671
PMI731	0.1186896	0.058277	2.04	0.0427	1.1863373
PTI796C	0.1610469	0.09236	1.74	0.0824	2.5751869
PQI700	0.0149964	0.00886	1.69	0.0917	1.147832
PrHK700C3R	-0.034945	0.253752	-0.14	0.8906	2.1083909

Table 4.17. Scaled estimates for the second-order model with Box Cox transform, OSB IB, n=300.

Term	Scaled Estimate	Plot Estimate	Prob>|t|
PZI701Ste4	13.449989		<.0001
MBCRFRFB	8.7800862		<.0001
(BnkSpdTCL)[2]	-17.76607		<.0001
MDBCLev	3.3308358		<.0001
PLI795	-8.101649		<.0001
PKI700JPTi	13.003986		0.0002
DryWeBin5	8.9205655		0.0002
(Dry3In-958.23)[2]	-8.176517		0.0005
(PLI795)[2]	-9.63946		0.0007
MSTSLOFSpA	-14.19673		0.0008
Dry4Out	-4.296684		0.0010
(PHK71DP3Ti)[2]	-6.395289		0.0013
(DryWeBin5)[2]	8.8196317		0.0014
PHK71DP3Ti	9.0368637		0.0016
DryWeBin2	-7.146946		0.0019
Dry3In	-5.869903		0.0020
MFLTerWegt	-4.482991		0.0024
PTI770	-3.778121		0.0034
BnkSpdTCL	-14.14665		0.0035
MDBBSLLev	-2.968064		0.0046
MPTemp	5.0527126		0.0117
(PrHK700C3R)[2]	-6.159825		0.0128
PZI701St13	-5.946544		0.0246
WeBi5Tot24	3.2818738		0.0247
MD3OutTem	-4.864141		0.0247
MD5OutTem	4.503253		0.0315
(DryWeBin2)[2]	-5.114356		0.0412
PMI731	4.3024978		0.0427
PTI796C	3.6235546		0.0824
PQI700	1.4546554		0.0917
PrHK700C3R	-0.227145		0.8906
PZI701Ste4	13.449989		<.0001

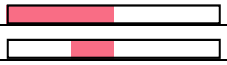
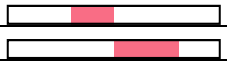
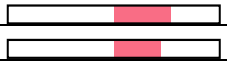
Table 4.18. Summary of fit for the first-order model, OSB Parallel EI, n=58.

RSquare	0.665715
RSquare Adj	0.585778
RMSE	2232.722
Mean of Response	58932.08
Observations (or Sum Wgts)	58

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	11	456666304	41515119	8.3279
Error	46	229312177	4985047.3	**Prob > F**
C. Total	57	685978481		<.0001

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
PMI733	-351.8691	57.58996	-6.11	<.0001	1.350916
MHOil2OilT	-190.0351	34.63287	-5.49	<.0001	1.2914359
Dr2OutMois	2319.4085	481.655	4.82	<.0001	1.3194214
PPI740	1469.5119	375.3968	3.91	0.0003	1.9847096
WeBi5Tot24	6.5940582	1.902065	3.47	0.0012	1.1809942
MSTCLLev	-219.564	67.31155	-3.26	0.0021	1.3454974
PTI700	-289.2639	94.62459	-3.06	0.0037	1.2149607
PZI701Ste2	230.67719	95.24907	2.42	0.0194	1.1874661
PMI748	0.8671941	0.366255	2.37	0.0222	1.7667314
Intercept	-777567.2	336454.4	-2.31	0.0254	.
Dry4Out	24.05691	11.44027	2.10	0.0410	1.2086515
MF1HMIWatd	130.61953	62.98994	2.07	0.0437	1.2524047

Table 4.19. Scaled estimates for the first-order model, OSB Parallel EI, n=58.

Term	**Scaled Estimate**	**Plot Estimate**	**Prob>|t|**
Intercept	58932.082		<.0001
MHOil2OilT	-5891.087		<.0001
PMI733	-2234.369		<.0001
Dr2OutMois	3649.9787		<.0001
PPI740	3159.4506		0.0003
WeBi5Tot24	2667.2965		0.0012
MSTCLLev	-3293.459		0.0021
PTI700	-2053.774		0.0037
PZI701Ste2	1499.4017		0.0194
PMI748	1682.7902		0.0222
Dry4Out	1876.439		0.0410
MF1HMIWatd	1240.8855		0.0437

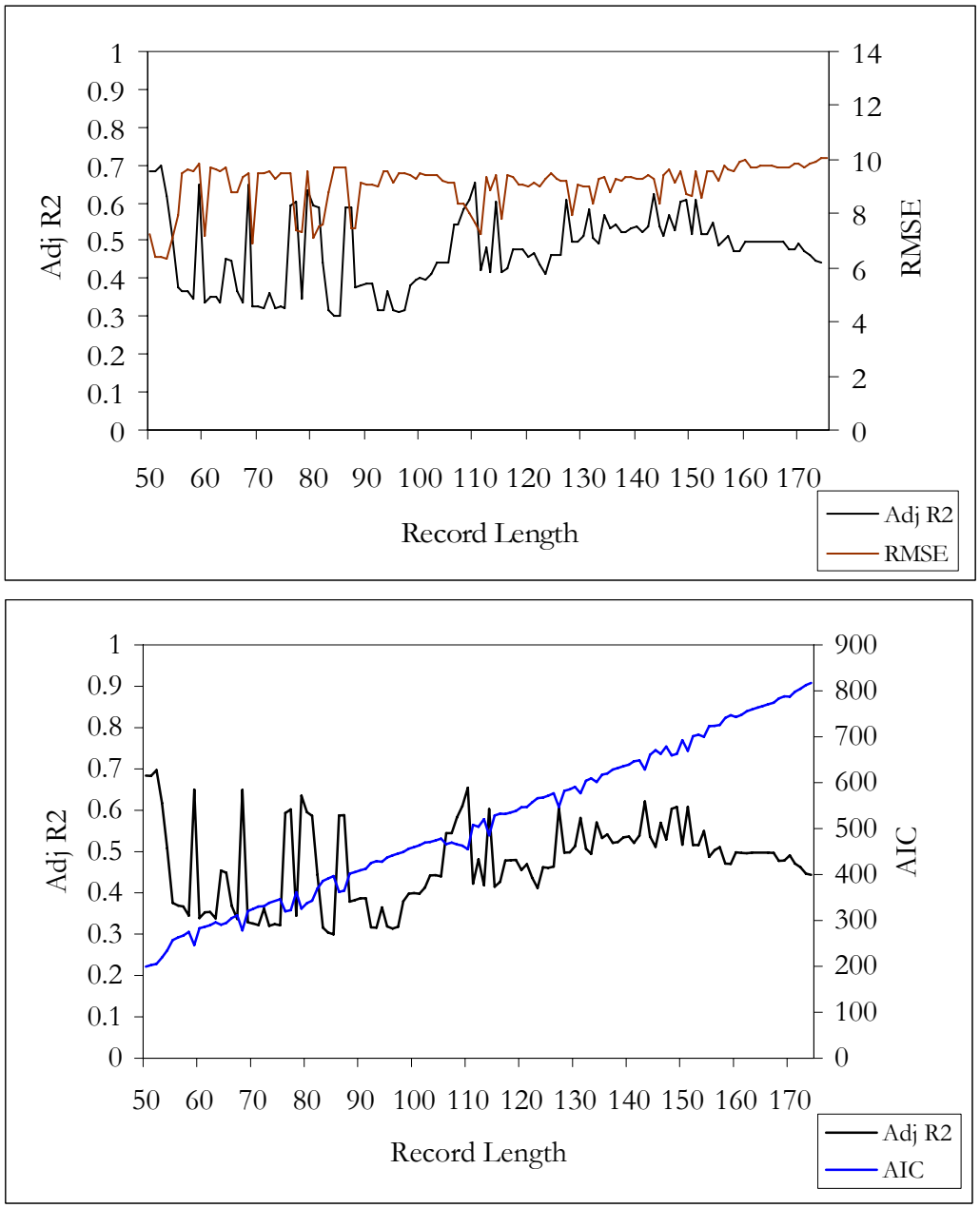
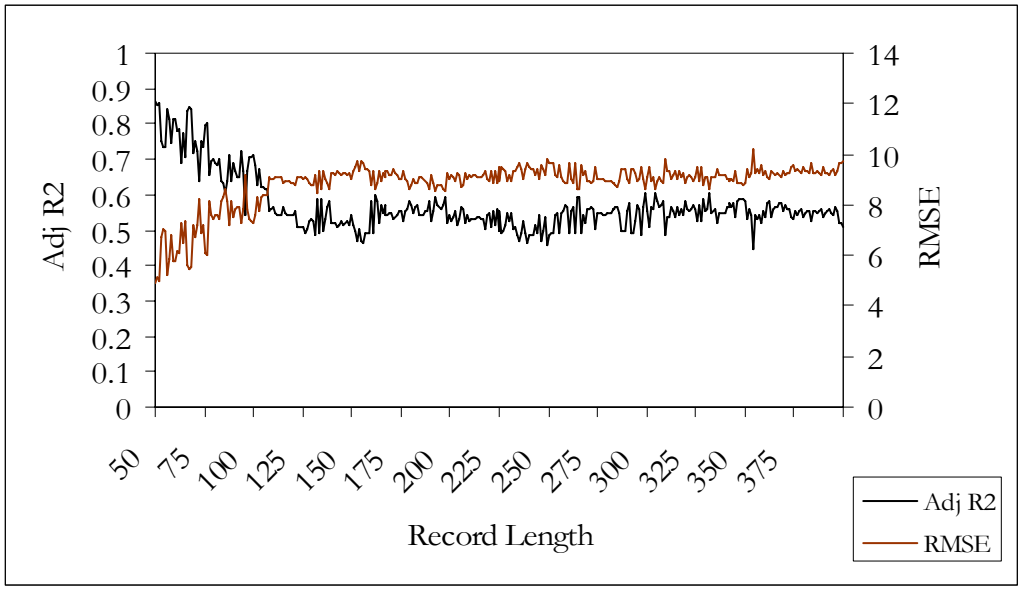


Figure 4.1. Comparison of Adjusted R^2 and AIC for first-order stepwise regression models for MDF 0.500" for every record length greater than 50.



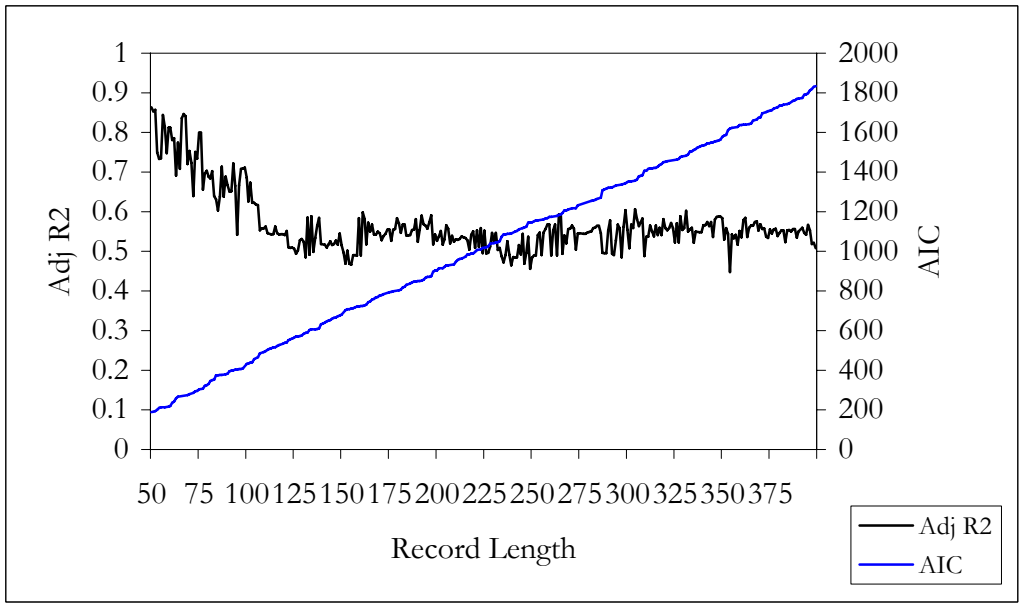


Figure 4.2. Comparison of Adjusted R^2 and AIC for first-order stepwise regression models for MDF 0.625" for every record length greater than 50.

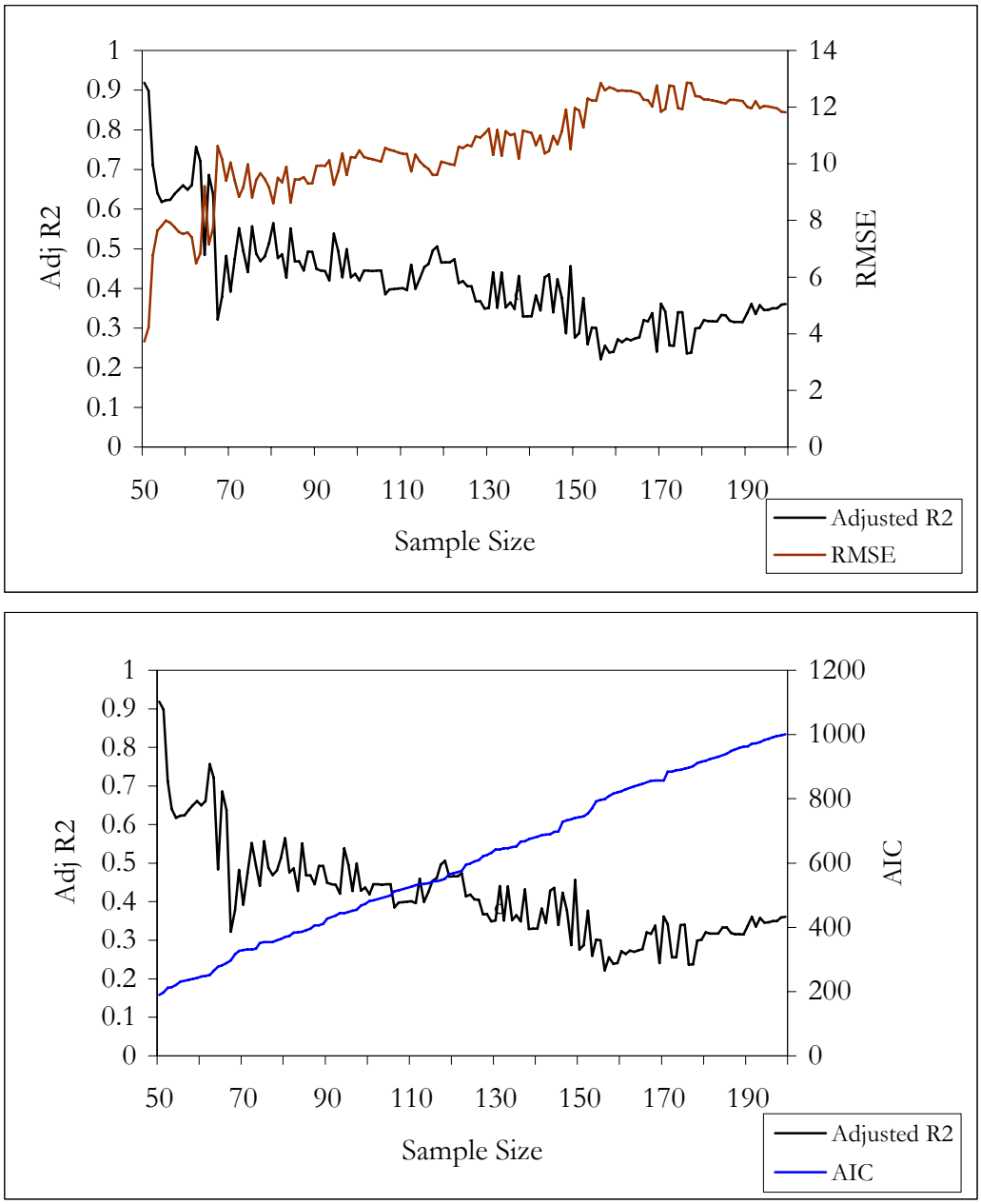


Figure 4.3. Comparison of Adjusted R^2 and AIC for first-order stepwise regression models for MDF 0.750" for every record length greater than 50.

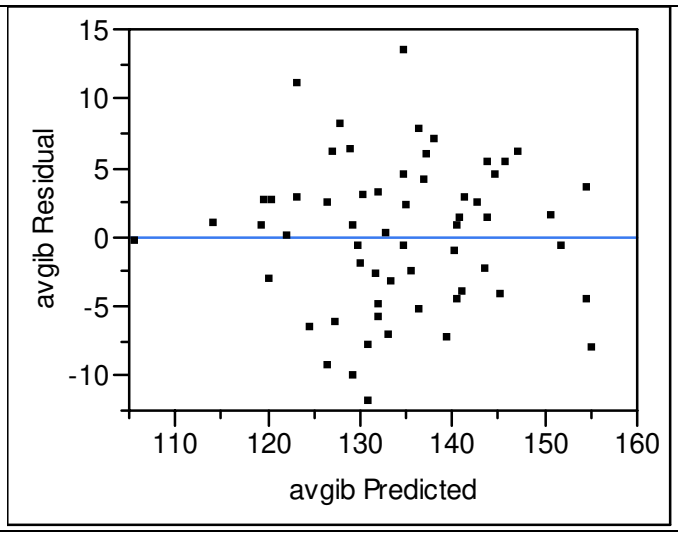


Figure 4.4. Residuals by predicted IB plot for the second-order model, MDF 0.500", n=60.

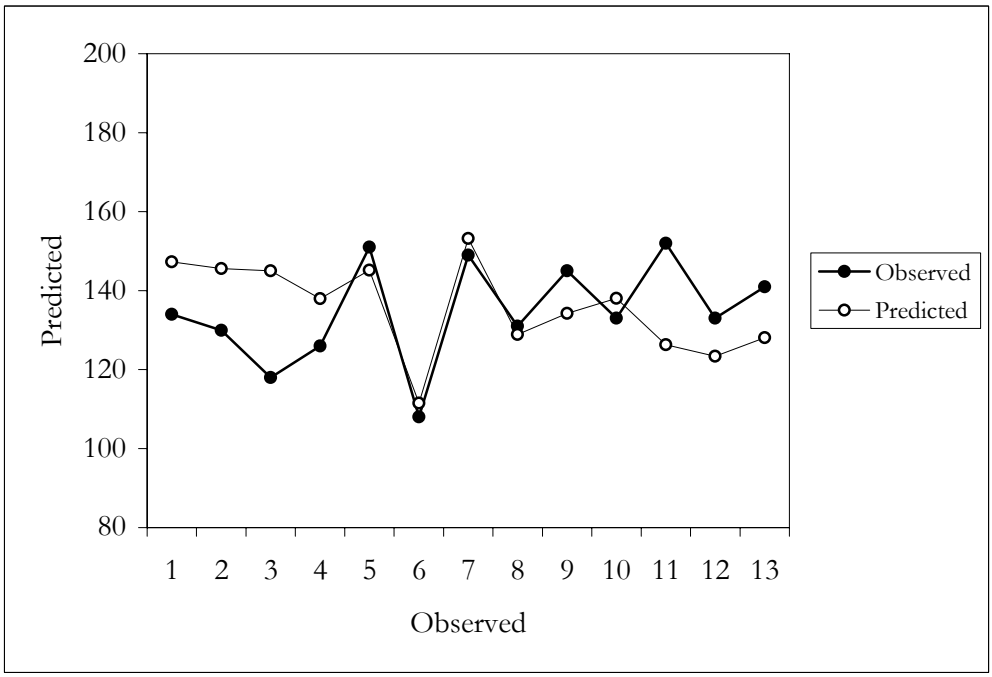


Figure 4.5. Time series graph of validation data set for the second-order model, MDF 0.500", n=15.

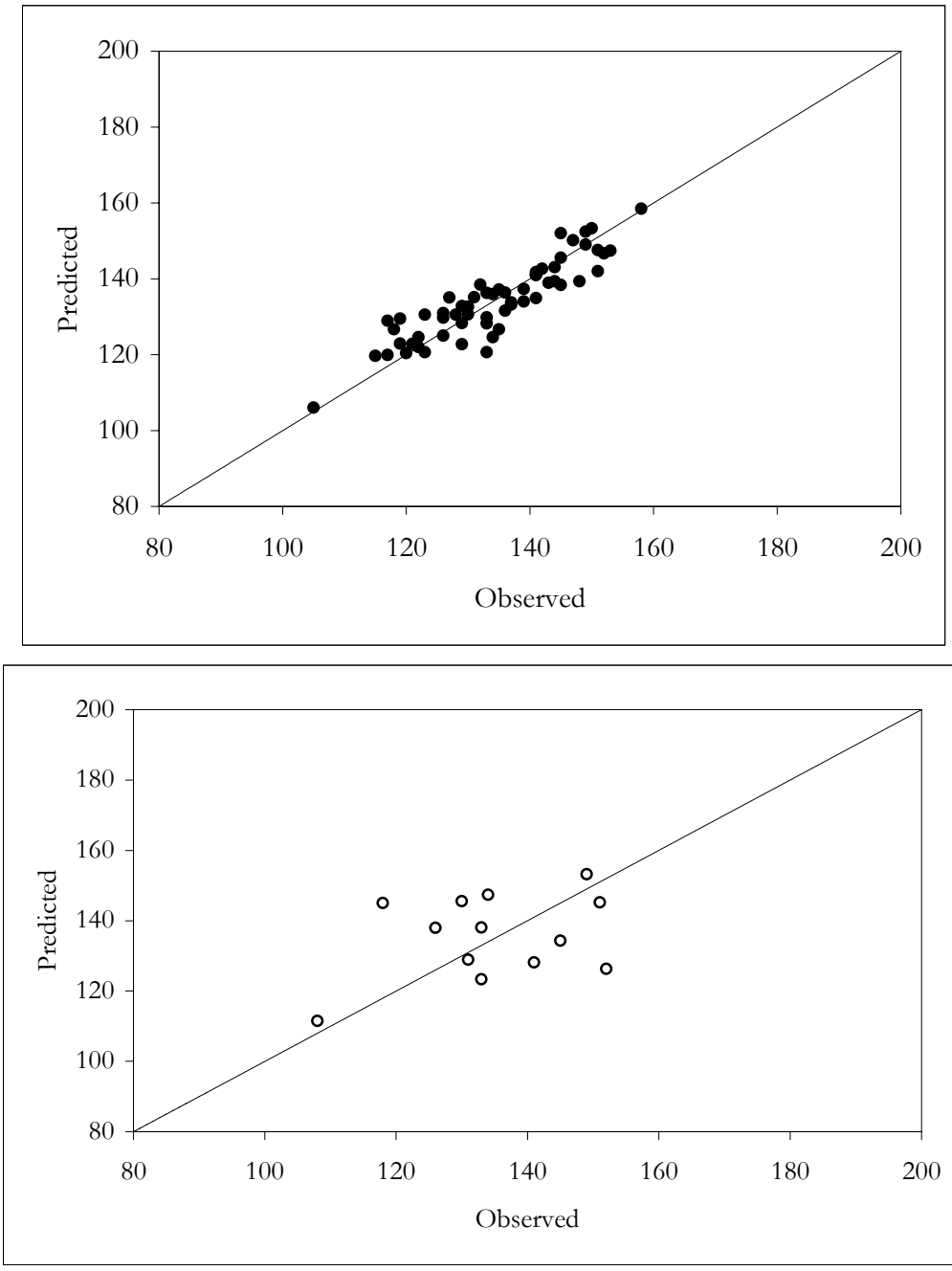


Figure 4.6. XY scatter plot of training (top) and validation (bottom) data sets for the second-order model, MDF 0.500", n=60.

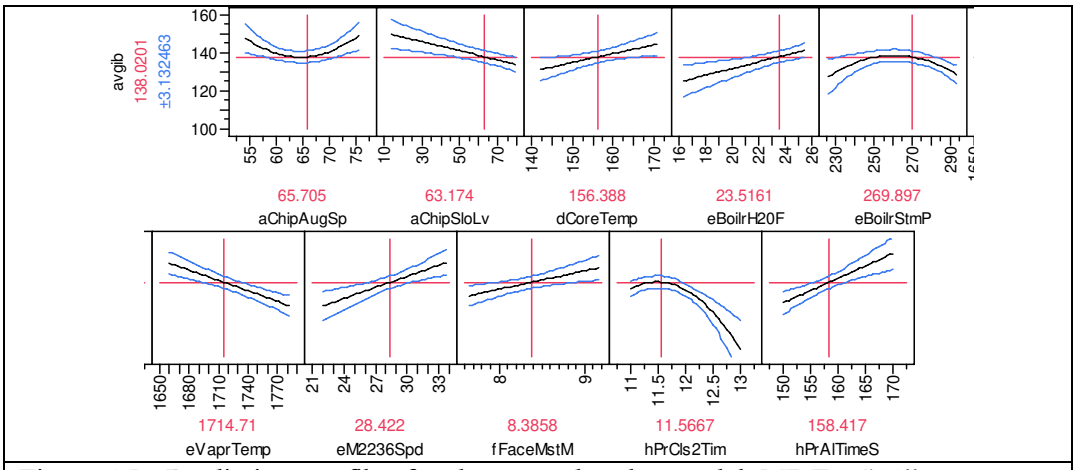


Figure 4.7. Prediction profiles for the second-order model, MDF 0.500", n=60.

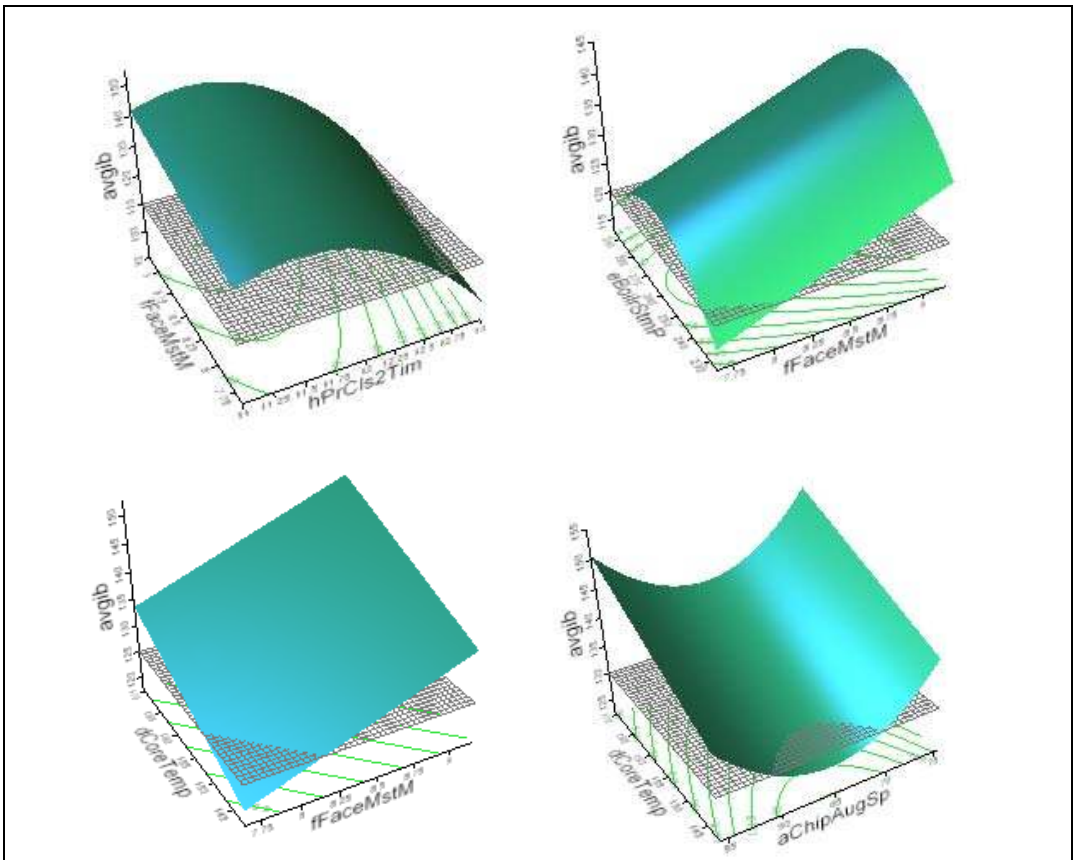


Figure 4.8. Response surface plots of IB for: "fFaceMstM" and "hPrCls2Tim" (upper left); "eBoilrStmP" and "fFaceMstM" (upper right); "dCoreTemp" and "fFaceMstM" (lower left); and "dCoreTemp" and "aChipAugSp" (lower right) for the second-order model, MDF 0.500", n=60.

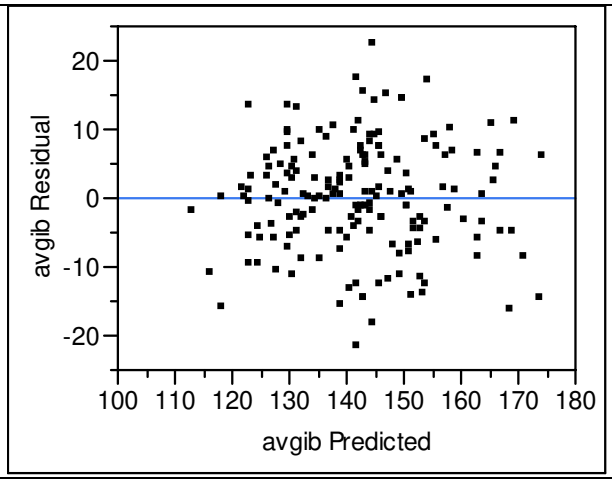


Figure 4.9. Residuals by predicted IB plot for the first-order model, MDF 0.500", n=175.

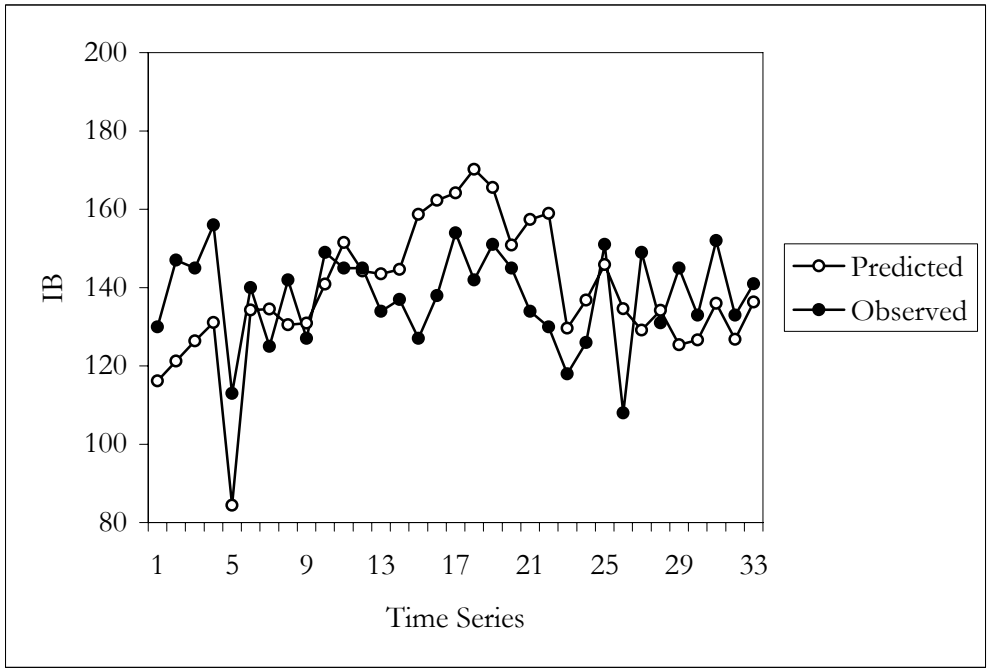


Figure 4.10. Time series graph of validation data set for the first-order model, MDF 0.500", n=33.

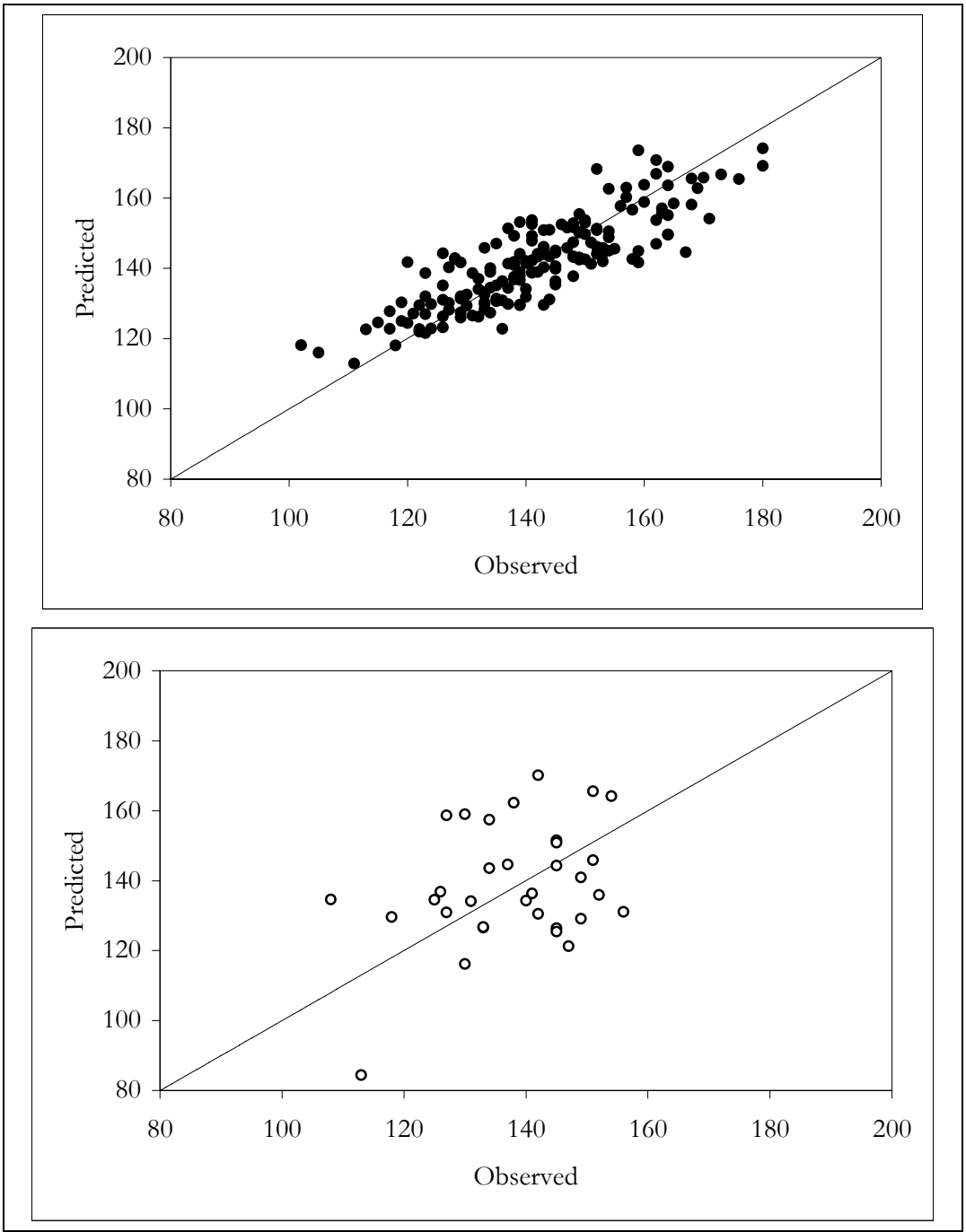


Figure 4.11. XY scatter plot of training (top) and validation (bottom) data sets for the first-order model, MDF 0.500", n=175.

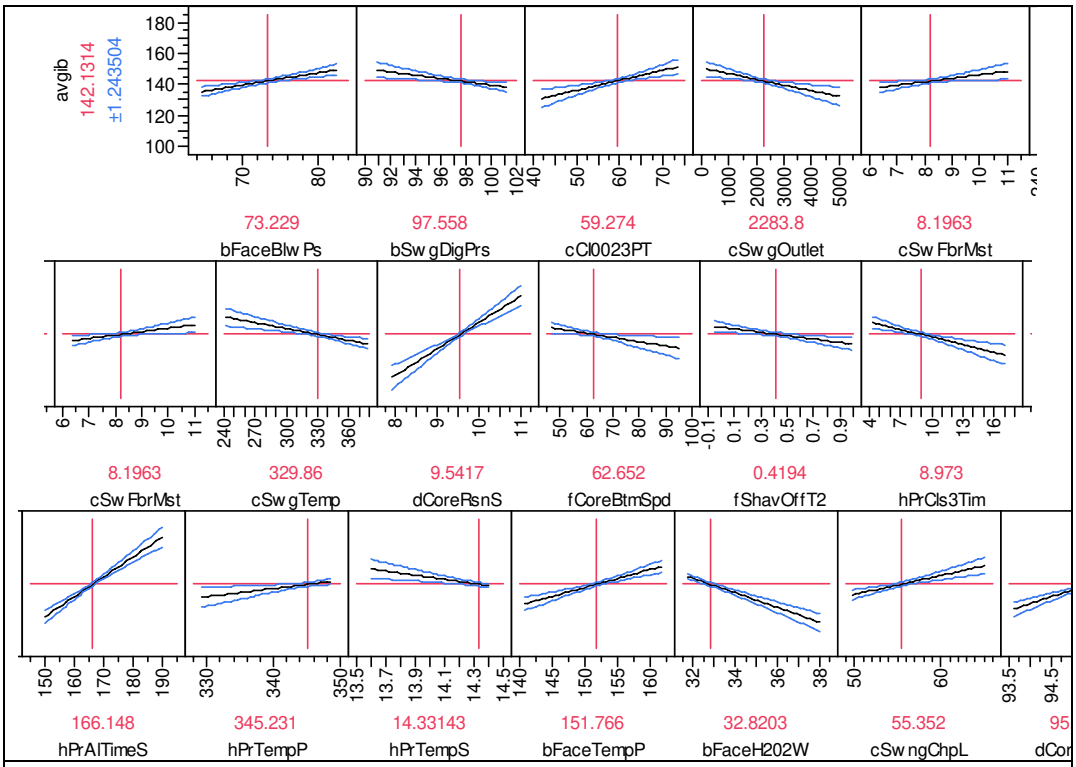


Figure 4.12. Prediction profiles for the first-order model, MDF 0.500", n=175.

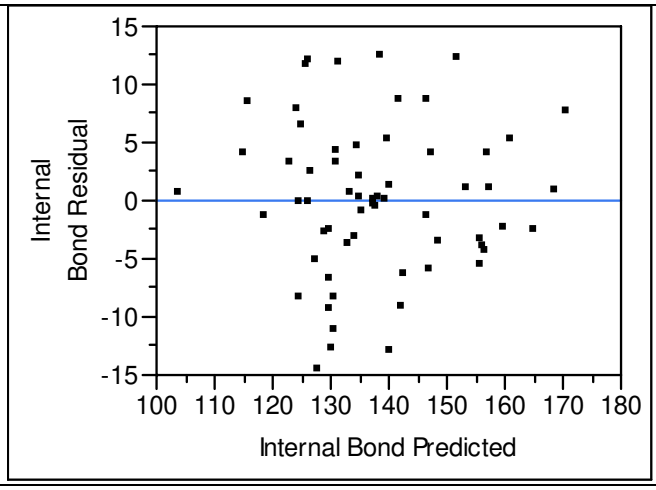


Figure 4.13. Residuals by predicted IB plot for the first-order model for MDF 0.625", n=62.

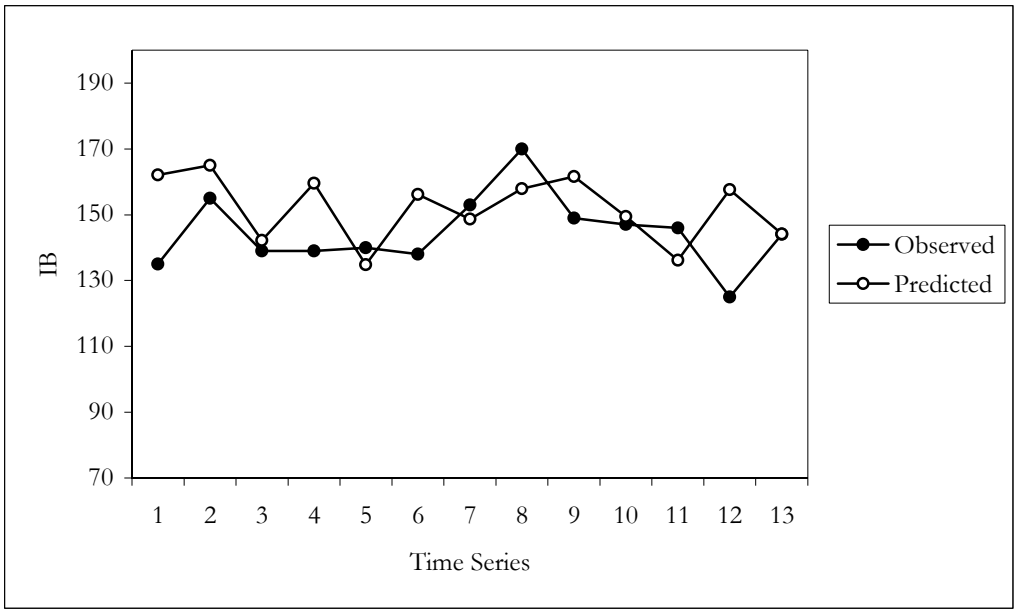


Figure 4.14. Time series graph of validation data set for the first-order model for MDF 0.625", n=13.

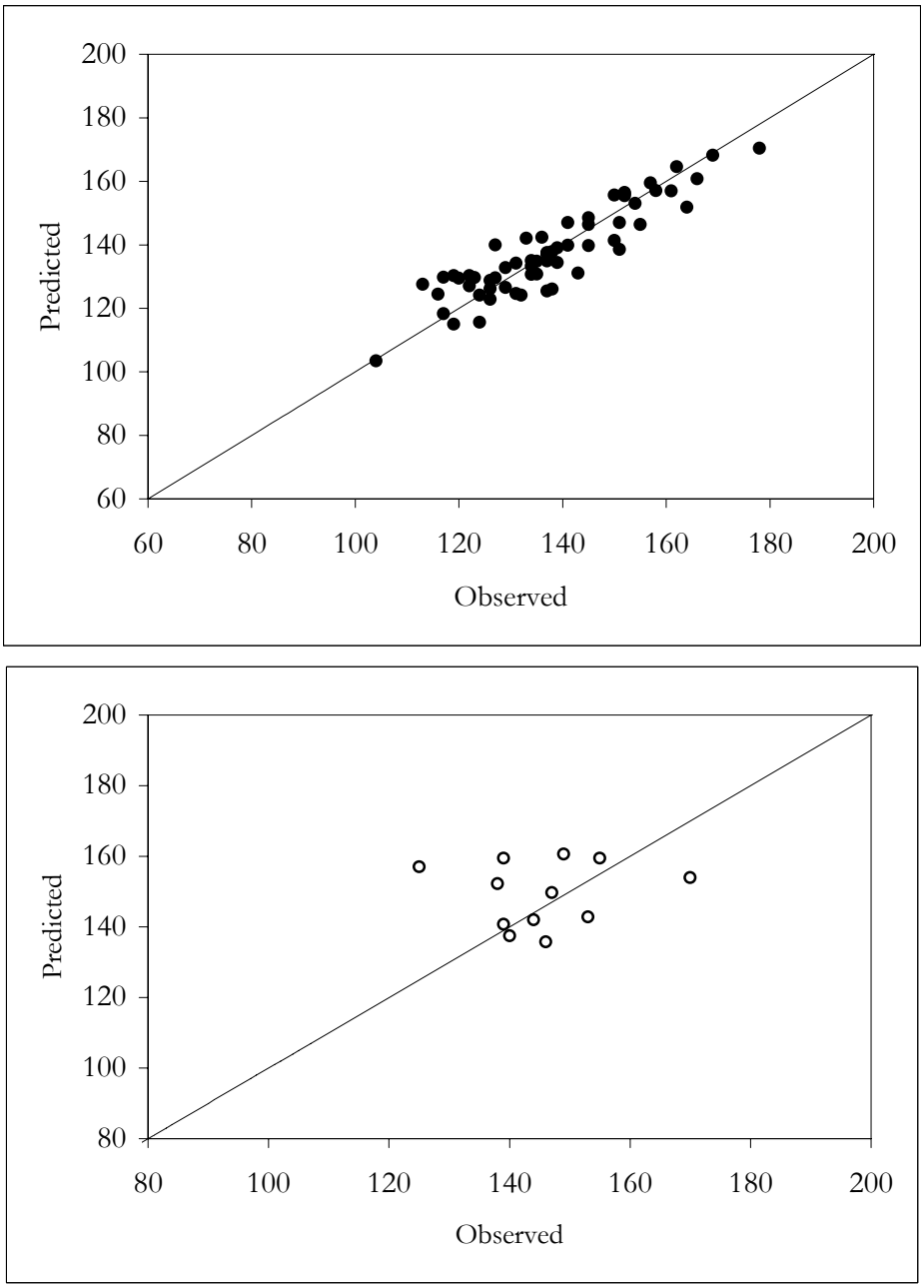


Figure 4.15. XY scatter plot of training (top) and validation (bottom) data sets for the first-order model for MDF 0.625", n=62.

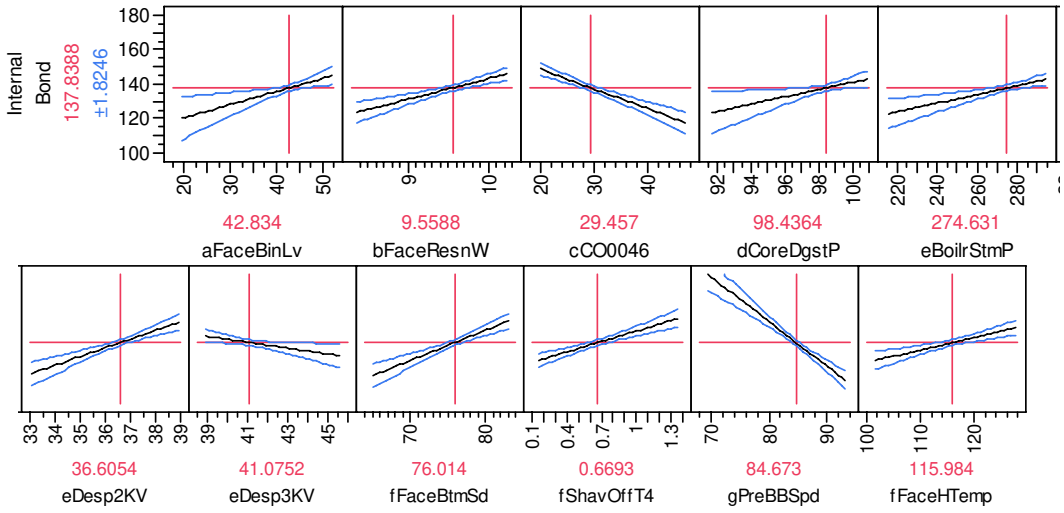


Figure 4.16. Prediction profiles the first-order model for MDF 0.625", n=62.

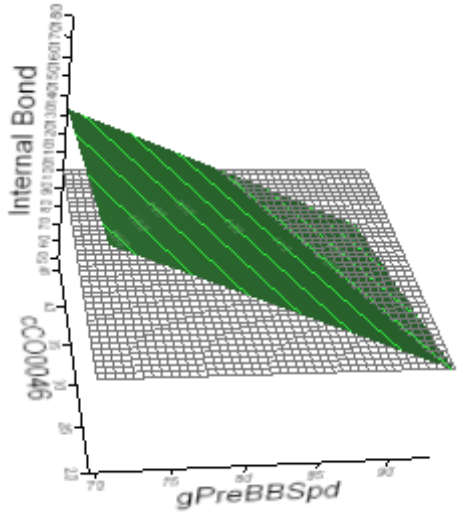


Figure 4.17. Response surface plot of IB for "gPreBBSpd" and "cC00046" for the first-order model for MDF 0.625", n=62. (grid-line 120 p.s.i.).

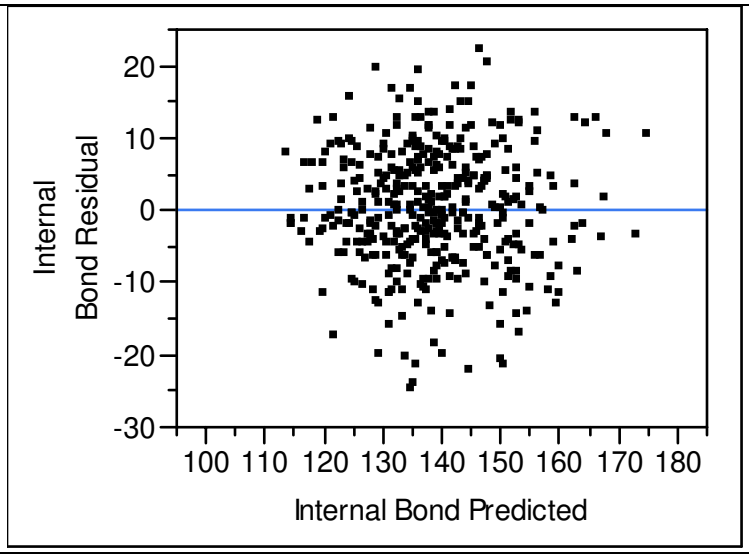


Figure 4.18. Residual by predicted IB plot for the second-order model for MDF 0.625", n=400.

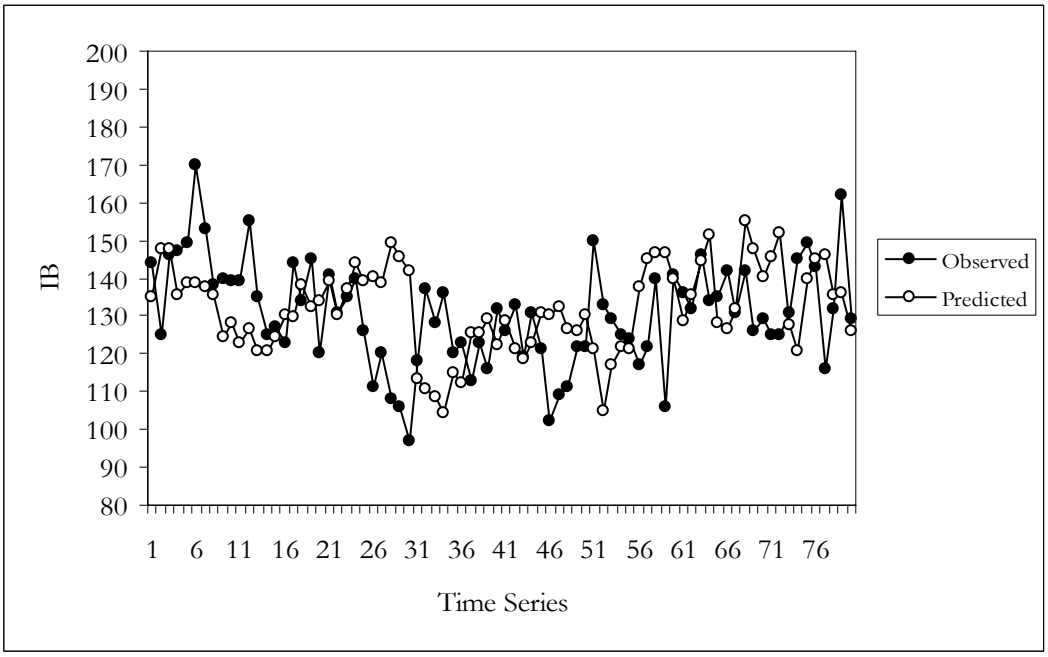


Figure 4.19. Time series graph of validation data set for the second-order model for MDF 0.625", n=80.

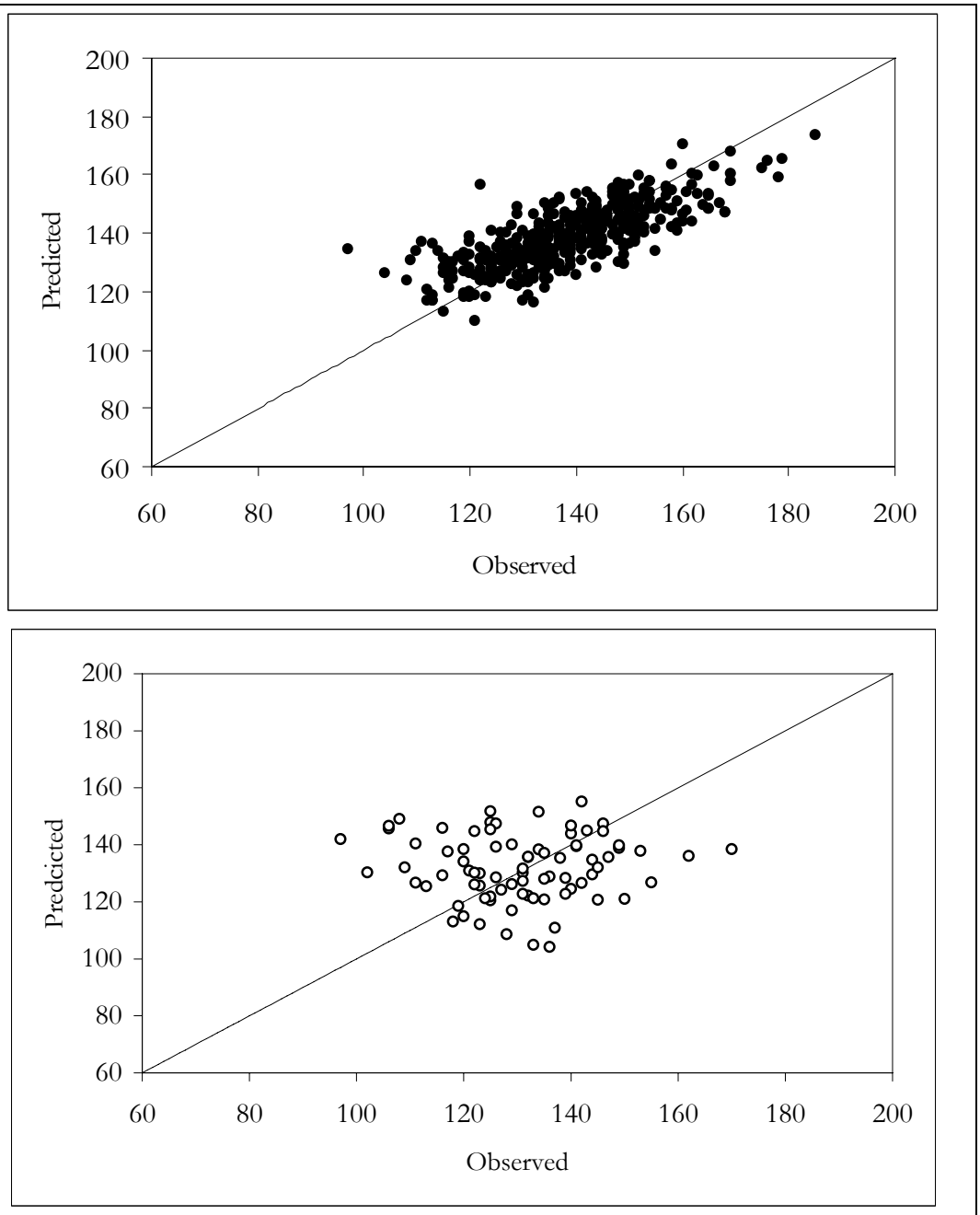


Figure 4.20. XY scatter plot of training (top) and validation (bottom) data sets for the second-order model for MDF 0.625", n=400.

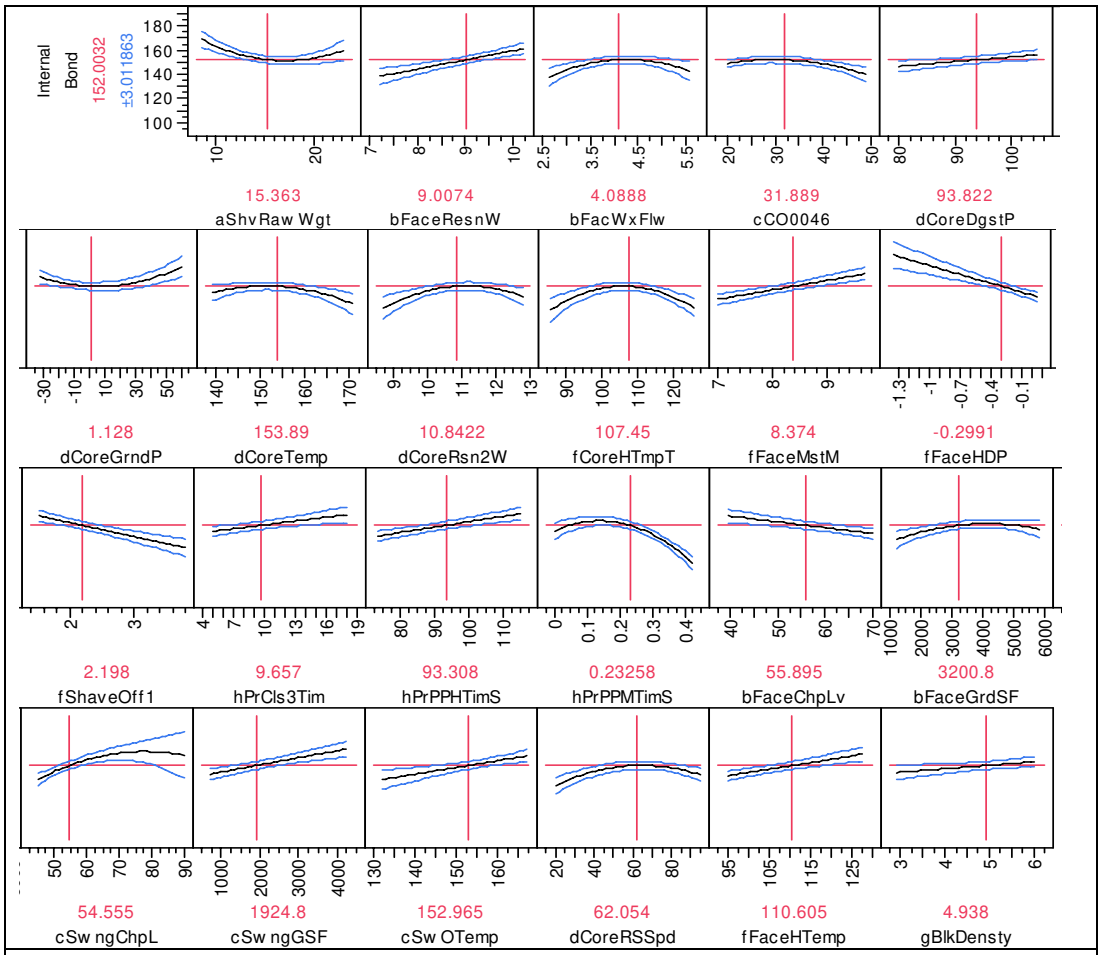


Figure 4.21. Prediction profiles for the second-order model for MDF 0.625", n=400.

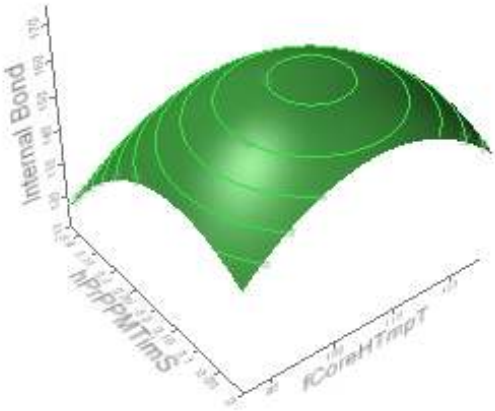


Figure 4.22. Response surface plot of IB for "hPrPPMTimS" and "fCoreHTmpT" for the second-order model for MDF 0.625", n=400.

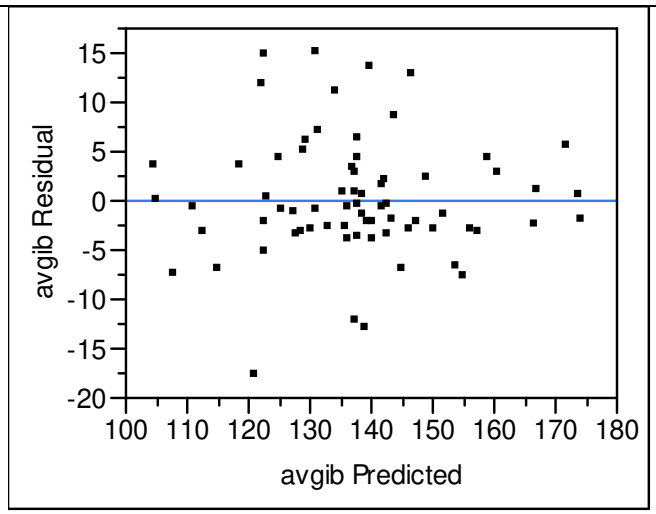


Figure 4.23. Residual by predicted IB plot for the second-order model for MDF 0.750", n=70.

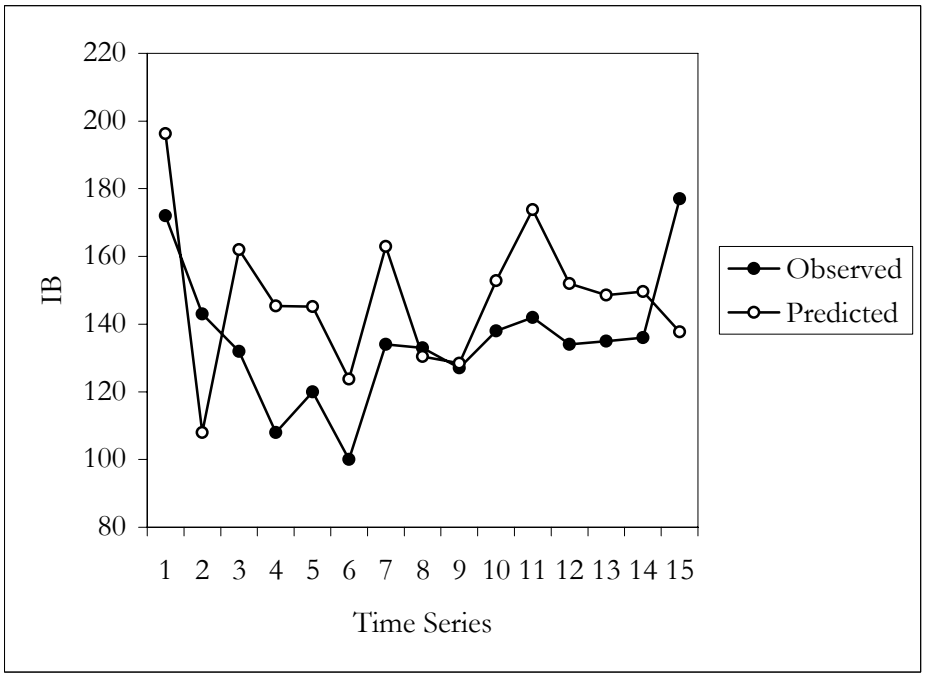


Figure 4.24. Time series graph of validation data set for the second-order model for MDF 0.750", n=15.

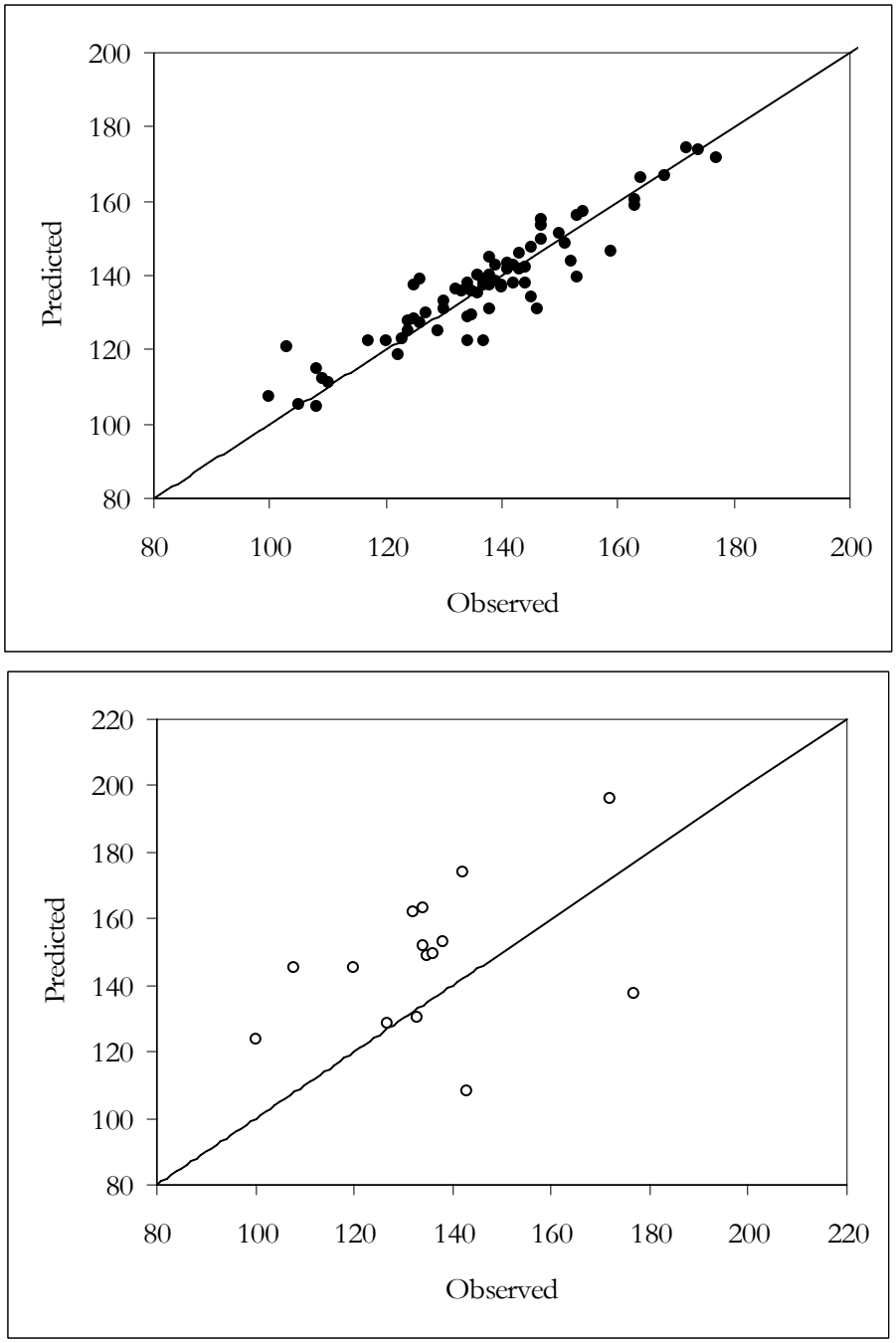


Figure 4.25. XY scatter plot of training (top) and validation (bottom) data sets for the second-order model for MDF 0.750", n=70.

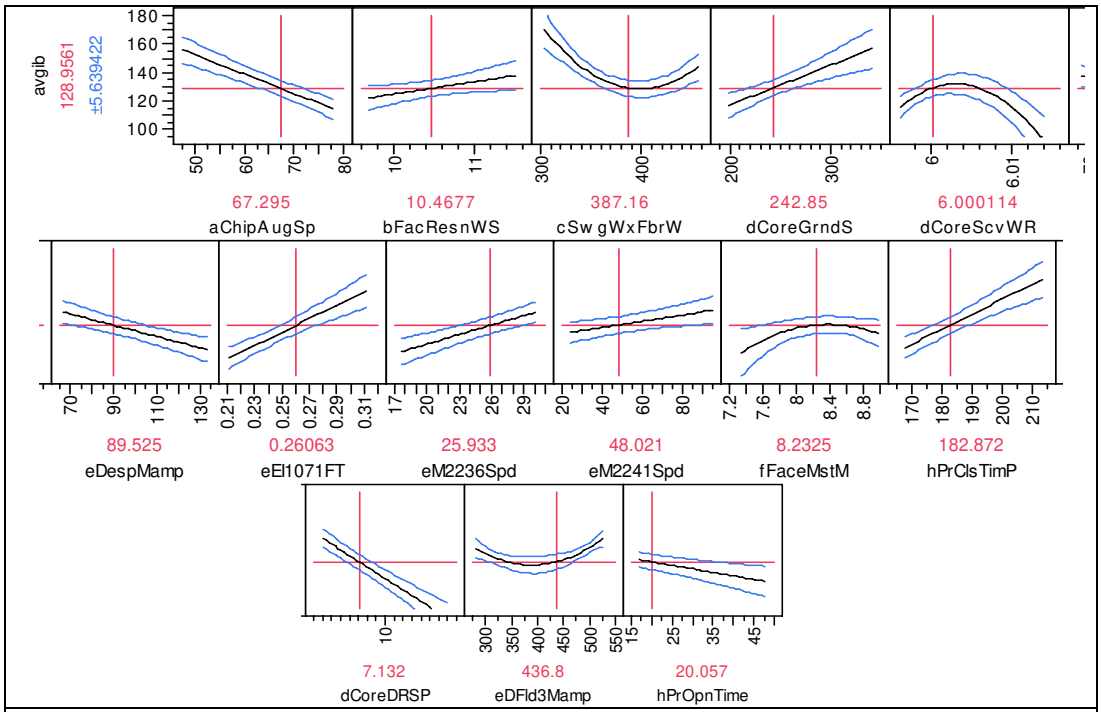


Figure 4.26. Prediction profiles for the second-order model for MDF 0.750", n=70.

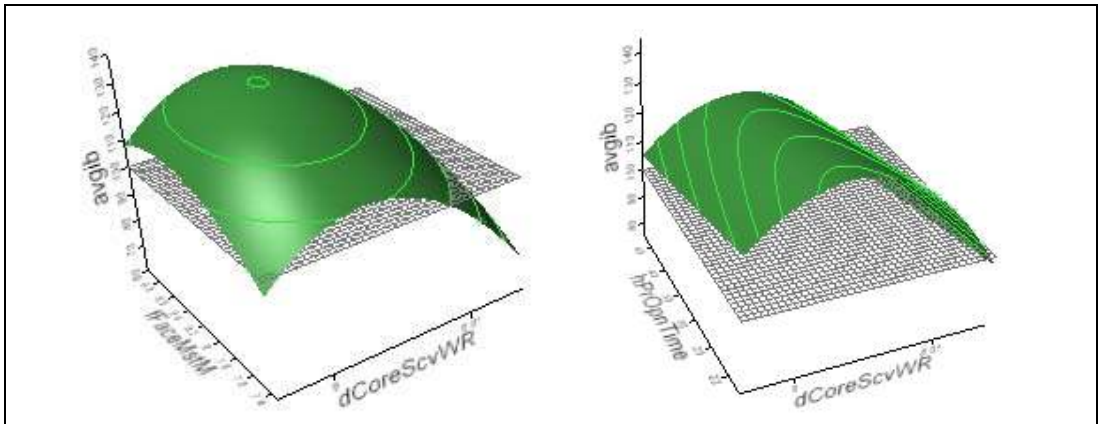


Figure 4.27. Response surface plot of IB for "fFaceMstm" and "dCoreScvWR" (left) and for "hPrOpnTime" and "dCoreScvWR" (right) for the second-order model for MDF 0.750", n=70.

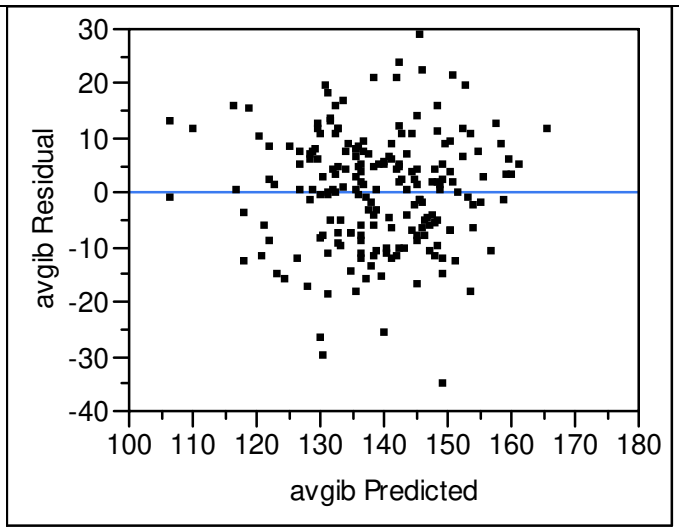


Figure 4.28. Residual by plot for the second-order model with interaction terms, MDF 0.750", n=200.

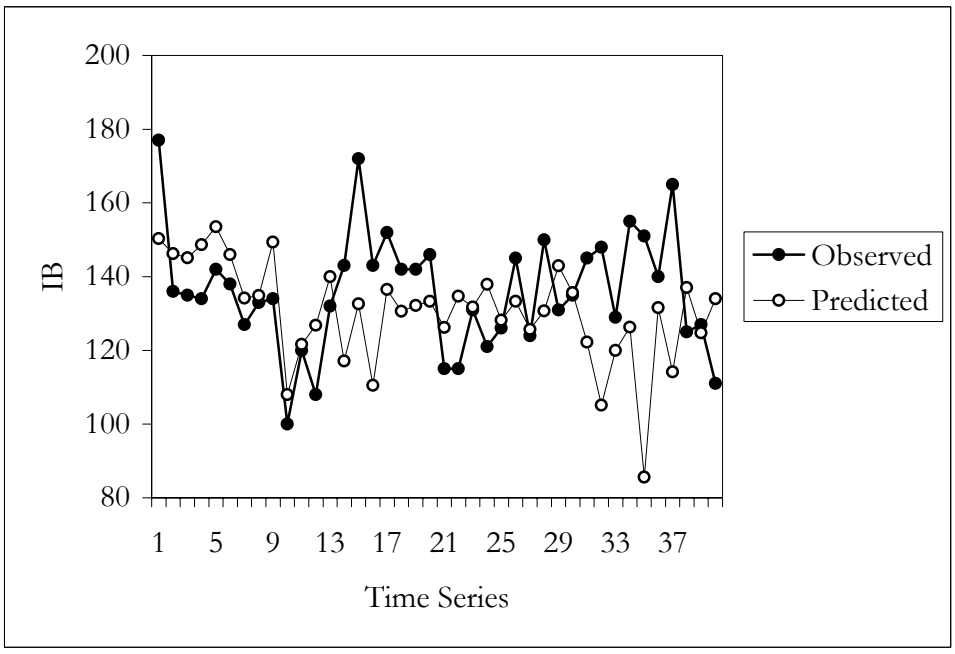


Figure 4.29. Time series graph of validation data set for the second-order model with interaction terms, MDF 0.750", n=40.

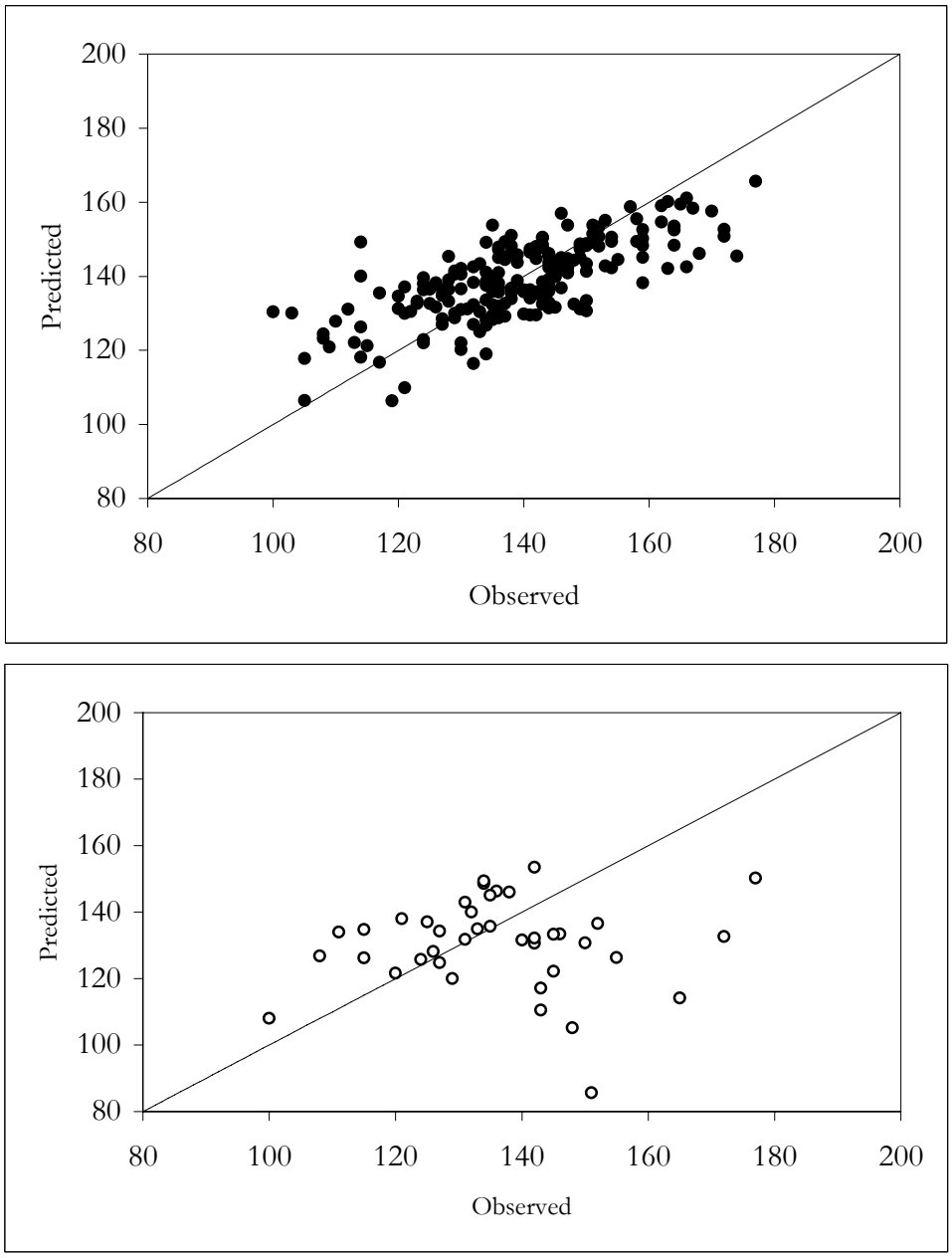
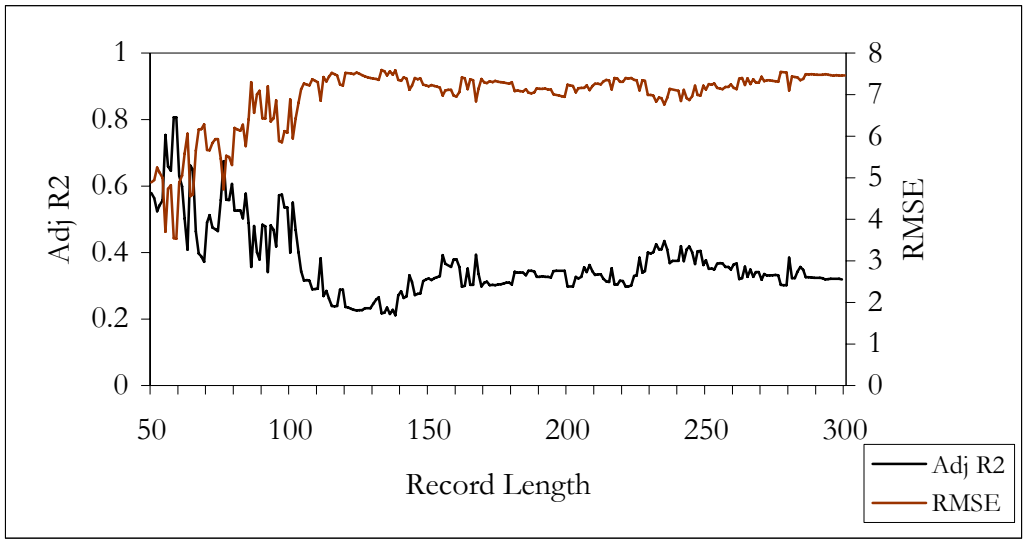


Figure 4.30. XY scatter plot of training (top) and validation (bottom) data sets for the second-order model with interaction terms, MDF 0.750", n=200.



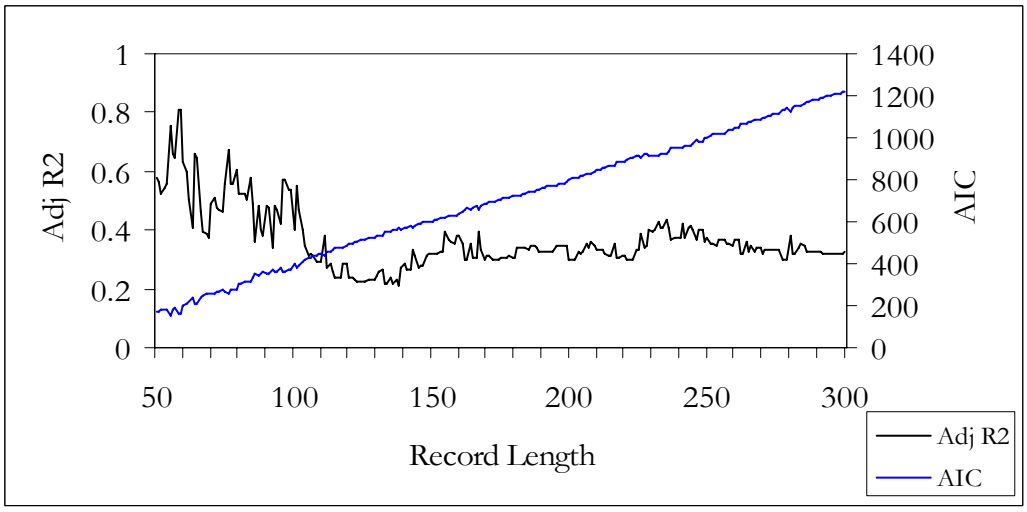


Figure 4.31. Comparison of Adjusted R^2 and RMSE (upper) and Adjusted R^2 and AIC (lower) for first-order models for OSB IB for every record length greater than 50.

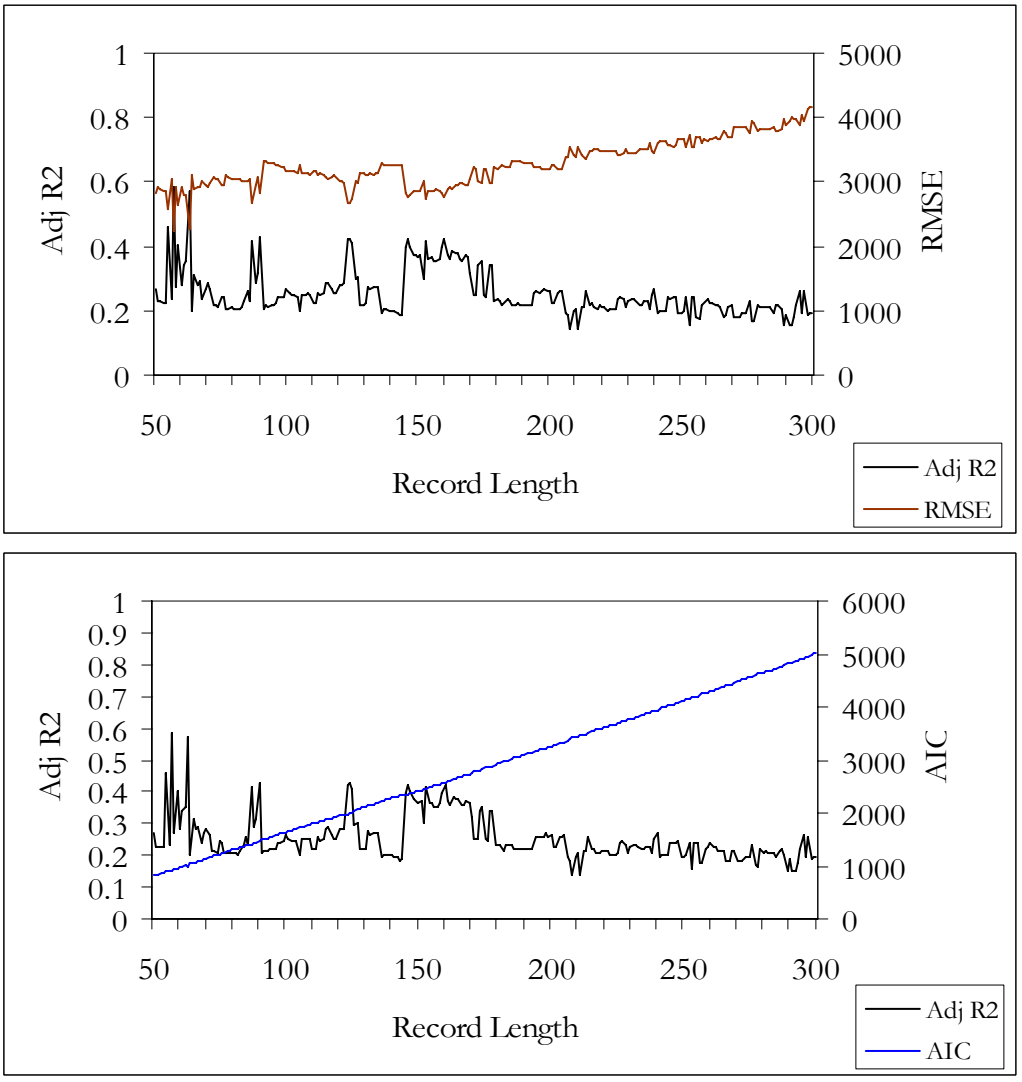


Figure 4.32. Comparison of Adjusted R^2 and RMSE (upper) and Adjusted R^2 and AIC (lower) for first-order models for OSB Parallel EI for every record length greater than 50.

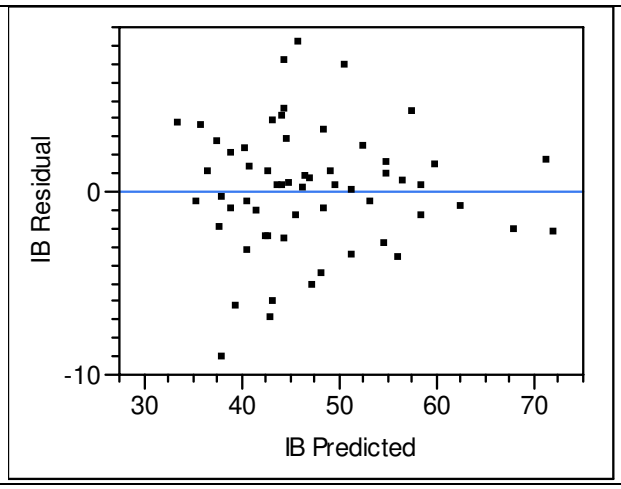


Figure 4.33. Residual by predicted IB plot for the second-order model with Box Cox transform, OSB IB, n=59.

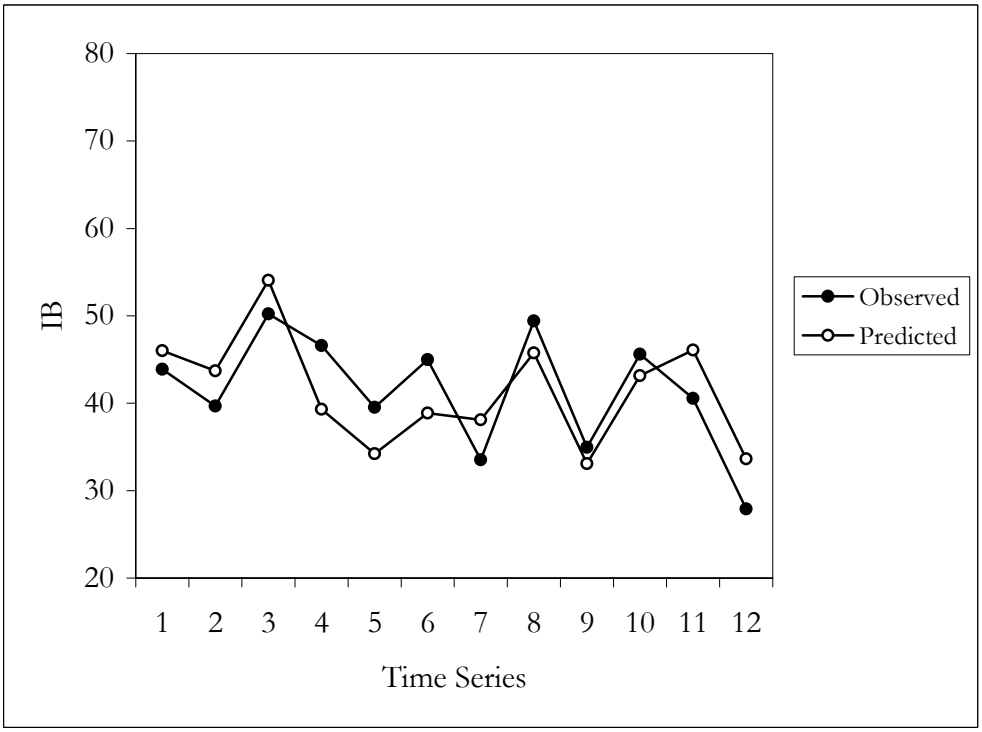


Figure 4.34. Time series graph of validation data set for the second-order model with Box Cox transform, OSB IB, n=12.

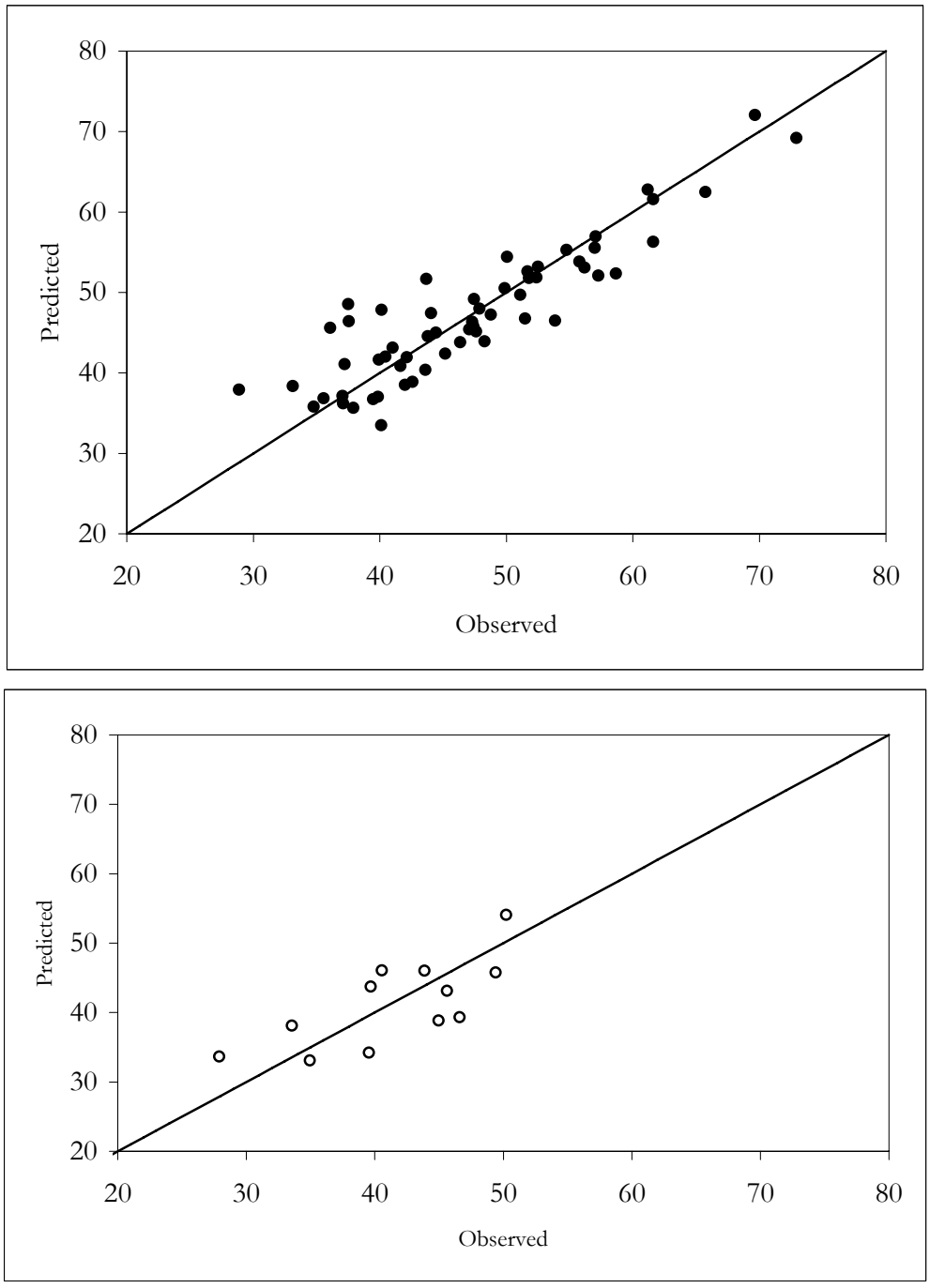


Figure 4.35. XY scatter plot of training (top) and validation (bottom) data sets for the second-order model with Box Cox transform, OSB IB, n=59.

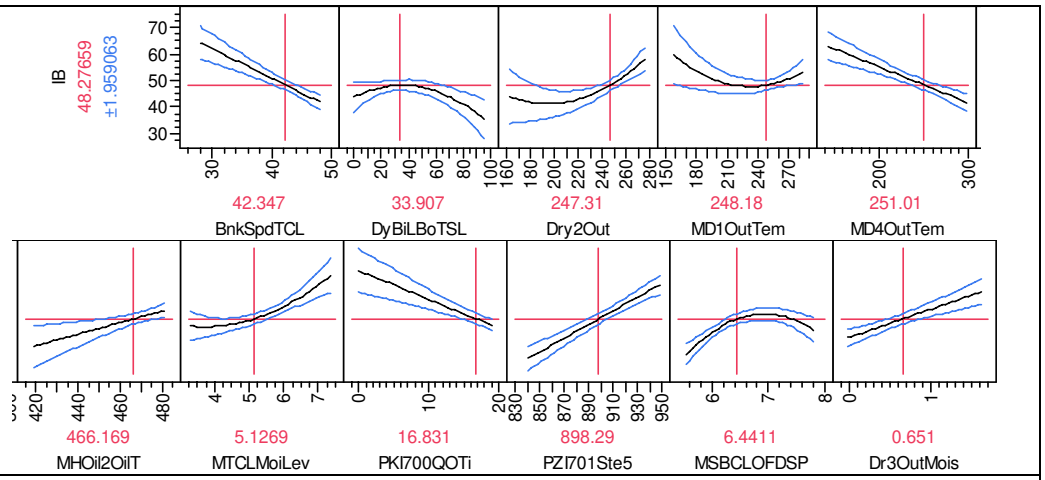


Figure 4.36. Prediction profiles for the second-order model with Box Cox transform, OSB IB, n=59.

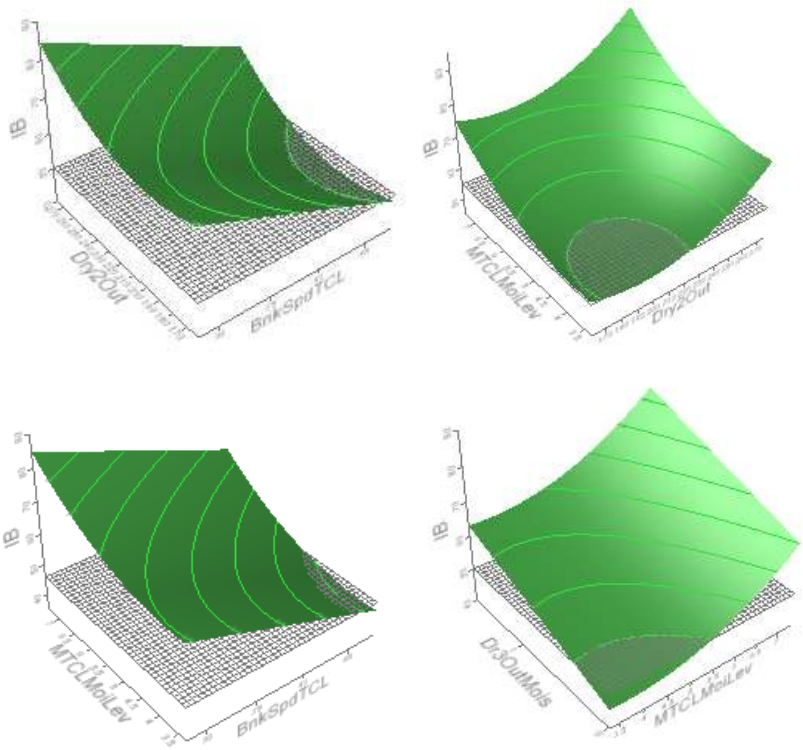


Figure 4.37. Response surface plots of OSB Parallel EI for: "Dry2Out" and "BnkSpdTCL" (upper left); "MTCLMoiLev" and "Dry2Out" (upper right); "MTCLMoiLev" and "BunkerSpdTCL" (lower left); and "Dr3OutMois" and "MTCLMoiLev" (lower right) for the second-order model with Box Cox transform, OSB IB, n=59.

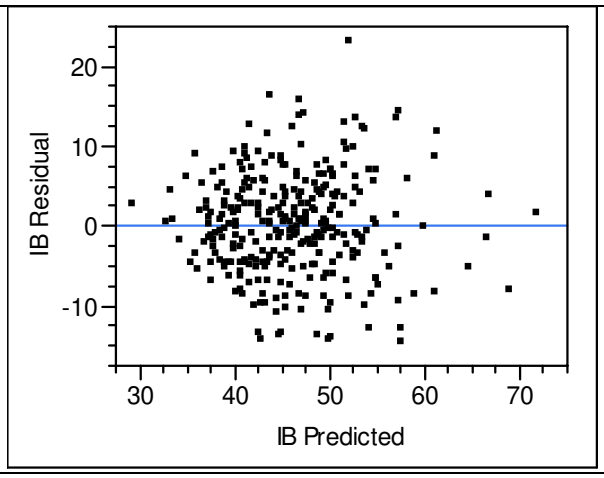


Figure 4.38. Residual by plot for the second-order model with Box Cox transform, OSB IB, n=300.

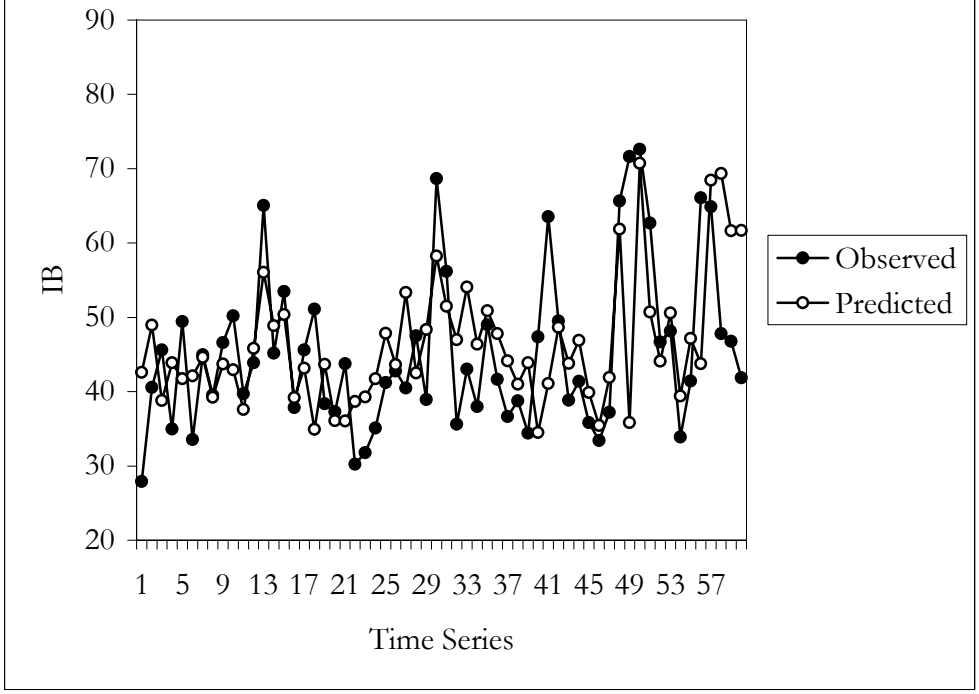


Figure 4.39. Time series graph of validation data set for the second-order model with Box Cox transform, OSB IB, n=60.

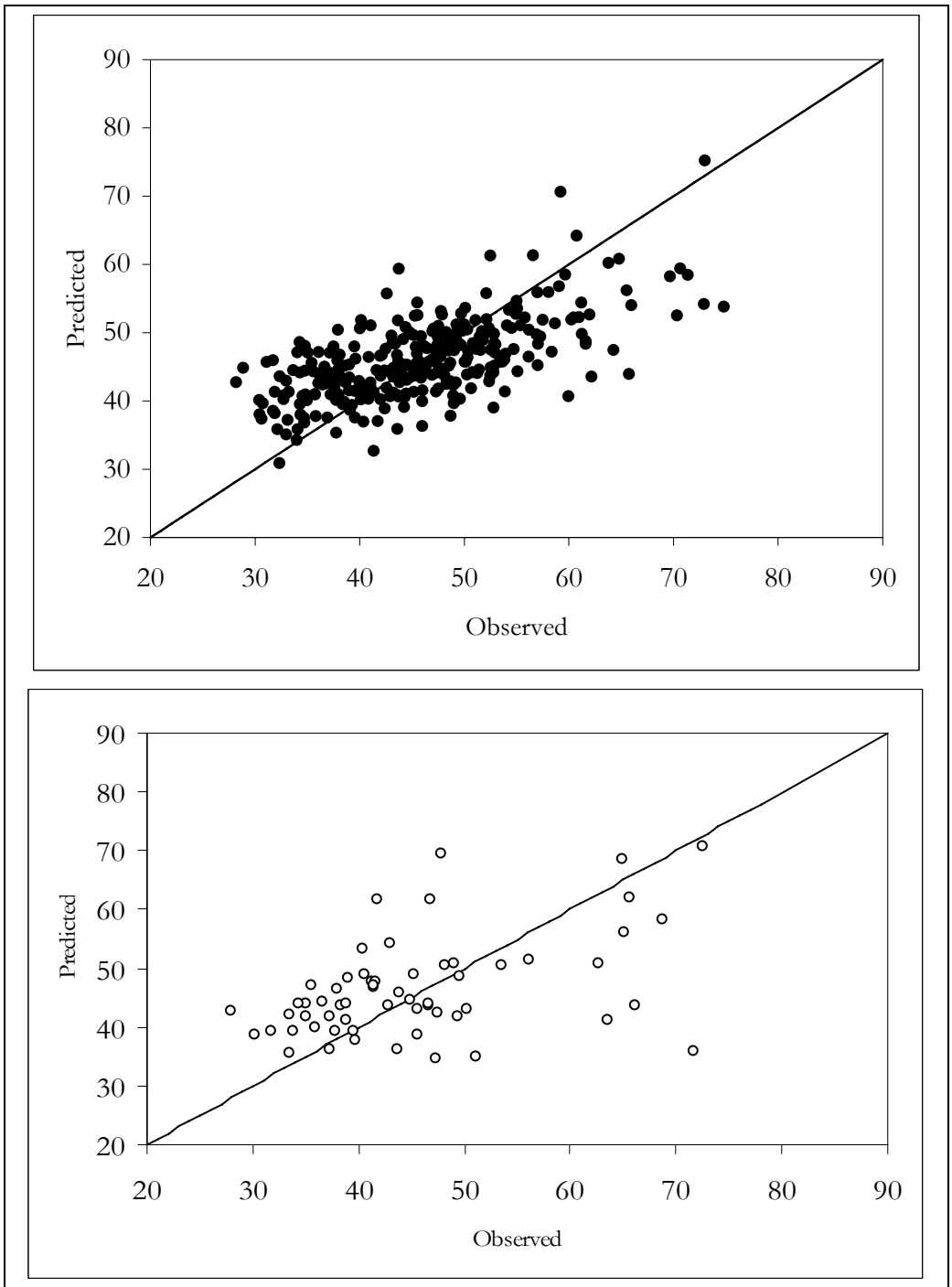


Figure 4.40. XY scatter plot of training (top) and validation (bottom) data sets for the second-order model with Box Cox transform, OSB IB, n=300.

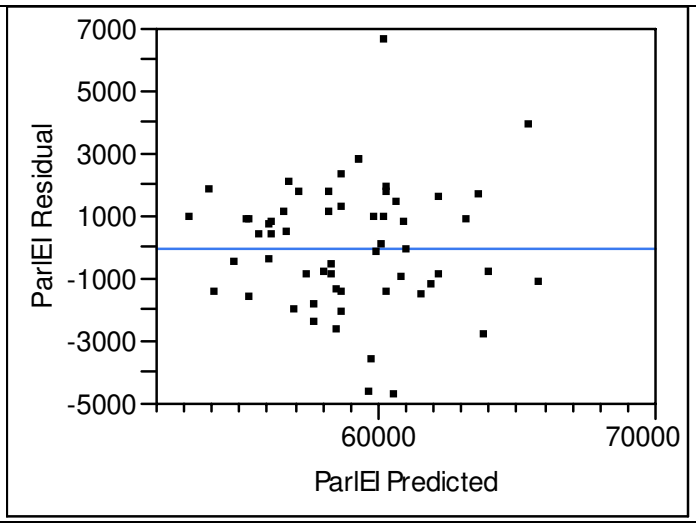


Figure 4.41. Residual by predicted Parallel EI plot for the first-order model, OSB Parallel EI, n=58.

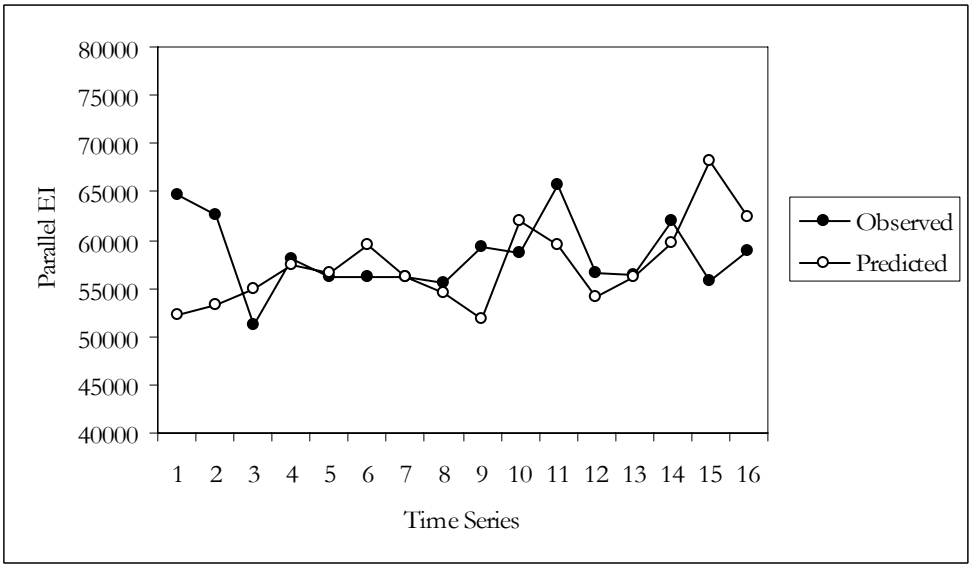
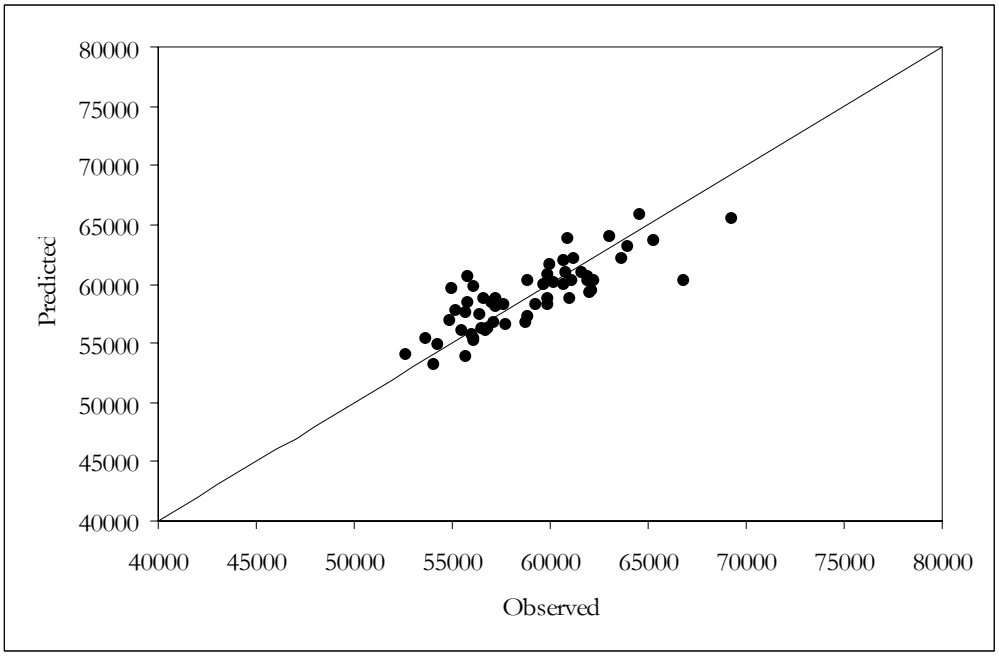


Figure 4.42. Time series graph of validation data set for the first-order model, OSB Parallel EI, n=16.



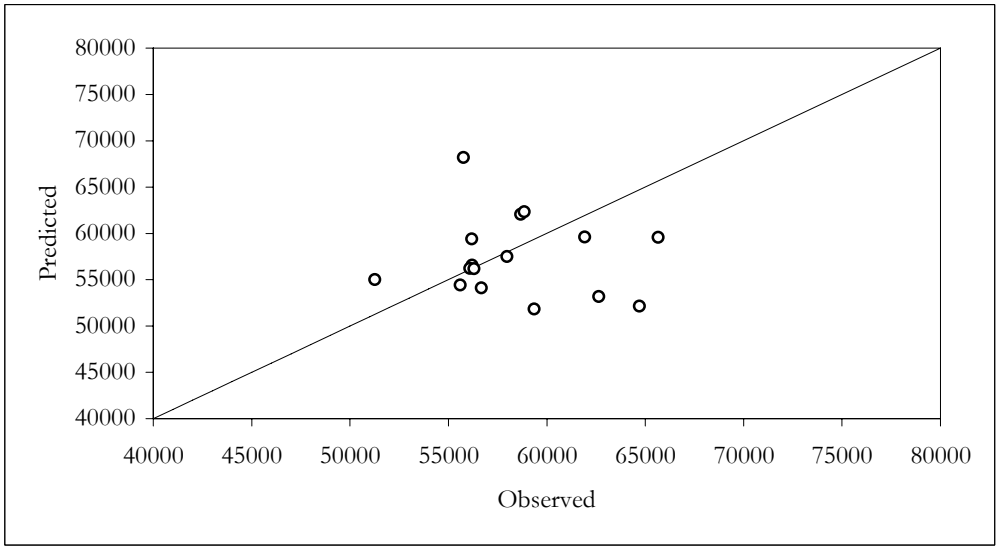


Figure 4.43. XY scatter plot of training (top) and validation (bottom) data sets for the first-order model, OSB Parallel EI, n=58.

CHAPTER V.

PARAMETRIC AND NON-PARAMETRIC REGRESSION TREE MODELS OF MDF AND OSB STRENGTH PROPERTIES

Results are presented in this chapter for the second and third objectives of the dissertation. Decision Tree (DT) or Regression Tree (RT) models are investigated that have explanatory value and predictive capability for the internal bond (IB) of medium density fiberboard (MDF) and oriented strand board (OSB). The Parallel Elasticity Index (EI), an important flexure strength property of OSB, is also explored using RT models. GUIDE (version 5.2) is used for RT model development (Loh 2006b). This research will hopefully advance the mathematical and industrial engineering sciences as applied to wood composites manufacture. Applications of this research for the practitioner are analytical methods that will improve the understanding of sources of variation of manufacturing processes that may facilitate improvements in product quality, product safety and lower costs.

To accomplish this objective, 56 parametric RT models and 32 non-parametric quantile RT models are investigated for the IB of MDF. Forty-eight parametric RT models and 24 non-parametric quantile RT models are further investigated for the IB and Parallel EI of OSB. Quantile regression RT models are investigated for the median or 50th percentile. For MDF, RT models for three nominal products are investigated, 0.500", 0.625" and 0.750". The OSB manufacturer produced predominately one product, 7/16" roof sheathing (RS) for which RT models are explored.

In the spirit of model-building and to advance the understanding of data set dimensionality and its influence on model development (Draper and Smith 1981, Myers 1990, Neter et al. 1996, and Kutner et al, 2004), RT models are developed for long and short record lengths, e.g., for 0.625" MDF RT models are developed for training data sets of 62 records, 100 records, 200 records, 300 records and 400 records (Table 5.1). The record length of 62 builds upon the results of Chapter IV. Validation data sets for both MDF and OSB are approximately 20% of the size of the training data sets (Draper and Smith 1981, Myers 1990, Neter et al. 1996, and Kutner et al. 2004). Note that all records are time-ordered and the validation data set are one continuous grouping of time-ordered records that start after the last record of the training data set. A challenge in this research in modeling industrial processes is low data set dimensionality, i.e., abundant data are available from on-line sensors but lack of data are available of destructive test strength properties due to infrequent sampling from the production line.

The MDF data set has 184 independent variables and the OSB data set has 234 independent variables (descriptions of variables are given in Appendix D). Many authors note that ideally record length should be six to ten times the number of independent variables (Draper and Smith 1981, Myers 1990, Neter et al. 1996, and Kutner et al, 2004). Unfortunately, the data set used for this research by Dawson et al. (2006) did not allow for this ideal to be met.

A strength of RT analysis is illustrated by the histograms for the MDF 0.625" process variables "core fiber temperature" and "swing refiner separator outlet pressure" (Figure 5.1). In RT analysis, models are investigated for these regressors as distinct sub-groups. This is further highlighted by the regression fits for each subgroup (or tree node)

of "swing refiner separator outlet pressure" (Figure 5.2). If analyzed as one subgroup (typical of MLR), the simple linear fit is not significant. When analyzed as two subgroups, the simple linear and polynomial regression fits have different slopes and intercepts. This illustration highlights some of the strengths of RT methods when modeling heterogeneous data (Loh 2007a, Kim et al. 2007b).

In general, pruning the RT models by v-fold cross-validation reduces the root mean square error of prediction (RMSEP) for MDF with the exception of the mixed stepwise model fits (Figures 5.3, 5.4 and 5.5). The improvement in the RMSEP from v-fold cross-validation pruning for MDF is more pronounced as the dimensionality of the data improve and approach a record length of 400 (Loh 2006b and Kim et al. 2007a). For OSB, a more difficult data set to model, the RMSEP is reduced by v-fold cross-validation for all RT models with the exception of two shorter record lengths for polynomial RT models (Figures 5.6 and 5.7). However, for both MDF and OSB, the RT models with the lowest RMSEP are not always the models with the most predictive capability. The coefficient of determination (R^2) for RT models by itself is not the best statistic in selecting a suitable RT model (Loh 2006b and Kim et al. 2007b). More in-depth analysis using XY scatter plots, time-ordered predictions of validation data sets, and plots of residuals are all important elements in model selection (refer to best model criteria previously discussed in Chapter III). The variable ranking feature of GUIDE (Loh 2007b) is used to pre-select significant regressors and improve the dimensionality of the data set as required for non-parametric (quantile) regression development.

Medium Density Fiberboard

Ranking of Key Regressors for all MDF Product Types

The ranking capability of GUIDE (Loh 2006) indicates that resin percentage, refining plate position and press set-points are all important in influencing IB variation (Table 5.2). "Waste fiber addition to refined core fiber" is a key variable for the thickest product (0.750") and may indicate that addition of waste fiber and its effect on IB needs further investigation by the manufacturer. A key process variable for the thinner 0.500" product is "mat weight at the Thayer scales." This may reflect difficulties with proper weight formation for thinner MDF products. "Crew" is a surprising source of IB variation for 0.625" MDF. This is an undesirable modeling regressor and makes the product more difficult to model. Operator induced variation can be a source of variation and usually indicates the need for additional training (Deming 1986 and 1993).

Medium Density Fiberboard RT Models

MDF 0.500" Thickness. -- Eighteen parametric and eight non-parametric RT models are explored for this product type. A candidate RT model **for a record length of 100** is a third-order quantile RT model with v-fold cross-validation node pruning (Figures 5.8, Tables 1c and 2c, Appendix C). The quantile RT model has five nodes and three terminal nodes. Key metrics of this RT model are a RMSE of 8.63, RMSEP of 32.56 p.s.i. and homogeneous residual pattern. Predicted IB does not approximate observed IB in the validation data set (Figure 5.9). Four possible outlier values of IB in the validation data set result in poor model predictability.

For a record length of 175, the best candidate model is a RT mixed stepwise regression model with v-fold cross-validation node pruning (Table 2c, Appendix C). Key

metrics for this RT model are a RMSE of 7.89 p.s.i., RMSEP of 16.26 p.s.i., R^2 of 0.72 and homogeneous residual pattern (Table 2c, Appendix C). Given the low dimensionality of the training data set, the RT consists of only one root node after v-fold cross-validation pruning (Table 5.3). Predicted IB tends to follow the general trend of observed IB in the validation data set with the exception of 12 contiguous records of overestimation in the middle of the validation data set (Figures 5.10 and 5.11). A quantile RT model of the median IB reduces the RMSE to 7.70 p.s.i. but increases the RMSEP to 26.11 p.s.i. and has an undesirable predicted IB trend.

A good candidate model using the shorter record length of 60 (investigated for MLR in Chapter IV) is a second-order RT model without node pruning. The RT model has good explanatory value and moderate predictive capabilities of IB (Figures 5.12 and 5.13). The RT model has 15 total nodes and eight terminal nodes. The RT model has a RMSE of 2.16 p.s.i., RMSEP of 15.74 p.s.i. with a tree model R^2 of 0.96 and a homogeneous residual pattern (Table 1c, Appendix C).

MDF 0.625" Thickness. -- Thirty-two parametric and 16 non-parametric quantile RT models are investigated for this product type (Tables 3c and 4c, Appendix C). **For a record length of 100**, the best candidate model is a mixed stepwise regression model with v-fold cross-validation node pruning (Table 5.4, Table 4c, Appendix C). Key metrics of this RT model are a RMSE of 9.42 p.s.i., RMSEP of 14.84 p.s.i., R^2 of 0.51 and a homogeneous residual pattern. The RT consists of only one root node after v-fold cross-validation pruning. Loh (2007) indicated in personal e-mails that pruning of a large number of nodes in GUIDE, even back to one node, is typical for data sets with low dimensionality. Predicted IB was not favorable and tends to over-estimate observed IB of

the time-ordered validation set. This outcome is not surprising given the low dimensionality of the data set and excessive node pruning.

The possible benefit of the model may be explanatory in the regressor variables of the mixed stepwise regression equation (Table 5.4). Regressor variables are related to chip digesting and refining of face fiber (fiber quality), resin percentage and face moisture content. A quantile RT model of the median IB reduces the RMSE to 8.79 p.s.i. but results in an unacceptable increase of the RMSEP to 21.47 p.s.i. and poor prediction of IB in the validation set.

For a record length of 200, the best possible candidate RT model is the piecewise simple linear model with v-fold cross-validation pruning. The predictability for this RT model declines after the 11^{th} record in the time-ordered validation data set. Key metrics for this RT model are a RMSE of 11.16 p.s.i., RMSEP of 11.44 p.s.i. and a homogeneous residual pattern. The RT model consists of three nodes with two terminal nodes (Figure 5.14). A quantile RT model of the median IB reduces the RMSE to 10.73 p.s.i. but results in an increase of the RMSEP to 12.90 p.s.i. and has poor predictability of the validation data set.

Even though model quality is not necessarily optimal, the explanatory value of RT models starts to reveal itself with a record length of 200. The split of cCI0023PT ("refiner steam pressure") indicates possible heterogeneous data and the simple linear regression models for each node indicate that "press pre-position move time set-point" has a large negative influence when "refiner steam pressure" > 48.1. "Core fiber moisture" had a negative influence when "refiner steam pressure" > 48.1. Surprisingly, the regressor variable "Lab Technician = JY" is statistically significant in both node models and has a

negative influence on IB. This may also explain some of the difficulty in developing a predictive model for this product. This is an unacceptable manufacturing outcome given that a person or crew should not result in a lower IB or increase IB variance. This source of variation should be vigorously investigated by the manufacturer and further training of this person and crew may be necessary. Recall Deming's (1986 and 1993) views on operator induced variance from "tampering" and his focus on training of operators and managers.

For a record length of 300, the quantile RT model has a RMSE of 9.31, RMSEP of 16.33 p.s.i. and a homogeneous residual pattern. The RT consists of only one root node after v-fold cross-validation pruning. Key regressors in the quantile regression equation are "press pre-position set-point" which has a strong negative influence on IB. "Pre-compressor belt speed" and surprisingly "swing wax percent" have positive influences on IB. "Forming line speed" has a negative influence on IB (Table 5.5). Predicted IB does not approximate observed IB in the validation data set but in some instances follows the time-ordered trend of IB (Figures 5.15 and 5.16).

For a record length of 400, the most suitable RT model is a piecewise simple linear model with v-fold cross-validation node pruning. Key metrics for this RT model are a RMSE of 11.33 p.s.i., RMSEP of 18.08 p.s.i., R^2 of 0.34 and the residual pattern is homogeneous (Table 4c, Appendix C). This RT model has 9 total nodes with 5 terminal nodes. The desirable attribute of this RT model is its explanatory value in the regression tree (Figure 5.17). The highest level of IB (mean = 145.7 p.s.i.) is for a "core fiber moisture > 10.0% and "core finer temperature" ≤ 142.4. The lowest level of IB (mean = 133.3 p.s.i.) is for "core fiber moisture > 10.0% and "core finer temperature" > 142.4.

The manufacturer should investigate these variables further as such variables are sources of IB variation. The predicted IB in the validation data set does not approximate observed IB after about the first six values (Figures 5.18 and 5.19). This further highlights the difficulty of predictive modeling of industrial processes (Bernardy and Scherff 1998, Benardy and Scherff 1999, Young and Guess 2002, Young and Huber 2004, Young et al. 2004, Dawson et al. 2006, Kim et al. 2007). In this case the RT model would have to be reasseseed after about six IB tests in the validation data set.

The quantile RT model candidate for this record length has a RMSE to 13.16 p.s.i. and RMSEP to 14.32 p.s.i. The predicted IB in the validation data set is worse than the piecewise simple linear model with v-fold cross-validation node pruning.

A candidate model using the shorter record length of 62 (investigated for MLR in Chapter IV) is a second-order RT model with v-fold cross-validation pruning. The RT model has good explanatory value and predictive capability (Figures 5.20 and 5.21). The RT model has three total nodes with two terminal nodes. The model has a RMSE of 9.55 p.s.i., RMSEP of 11.29 p.s.i. with a tree model R^2 of 0.60 and a homogeneous residual pattern (Table 4c, Appendix C).

MDF 0.750" Thickness. -- Sixteen parametric and eight non-parametric quantile RT models are investigated for this product type (Tables 5c and 6c, Appendix C). **For a record length of 100**, a suitable candidate model is a non-parametric quantile RT model without node pruning. Key metrics for this RT model are a RMSE of 8.54 p.s.i., RMSEP of 22.17 p.s.i. and homogeneous residual pattern. The RT model has 13 total nodes and seven terminal nodes. Attractive features of this model are its explanatory value from the regression tree and also its predictive capability in the near term (Figures 5.22 and 5.23).

The highest level of IB (mean = 151) occurs with "refiner steam pressure" > 54.6 and "press start control" ≤ 933.0. The lowest level of IB (mean = 132) occurs with "refiner steam pressure" ≤ 54.6 and "dry fuel bin speed" ≤ 27.7. This model is good example of the explanatory value of RT models (Loh 2007, Kim et al. 2007, Kim et al. 2007), Figure 5.23. "Core scavenger resin rate" has a significant influence on IB within each sub-tree as indicated by the quantile regression coefficients. However, the level and direction of influence of "core scavenger resin rate" is dependent on other process variables and levels of the process variable within a sub-tree.

Predictions of IB follow the trend of observed IB with a high degree of accuracy for the first eight records of the validation data set (Figure 5.22). With the exception of the later records of the validation data set (17 through 20), the XY scatter plot reveals that IB predictions follow the scale of observed IB (Figure 5.24).

For a record length of 200, the best candidate model of the 12 models investigated is a piecewise simple linear RT model without node pruning. The RT model is quite large and consists of a total of 75 nodes with 38 terminal nodes. Given the length of this RT model, it is in presented in GUIDE format as Illustration 1B, Appendix B. Key metrics for this RT model are a RMSE of 2.97 p.s.i., RMSEP of 20.03 p.s.i., R^2 of 0.96 and a homogeneous residual pattern. The RT model is over fit without node pruning but it is the only model for this record length that has any approximation to observed IB in the validation data set (Figures 5.25 and 5.26).

The highest level of IB (mean = 156) for this model occurs for "core dust ratio set-point" ≤ 7.5, "mat shave off height #1" > 0.14, "mat shave off height #1" ≤ 0.52, "fiber bulk density" ≤ 5.0, "core fiber humidifier temperature" > 102.6, "swing dyer outlet

temperature" ≤ 154, "eDesp_Field3_Millamps" > 494.5 and "swing refiner fiber moisture" ≤ 8.0. The significant ($\alpha <$ 0.01) linear regressor within this sub-tree is "press position time" with a coefficient of 3.17, i.e., for every second increase in "press position time" IB increases by 3.17 p.s.i. if all other regressors are held constant.

The lowest level of IB (mean = 112) occurs for "core dust ratio set-point" > 7.5, "core scavenger rate" ≤ 6.0, "press temperature" > 347.2, and "face fiber humidity" > 50.5. The significant ($\alpha <$ 0.01) linear regressor for this tree is "mat shave off height #1" with a positive coefficient of 53.9.

A second-order RT model with v-fold cross-validation pruning using the shorter record length of 70 (investigated for MLR in Chapter IV) improved the ability to model this product type. The RT model has one node, RMSE of 14.69 p.s.i., RMSEP of 15.87 p.s.i. with a tree model R^2 of 0.20 and a non-homogeneous residual pattern (Figures 5.27 and 5.28, Table 6c, Appendix C).

Oriented Strand Board

Ranking of Key Regressors for the Parallel EI and IB of OSB

The ranking capability of variables from GUIDE (Loh 2006) indicate that variation in top and bottom core layer moistures are significant factors influencing both Parallel EI and IB (Table 5.6). Process variables related to press time also appear to be important but to a lesser extent. Process variables related to forming speed also appear to be important in influencing Parallel EI strength variation. The manufacturer should investigate these sources of variation in the process as such variation limits product quality and safety improvements, and potential cost savings (Deming 1986, 1993).

Oriented Strand Board RT Models

Internal Bond. -- Twenty-four parametric and 12 non-parametric quantile RT models are investigated (Tables 7c and 8c, Appendix C). **For a record length of 100**, the best candidate model is mixed stepwise all possible subsets RT model without node pruning. Key metrics for this RT model are a RMSE of 5.79 p.s.i., RMSEP of 9.24 p.s.i., R^2 of 0.64 and a homogenous residual pattern. The RT model consists of a total of 23 nodes with 12 terminal nodes (Illustration 2B, Appendix B). Predicted IB for the validation data set approximates the observed IB in the time-ordered validation data set (Figures 5.29 and 5.30). Development of a quantile RT model of the median IB reduces the RMSE to 3.95 p.s.i. but unsatisfactorily increases the RMSEP to 15.48 p.s.i. with unsatisfactory predicted IB in the validation data set.

The subgroup with the highest mean IB (61.1 p.s.i.) for this model occurs for a "top core layer moisture content" > 5.4% and "press close time" > 56.5 seconds. The significant ($\alpha < 0.01$) linear regressors for this sub-tree are "bunker speed bottom surface layer" with a coefficient of 2.05 and "bunker speed top surface layer" with a coefficient of -3.41, i.e., for every unit increase in "bunker speed top surface layer" IB decreases by 3.41 p.s.i. with everything else in the equation held constant (Illustration 2B, Appendix B)

The subgroup with the lowest mean IB (38.2 p.s.i.) for this model occurs for "top core layer moisture content" ≤ 5.2%, "main spreader top surface layer speed" > 54.9, and "main flaker 3 pass counter" ≤ 249.7. The significant linear regressors for this sub-tree are "bunker speed top core layer" with a coefficient of -2.84 and "bunker speed top surface layer" with a coefficient of 2.03.

For a record length of 200, the best candidate model is mixed stepwise all possible subsets RT model without node pruning. Key metrics for this RT model are a RMSE of 7.03 p.s.i., RMSEP of 7.44 p.s.i., tree model R^2 of 0.73 and a homogeneous residual pattern. The RT model consists of a total of 33 nodes with 17 terminal nodes (Illustration 3B, Appendix B). Predicted IB did not approximate the general trend of observed IB for the validation data set with some central tendency of predictions (Figures 5.31 and 5.32). Development of a quantile RT model increases the RMSE to 7.53 p.s.i. and the RMSEP to 8.12 p.s.i. and did not improve predictability.

The subgroup with the highest mean IB (63.6 p.s.i.) for this model occurs for "press KI700A close 1 time" ≤ 16.5, "top core layer moisture content" > 4.9%, "main dryer bin surface layer level" ≤ 62.5, and "main blender bottom surface layer wood flow" ≤ 25954. There is no significant linear equation within this sub-tree.

The subgroup with the lowest mean IB (34.2 p.s.i.) for this model occurs for "press KI700A close 1 time" > 16.5, "press PC 741A" ≤ 97.0, "main flaker speed as a percent" > 97.8, "bottom surface layer bunker speed" > 66.5 and "press KI700S full open time" ≤ 16.5. The two significant ($\alpha <$ 0.01) linear regressors within this sub-tree are "bunker speed bottom core layer" with a coefficient of -2.61 and "bunker speed bottom surface layer" with a coefficient of -6.19 (Illustration 3B, Appendix B).

For a record length of 300, the best possible candidate model is again a mixed stepwise all possible subsets RT model without node pruning (Tables 7c and 8c, Appendix C). Key metrics for this RT model are a RMSE of 5.07 p.s.i., RMSEP of 13.35 p.s.i., tree model R^2 of 0.69 and a homogeneous residual pattern. The RT model consists of a total of 39 nodes with 20 terminal nodes. The predictions for this RT model do not

approximate observed IB in the validation data set (Figures 5.33 and 5.34). Development of a quantile RT model of the median IB increases the RMSE to 7.53 p.s.i. and reduces the RMSEP to 8.12 p.s.i. and does not approximate IB in the validation data set.

The subgroup with the highest level of mean IB (55.1 p.s.i.) for this model occurs for "main dry bin core level" > 37.5, "dryer wet bin #2" > 24.5, "press ZI 701 Step 3" > 972.5, "press LI795" ≤ 15.1, "main blender top surface layer wood flow" > 35.5, "press PC 741B" > 747.0, and "press UY741B output" ≤ 97.0. There is no significant linear equation within this sub-tree.

The lowest subgroup with a mean IB of 38.2 p.s.i. occurs for "main dryer bin core level" ≤ 37.5, "dryer wet bin #2" > 24.5, "press ZI 701 step3" ≤ 972.5, and "press MI 780" > 1.1. The three significant ($\alpha < 0.01$) linear regressors for this sub-tree are "bunker speed top surface layer" with a coefficient of -0.18, "dryer 1 inlet temperature" with a coefficient of -0.015, and "dryer 1 outlet temperature" with a coefficient of 0.159.

A candidate model using the shorter record length of 59 (investigated for MLR in Chapter IV) is a second-order RT model without node pruning. The RT model has good explanatory value and with the exception of two points in the validation data set, moderately good predictive capabilities (Figures 5.35 and 5.36). The RT model has 15 total nodes and eight terminal nodes. The RT model has a RMSE of 2.30 p.s.i., RMSEP of 11.01 p.s.i. with a tree model R^2 of 0.94 and a homogeneous residual pattern (Table 7c, Appendix C).

Parallel EI. -- Twenty-four parametric and 12 non-parametric quantile RT models are explored for this OSB strength property (Tables 9c and 10c, Appendix C). **For a record length of 100**, the best candidate model is a third-order RT model without node pruning.

Key metrics for this RT model are a RMSE of 920 in-lb^2/ft, RMSEP of 4361 in-lb^2/ft, tree model R^2 of 0.94 and a homogeneous residual pattern. The RT model consists of a total of 21 nodes with 11 terminal nodes (Illustration 4B, Appendix B). Predicted IB for the validation data set when lagged one time period approximated the time-ordered observed IB (Figures 5.37 and 5.38). Note that the Parallel EI is the only product in the study where lagging of predicted values improved predictions of IB. This may be an indication that the lagging of sensor data for Parallel EI is not accurate and requires further investigation in the fusion database.

The subgroup with the highest level of mean Parallel EI (63,030 in-lb^2/ft) occurs for "press MI 747" ≤ 2370, "flaker 2 stroke speed" > 3.5, and "flaker 3 stroke speed" > 1.5. The significant ($\alpha < 0.01$) regressors for this sub-tree are:

$$\begin{aligned}\text{Parallel EI} = &-- 2.35 \times 10^6 + 1.37 \times 10^6 (\text{"Top core layer moisture content"}) \\ &-- 2.57 \times 10^5 (\text{"Top core layer moisture content"})^2 \\ &+ 1.59 \times 10^4 (\text{"Top core layer moisture content"})^3. \end{aligned} \quad [5.1]$$

The subgroup with the lowest mean Parallel EI (56,207 in-lb^2/ft) occurs for "press MI 747" > 2370", "main spreader top core level" > 54.5, and "main blender bottom surface layer wood flow" > 48.5. The significant ($\alpha < 0.01$) regressors for this sub-tree are:

$$\begin{aligned}\text{Parallel EI} = &-- 3.76 \times 10^7 + 4.41 \times 10^6 (\text{"Dryer wet bin \#2"}) \\ &-- 1.72 \times 10^5 (\text{"Dryer wet bin \#2"})^2 \\ &+ 2.24 \times 10^3 (\text{"Dryer wet bin \#2"})^3. \end{aligned} \quad [5.2]$$

For a record length of 200, the most suitable candidate model is a third-order RT model without node pruning (Table 9c, Appendix C). Key metrics for this RT model are a RMSE of 946 in-lb^2/ft, RMSEP of 5337 in-lb^2/ft, tree model R^2 of 0.94 and a

homogeneous residual pattern. The RT model consists of a total of 45 nodes with 23 terminal nodes. Predicted Parallel EI (lagged one time period) for the validation data set does not approximate the time-ordered records with underestimation of large Parallel EI values (Figures 5.39 and 5.40). Note predictions of Parallel EI are worse when not time-lagged.

The highest level of mean Parallel EI (63,696 in-lb^2/ft) occurs for "core layer moisture" > 6.3%, "press main pressure" ≤ 3387.5 and "press PC 741B" > 97.0. The significant ($\alpha < 0.01$) regressors for this sub-tree are:

$$\begin{aligned}\text{Parallel EI} = &-- 1.10 \times 10^7 + 3.70 \times 10^5 \text{ (“Press position time”)} \\ &-- 4.12 \times 10^3 \text{ (“Press position time”)}^2 \\ &+ 1.52 \times 10^1 \text{ (“Press position time”)}^3. \end{aligned} \quad [5.3]$$

The subgroup with the lowest mean Parallel EI (56,853 in-lb^2/ft) occurs for "core layer moisture" ≤ 6.3%, "press MI 734" ≤ 13.3, "main dryer 5 outlet temperature" > 250.7, "press PMI 740" ≤ 3600, "top surface layer moisture content" ≤ 6.7%, "press PI 700 B" ≤ 1141 and "main flaker 2 HMI" > 28.3. The significant ($\alpha < 0.01$) regressors for this sub-tree are:

$$\begin{aligned}\text{Parallel EI} = 1.49 \times 10^7 &-- 2.14 \times 10^5 \text{ (“Press button-to button time”)} \\ &+ 1.01 \times 10^3 \text{ (“Press button-to button time”)}^2 \\ &-- 1.53 \times 10^1 \text{ (“Press button-to button time”)}^3. \end{aligned} \quad [5.4]$$

For a record length of 300, the best predictive model is a piecewise simple linear RT model without node pruning. Key metrics for this RT model are a RMSE of 1360 in-lb^2/ft, RMSEP of 6679 in-lb^2/ft, tree model R^2 of 0.91 and residual pattern with some heteroscedasticity. The RT model is large and consists of a total of 97 nodes with 49

terminal nodes. Predicted IB for the validation data set did not fit the observed IB very well even after time lagging.

A candidate model using the shorter record length of 58 (investigated for MLR in Chapter IV) is a second-order RT model without node pruning. The RT model has good explanatory value and good predictive capabilities of Parallel EI when predictions are lagged one time period (Figures 5.41 and 5.42). The RT model has 15 total nodes and eight terminal nodes. The RT model has a RMSE of 2.30 in-lb^2/ft, RMSEP of 11.01 in-lb^2/ft with a tree model R^2 of 0.94 and a homogeneous residual pattern (Table 9c, Appendix C).

Chapter V Summary

The investigation of modeling the IB of MDF and OSB, and the Parallel EI of OSB reveals RT models of the industrial process that have strong explanatory value and in some instances good predictability of the validation data sets. A challenge of the research is to develop useful RT models from industrial data sets with low dimensionality. Of the 160 models investigated using GUIDE (Loh 2006), 15 models have strong explanatory value of IB and four of these 15 models have acceptable predictability in the validation data sets (Tables 5.7 and 5.8). RT Models from shorter record lengths outperformed RT models with longer record lengths. Recall Box's (1979) famous quote, "All models are wrong but some are useful."

A comparison of the RMSEP for the RT models and similar MLR models (Chapter IV) reveals that the RMSEP is smaller for three of the six RT models for MDF (Figure 5.43). For OSB, the RMSEP for RT models is smaller than the MLR models in only one of the four models. The RMSEP is not the only metric associated with a best

model criteria and in most cases, predictability of IB in the validation data set is superior in RT models relative to similar MLR models (e.g., OSB Parallel EI, Figures 4.45 and 5.41; OSB IB, Figures 4.42 and 5.35; MDF 0.500" Figures 4.11 and 5.10; MDF 0.750" Figures 4.30 and 5.27). Note that low RMSEP did not always have acceptable prediction along the scale of the validation data set.

The in-depth investigation of RT models for MDF and OSB may have practical value to wood composite manufacturers. The analyses reveal opportunities for product quality and safety improvements and cost savings, by identifying statistical sources of product strength variation that may otherwise go undetected using traditional MLR methods with assumptions of homogeneous data. For MDF, process variables related to overall pressing time, press position times and core fiber moisture are highly significant ($\alpha < 0.01$) in influencing IB variation (Table 5.2). Even though such variables are commonly cited in the wood composites literature (Maloney 1977, Suchsland and Woodson 1986), the detail interactions identified by the RT sub-trees are helpful for discovering possible unknown sources and levels of process variation. Surprisingly, the RT analysis identified a variable related to crew that is negatively correlated with the IB of MDF. The RT models identified operator induced variation and the need for additional training of operators and managers (Deming 1986, 1993).

Significant ($\alpha < 0.01$) process variables related to the IB and Parallel EI of OSB are top and bottom core layers moisture content (Table 5.6). Press position and overall press times also influence the IB variation of OSB but to a lesser extent. Forming related variables are highly significant for the Parallel EI of OSB.

Appendix to Chapter V

Table 5.1. Description of training and validation data sets for longest record lengths.

Product	Type	Data	n	Time Length	Months
MDF	0.625"	Training	400	10/27/05 to 8/6/06	9 months
		Validation	80	8/6/06 to 10/5/06	2 months
MDF	0.750"	Training	200	10/28/05 to 8/21/06	10 months
		Validation	40	8/21/06 to 9/30/06	1 month
MDF	0.500"	Training	175	11/4/05 to 8/10/06	9 months
		Validation	33	8/10/06 to 10/23/06	2 months
OSB IB	7/16" RS	Training	300	6/1/06 to 10/13/06	4.5 months
		Validation	60	10/13/06 to 11/10/06	1 month
OSB Parallel EI	7/16" RS	Training	300	6/1/06 to 10/13/06	4.5 months
		Validation	60	10/13/06 to 11/10/06	1 month

Table 5.2. Ten most important independent variables for MDF using GUIDE scoring.

0.750" Thickness (n=200)		0.625" Thickness (n=400)		0.500" Thickness (n=175)	
Score	Description	Score	Description	Score	Description
100.00	Waste fiber addition to refined core fiber	100.00	Face refiner resin percentage set-point	100.00	Mat weight at the Thayer scales
92.52	Press overall time set-point	97.76	Swing refiner resin percentage set-point	87.22	Boiler temperature
91.18	Dry fuel bin #237 speed	93.39	Swing refiner plate position	81.41	Press final position time set-point
90.62	Dry fuel bin #236 speed	91.57	Press Pre-position time set-point*	81.41	Press final position time actual
77.87	Electrostatic precipitator Milliamps	89.99	Crew	75.86	Press pre-position time set-point
76.62	Core refiner plate pressure	89.90	Face Refiner Plate Position	72.79	Electrostatic precipitator Milliamps
76.41	Face refiner steam temperature	84.69	Face Bin Speed, i.e., time in forming bin	72.33	Core fiber digester pressure
74.92	Press total cycle time	84.52	Press pre-position thickness set-point	72.01	Press pre-position thickness set-point
74.85	Core refiner auger feeder screw speed	83.33	Percentage wood chips	71.96	Bar #2 shave off of fiber mat at formers
74.17	Core fiber moisture	82.18	Press final position time actual	70.78	Core fiber moisture

*Process variables highlighted in blue are common across all product types.

Table 5.3. Mixed stepwise regression equation with v-fold cross-validation node pruning for 0.500" MDF, n=175.

	GUIDE Regressor	Description	Coefficient	t	p-value
1	Constant		-1098.30	-7.34	< 0.0001
2	bFaceBlwPs	Face refiner blow valve position	0.78	4.44	< 0.0001
3	bSwgDigPrs	Swing refiner digester pressure	-1.08	-2.96	0.005
4	cCI0023PT	Refiner steam pressure	0.64	3.96	< 0.0001
5	cSwgOutlet	Swing Separator outlet	0.00	-3.32	0.0005
6	cSwFbrMst	Swing finer moisture	2.24	2.45	0.01
7	cSwgTemp	Swing refiner hot gas temp.	-0.13	-3.95	< 0.0001
8	dCoreRsnS	Core fiber resin set-point	17.32	7.35	< 0.0001
9	fCoreBtmSp	Core fiber forming speed	-0.28	-2.53	0.01
10	fShavOffT2	Mat shave off height #2	-11.09	-3.23	0.001
11	hPrCls3Tim	Press close #3 time	-1.66	-4.17	< 0.0001
12	hPrAlTimeS	Press overall time set-point	1.31	9.76	< 0.0001
13	hPrTempP	Press temperature	0.53	2.57	0.005
14	hPrTempS	Press temperature set-point	-13.71	-3.26	0.001
15	bFaceTempP	Face dryer outlet temperature	1.18	6.01	< 0.0001
16	bFaceH202W	Face refiner water add percent	-4.82	-8.89	< 0.0001
17	cSwngChpL	Swing chip chute level	1.25	4.51	< 0.0001
18	dCoreEFCur	Core dryer fan current	9.01	7.12	< 0.0001

Table 5.4. RT model mixed stepwise regression equation for 0.625" MDF, n=100.

Regressor	Coefficient	t	p-value[18]
Constant	1301.00	4.66	< 0.0001
Face Fiber Plate Position	-0.019	-3.68	0.0001
Face Resin to Wood Percentage	34.86	6.71	< 0.0001
Face Scavenger to Wood Percentage	-160.03	-3.59	0.0001
Swing Digester Pressure	-3.26	-7.00	< 0.0001
Core Resin to Wood Percentage	-18.52	-5.22	< 0.0001
Face Fiber Moisture Content	8.89	2.83	0.005

[18] Note that GUIDE does not give p-values. P-values are derived from Milton and Arnold (2003) and represent the minimum level of significance for a given t statistic using Table V of the T distribution, p. 700.

Table 5.5. Key regressors in multiple linear quantile RT model for 0.625" MDF, n=300

GUIDE Variable	Industrial SQL Tag	Description	Quantile coefficient
hPrPPXSP	hPressPrePostitionX_SP	Press pre-position set-point	-581.83
gPreBBSpd	gPrecompBottomBeltSpeed_SI	Pre-compressor belt speed	99.37
gFrmSpd	gFormingWireSpeed_SI	Forming line speed	-99.34
cSwgWx2W	cSwingWaxToWood_Act	Swing wax percent	88.34
bFacePltPs	bFacePlatePosition_ZI	Face refiner plate position	-25.28
dCoreRsnS	dCoreResin_SP	Core resin percent set-point	-16.61
bFaceWFCur	bFaceWestFanCurrent_PV	Face dryer fan current	-14.85
hPrPPMTimS	hPressPPMoveTime_SP	Press pre-position move time SP	-12.19
cSwgRsn2Wd	cSwingResinToWood_Act	Swing refiner resin percent	10.49
dCoreRsn2W	dCoreResinToWood_Act	Core resin percent	8.75
hPrCls2Tim	hPressClosetwoTime_PV	Press Close #2 time	6.42

Table 5.6. Ten most important independent variables for OSB using GUIDE scoring.

Internal Bond (n=300)		**Parallel EI (n=300)**	
Score	Description	Score	Description
100.00	Top core layer moisture content	100.00	Bottom surface layer forming spreader arm speed
95.65	Bottom core layer moisture content	94.97	Top core layer moisture content
72.85	Press position time	92.50	Top surface layer moisture content
63.15	Press overall time step movement time	92.15	Former bottom surface layer forming speed
62.78	Press overall time	91.95	Bottom core layer moisture content
56.89	Bottom core layer density set-point	89.44	Main dryer top surface layer moisture content
54.94	Top core layer forming speed	79.99	Dryer #5 Outlet Temperature
53.80	Blender bottom surface layer wood total	79.57	Press step position 734
52.32	Total time of press open to press close	78.54	Press step position 770
70.78	Blender bottom surface layer resin total	75.43	Former top surface layer forming speed

Table 5.7. Candidate RT models for all MDF and OSB RT products, not including Chapter IV shorter record lengths.

Product	RT Model	n	Total Nodes	Terminal Nodes	RMSE	RMSEP	R^2	Explanatory Value	Predictive Capability
MDF 0.500"	Mixed Stepwise CV prune	175	1	1	7.89	16.26	0.72	Moderate	Moderate
MDF 0.500"	Second-order Quantile CV prune	100	5	3	8.63	32.56	--	Good	Poor
MDF 0.625"	Mixed Stepwise CV prune	400	3	2	7.88	23.66	0.61	Good	Poor
MDF 0.625"	MLR Quantile CV prune	300	1	1	9.31	16.33	--	Moderate	Moderate
MDF 0.625"	PW Simple Linear CV prune	200	3	2	11.16	10.44	0.29	Poor	Poor
MDF 0.625"	Mixed Stepwise CV prune	100	1	1	9.42	14.84	0.51	Moderate	Poor
MDF 0.750"	PW Simple Linear no pruning	200	75	38	2.97	20.03	0.96	Good	Moderate
MDF 0.750"	MLR Quantile no pruning	100	13	7	8.54	22.17	--	Good	Poor
OSB IB	Mixed Stepwise All Subsets no pruning	300	39	20	5.07	13.35	0.69	Good	Poor
OSB IB	Mixed Stepwise All Subsets no pruning	200	23	12	5.79	9.24	0.64	Good	Good
OSB IB	Mixed Stepwise All Subsets no pruning	100	33	17	7.03	7.44	0.73	Good	Moderate
Parallel EI	Piecewise Simple Linear	300	97	49	1360	6679	0.91	Good	Poor
Parallel EI	Second-order Model no pruning	200	45	23	946	5337	0.94	Good	Moderate
Parallel EI	Second-order Model no pruning	100	21	11	920	4361	0.94	Good	Good

Table 5.8. Candidate RT models for MDF and OSB products for Chapter IV shorter record lengths.

Product	RT Model	n	Total Nodes	Terminal Nodes	RMSE	RMSEP	R^2	Explanatory Value	Predictive Capability
MDF 0.500"	Second-order no pruning	60	15	8	2.16	15.74	0.96	Good	Moderate
MDF 0.500"	Second-order CV prune	60	3	2	8.01	13.08	0.50	Moderate	Poor
MDF 0.625"	Second-order no pruning	62	17	9	2.36	18.31	0.98	Good	Poor
MDF 0.625"	Second-order CV prune	62	3	2	9.55	11.29	0.60	Moderate	Good
MDF 0.750"	Second-order no pruning	70	23	12	2.01	30.10	0.98	Good	Poor
MDF 0.750"	Second-order CV prune	70	1	1	14.69	15.87	0.20	Poor	Poor
OSB IB	Second-order no pruning	59	15	8	2.30	11.01	0.94	Good	Moderate
OSB IB	Second-order CV prune	59	1	1	7.88	9.17	0.26	Poor	Poor
Parallel EI	Second-order Model no pruning	58	15	8	692	2479	0.96	Good	Good
Parallel EI	Second-order Model no pruning	58	1	1	3255	5988	0.10	Poor	Poor

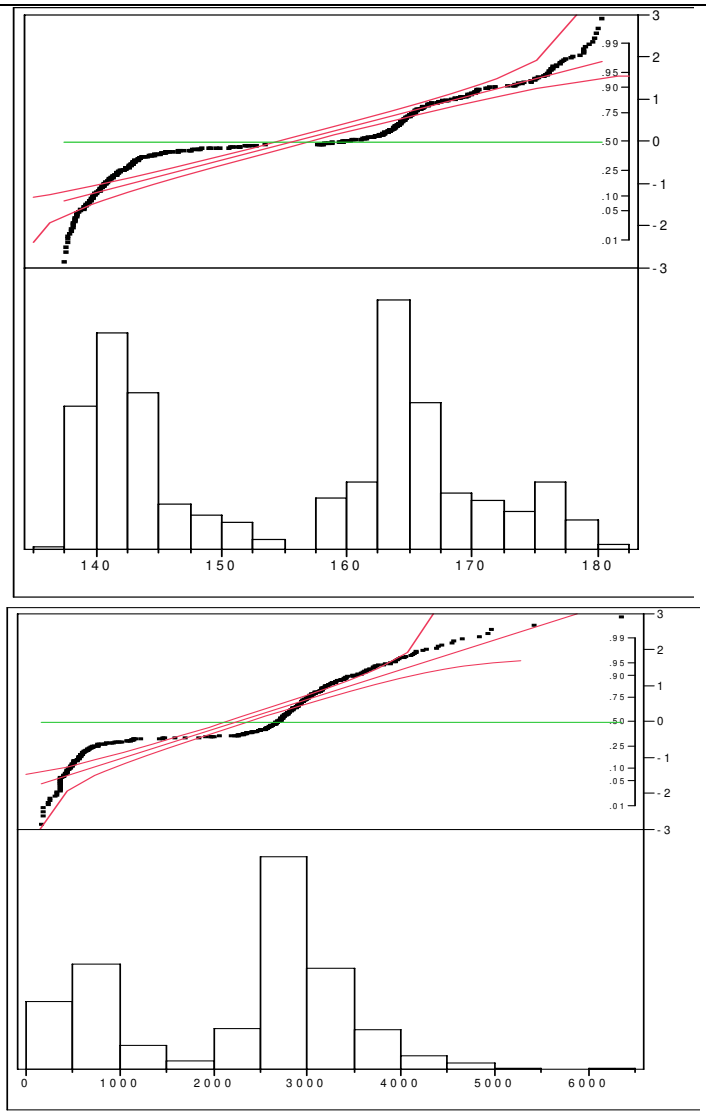


Figure 5.1. Histograms and quantile plots of "core fiber temperature" (top) and "swing refiner separator outlet pressure" (bottom) for 0.625" MDF, n=400.

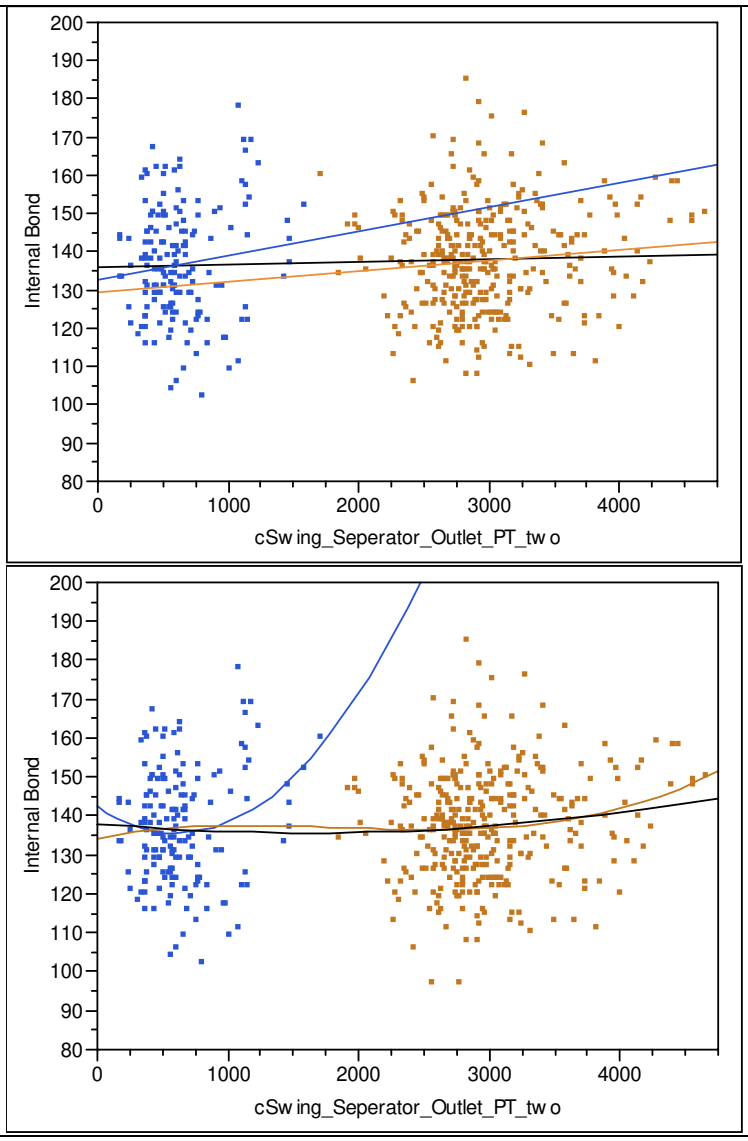


Figure 5.2. Linear and polynomial regression fits for IB to the sub-spaces of "swing refiner separator outlet pressure" for 0.625" MDF, n=400 (blue line fits the blue points; red line fits the red points; black line fits all of the data).

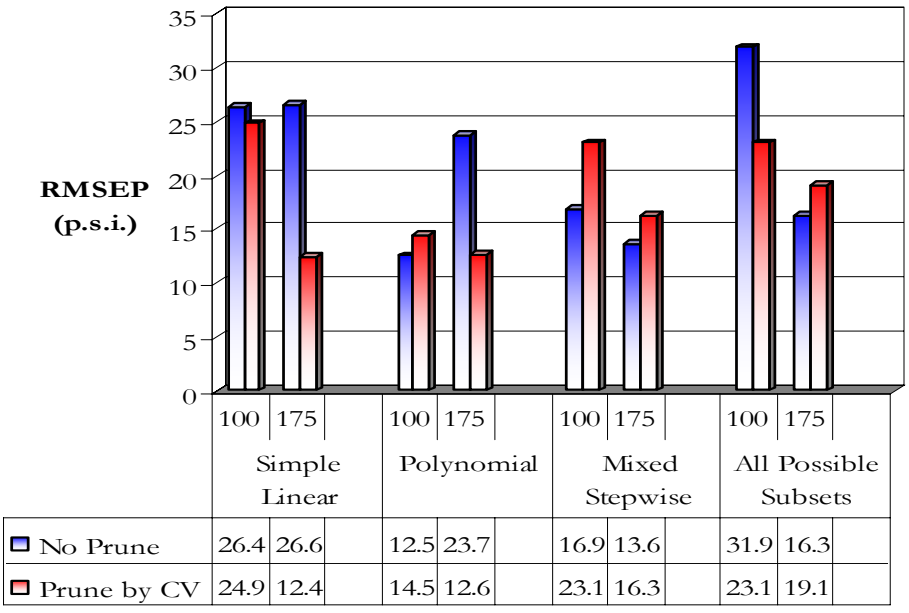


Figure 5.3. RMSEP by record length and modeling type for 0.500" MDF IB.

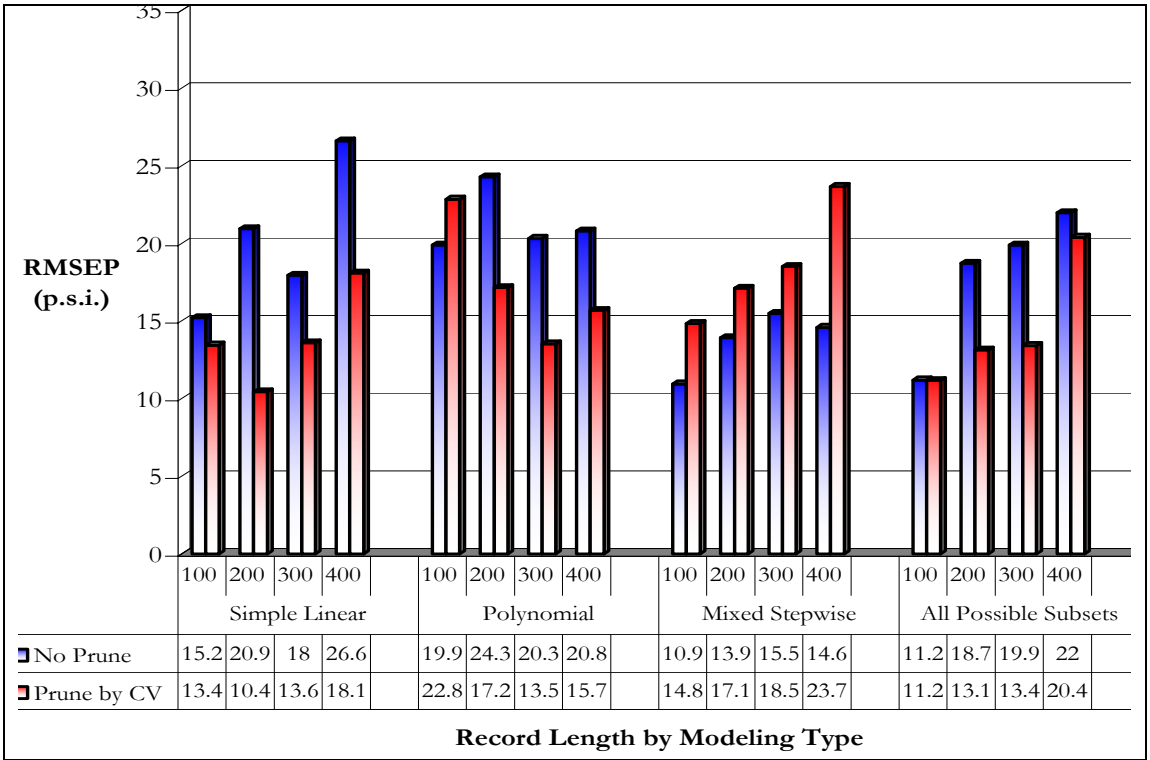


Figure 5.4. RMSEP by record length and modeling type for 0.625" MDF IB.

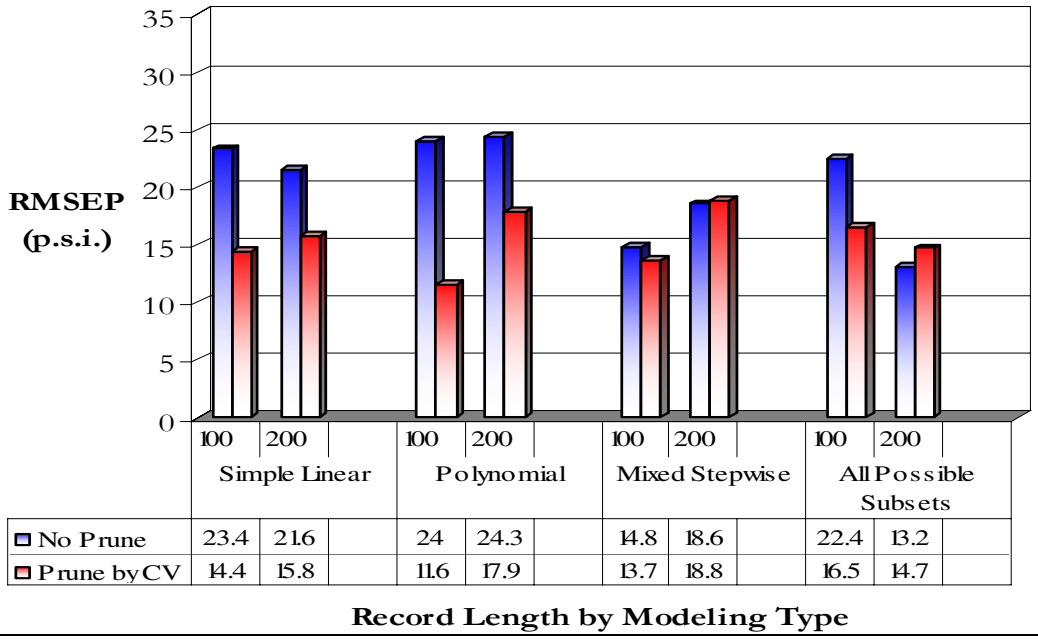


Figure 5.5. RMSEP by record length and modeling type for 0.750" MDF IB.

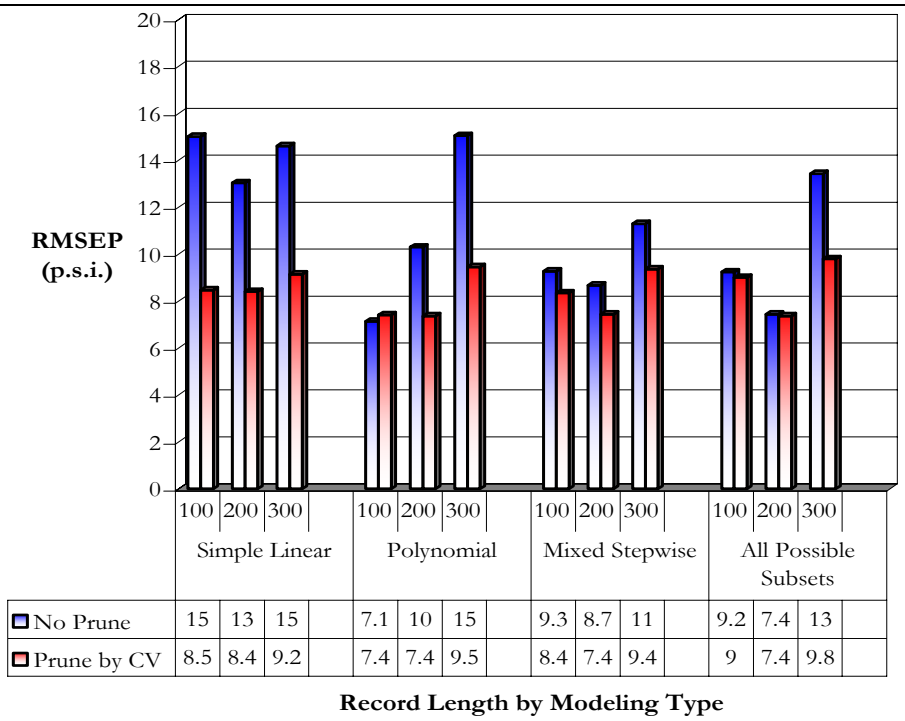


Figure 5.6. RMSEP by record length and modeling type for the IB of OSB.

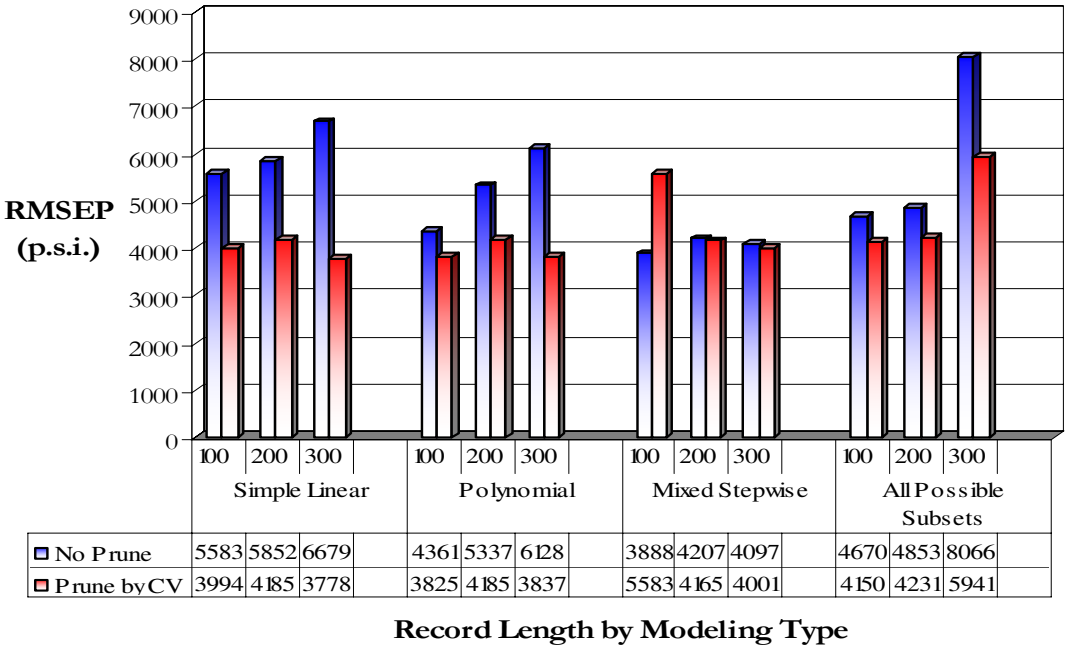
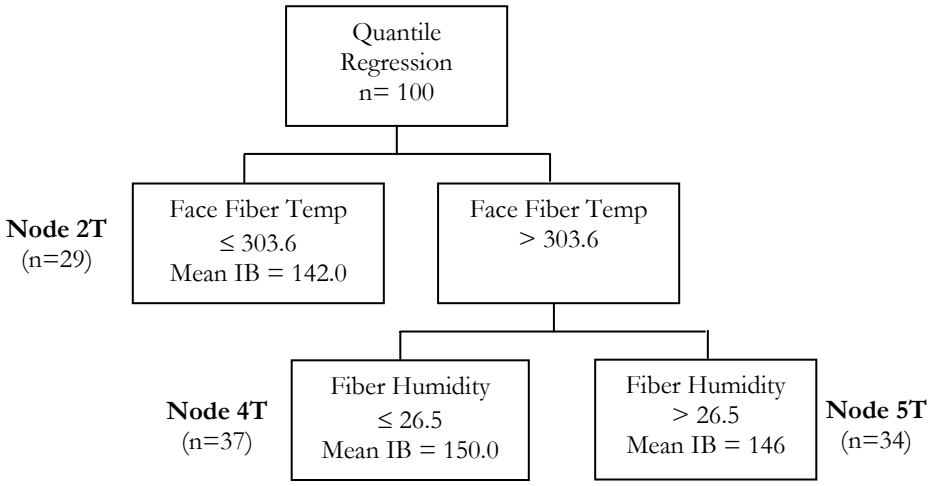


Figure 5.7. RMSEP by record length and modeling type for the Parallel EI of OSB.



Node 2T:

Quantile Regressor	Coefficient
Constant	158.01
Press pre-position time set-point	175.06
(Press pre-position time set-point)^2	-1077.90
(Press pre-position time set-point)^3	976.86

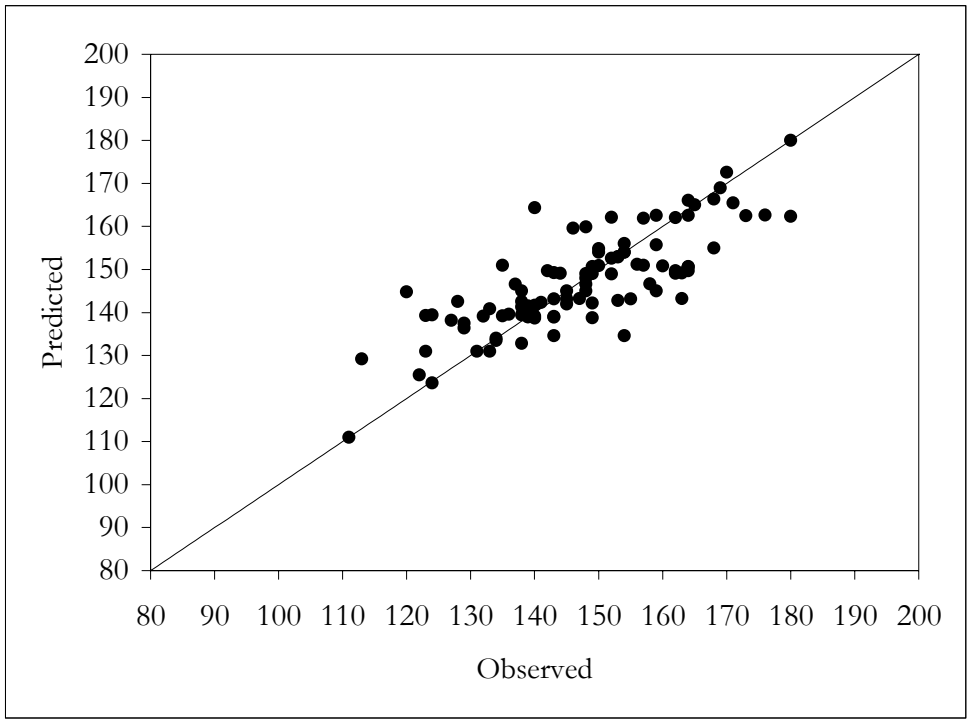
Node 4T:

Quantile Regressor	Coefficient
Constant	-1360.9
Swing refiner chip chute level	79.31
(Swing refiner chip chute level)^2	-1.42
(Swing refiner chip chute level)^3	0.009

Node 5T:

Quantile Regressor	Coefficient
Constant	167.88
Face refiner grinding pressure	-7.06
(Face refiner grinding pressure)^2	0.62
(Face refiner grinding pressure)^3	-0.018

Figure 5.8. Quantile third-order RT model with v-fold cross-validation for 0.500" MDF, n=100.



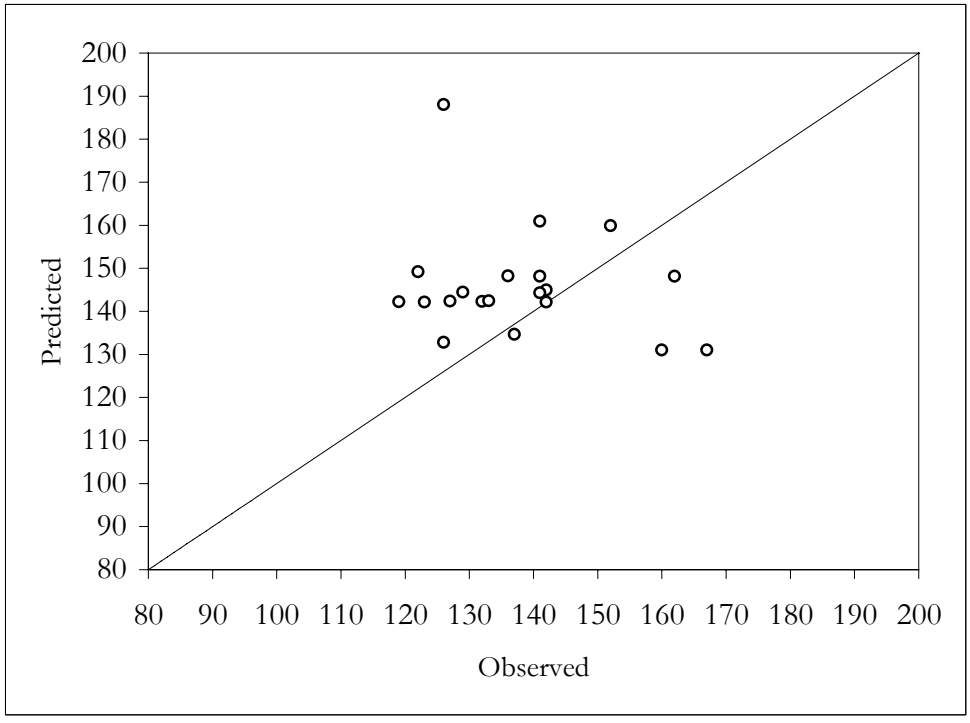
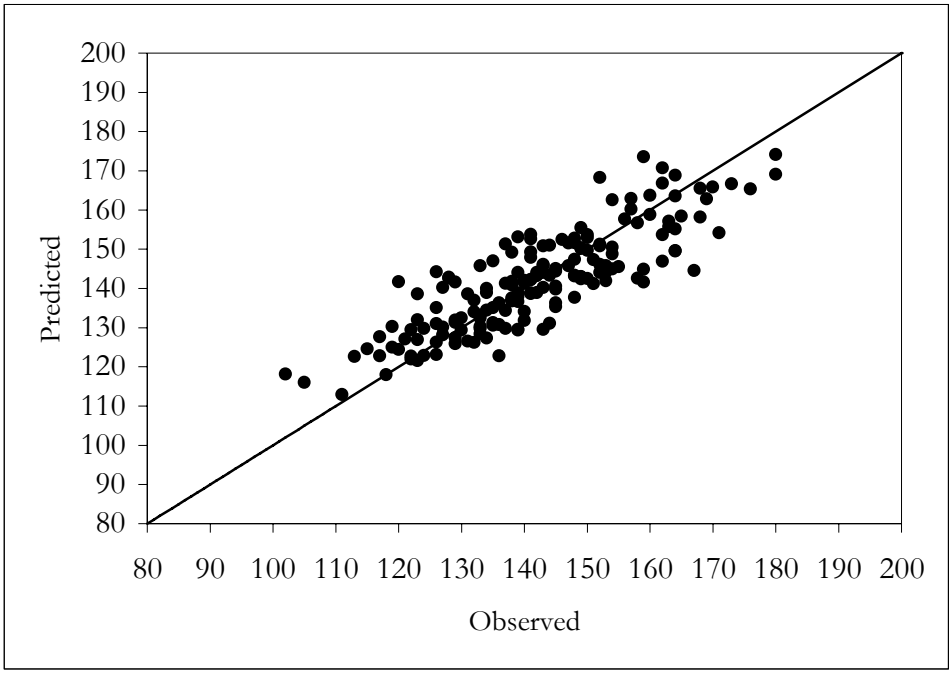


Figure 5.9. XY scatter plot of training (top) and validation data sets (bottom) for the third-order quantile regression RT model with v-fold cross-validation node pruning for 0.500" MDF, n=100.



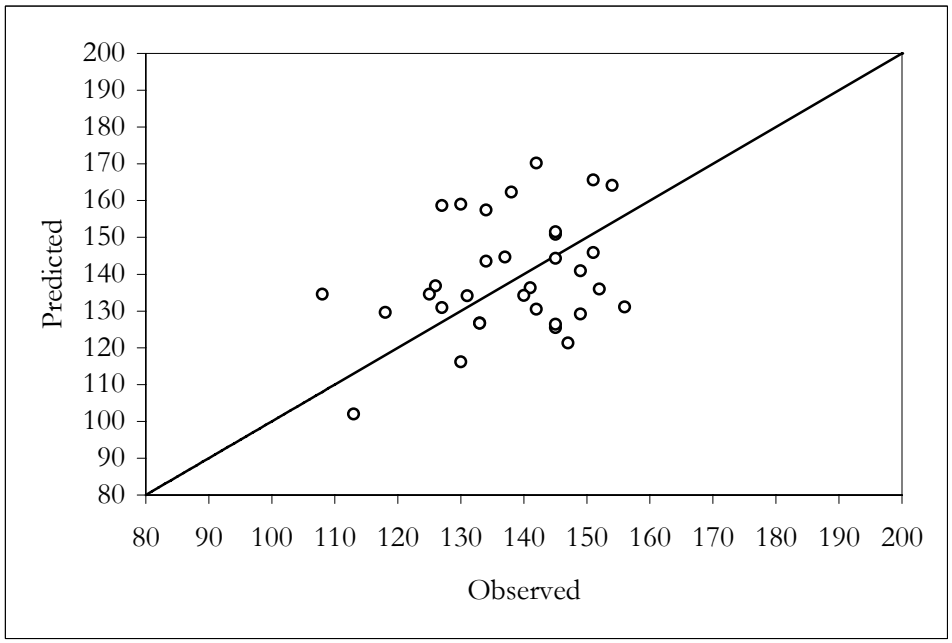


Figure 5.10. XY scatter plot of training (top) and validation data sets (bottom) for the stepwise regression RT model with v-fold cross-validation node pruning for 0.500" MDF, n=175.

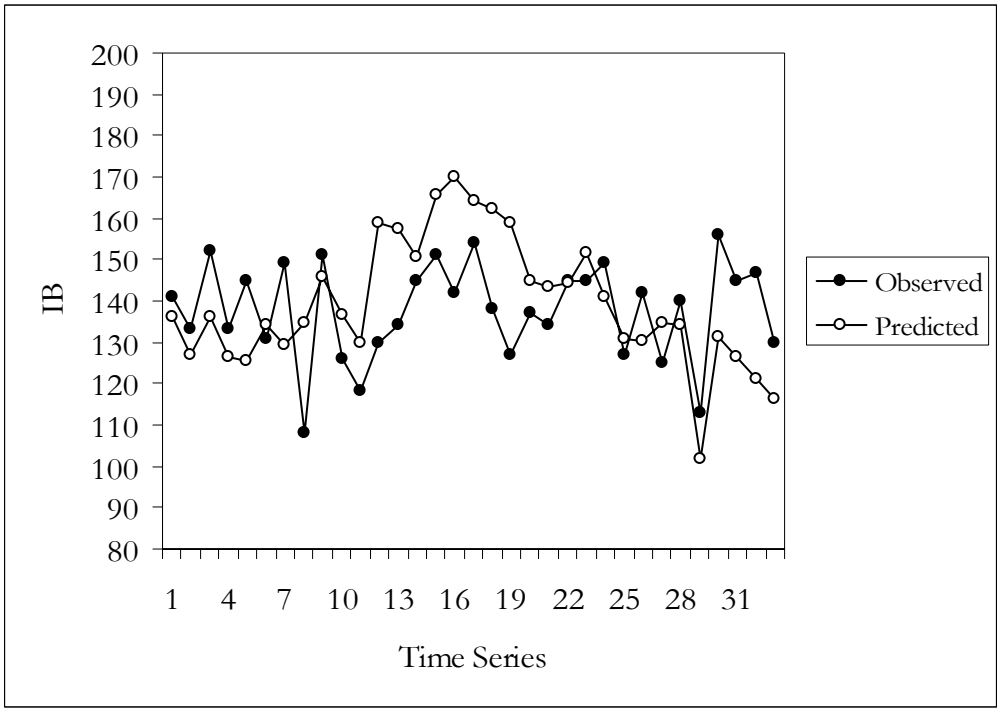


Figure 5.11. Time series graph of validation data set for the stepwise regression RT model with v-fold cross-validation node pruning for 0.500" MDF, n=35.

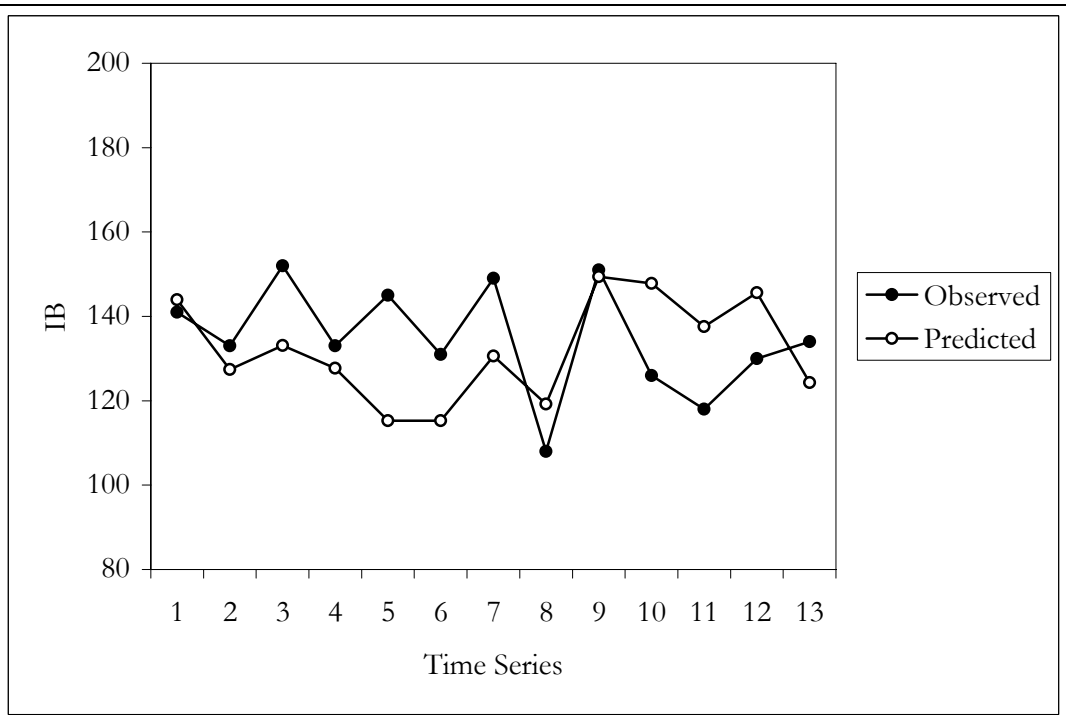


Figure 5.12. Time series graph of validation data set for the second-order RT model without node pruning for 0.500" MDF, n=13.

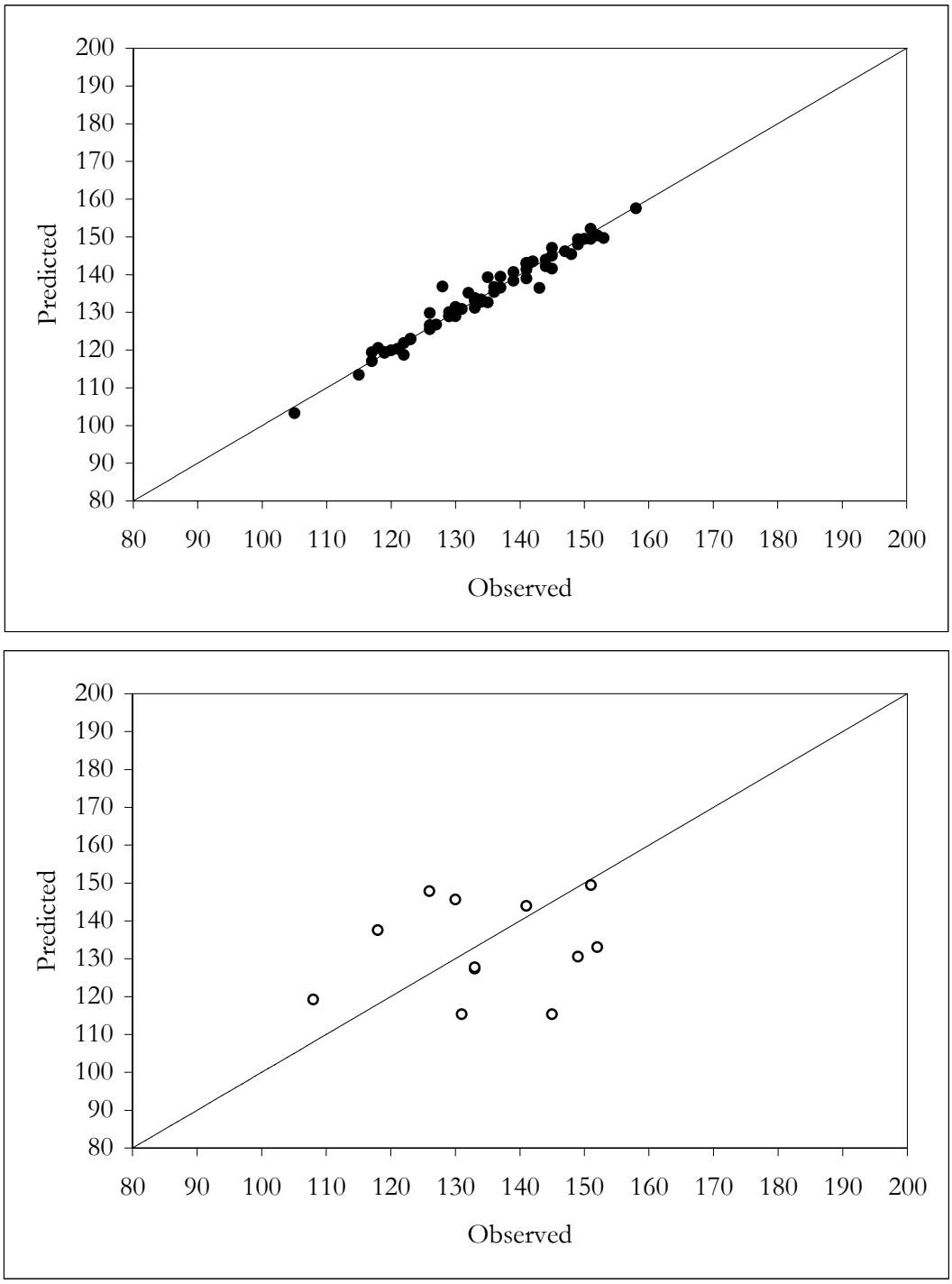
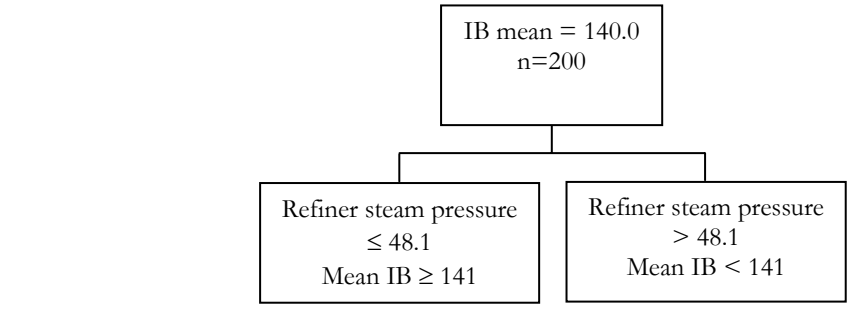


Figure 5.13. XY scatter plot of training (top) and validation data set (bottom) for the second-order RT model without node pruning for 0.500" MDF, n=60.



Regressor	Coefficient	t
Constant	232.75	17.79
Core fiber moisture	-9.24	-7.00
Tech JY	-6.29	-2.13

RMSE	12.72
n	68
R^2	0.43

Regressor	Coefficient	t
Constant	158.61	45.85
Press pre-position move time set-point	-70.33	-5.21
Tech JY	-9.58	-3.58

RMSE	11.45
n	132
R^2	0.20

Figure 5.14. RT and mixed stepwise regression equations for 0.625" MDF, n=200.

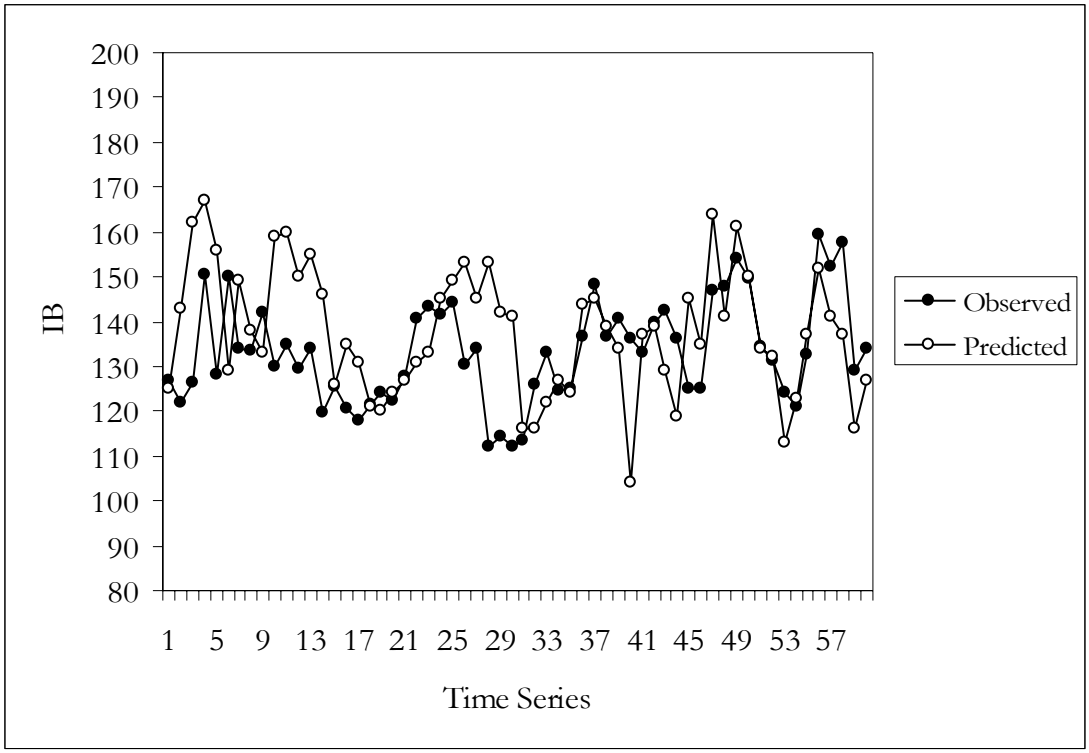


Figure 5.15. Time series graph of validation data set for the multiple linear quantile RT model with v-fold cross-validation node pruning for 0.625" MDF, n=300.

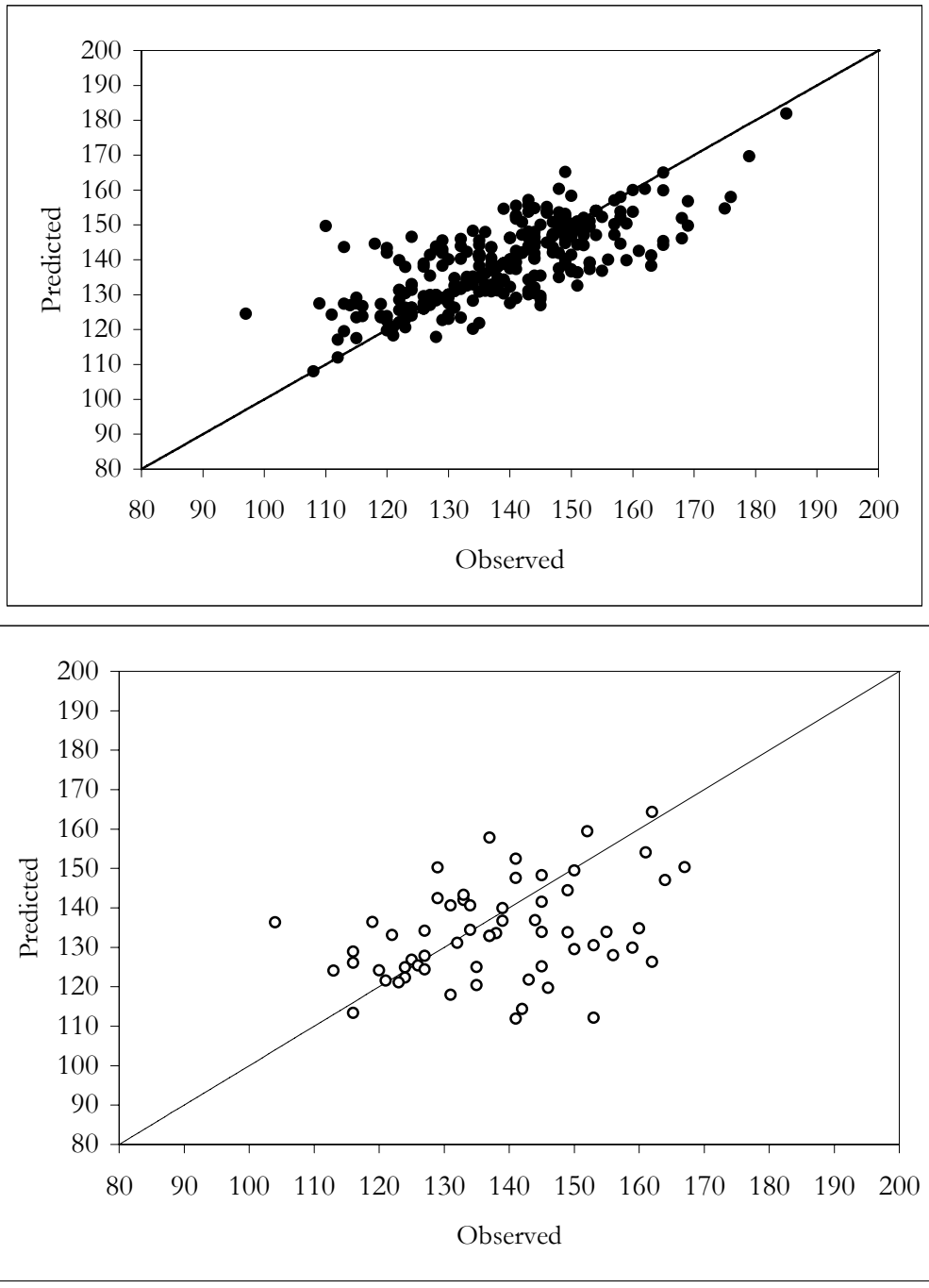
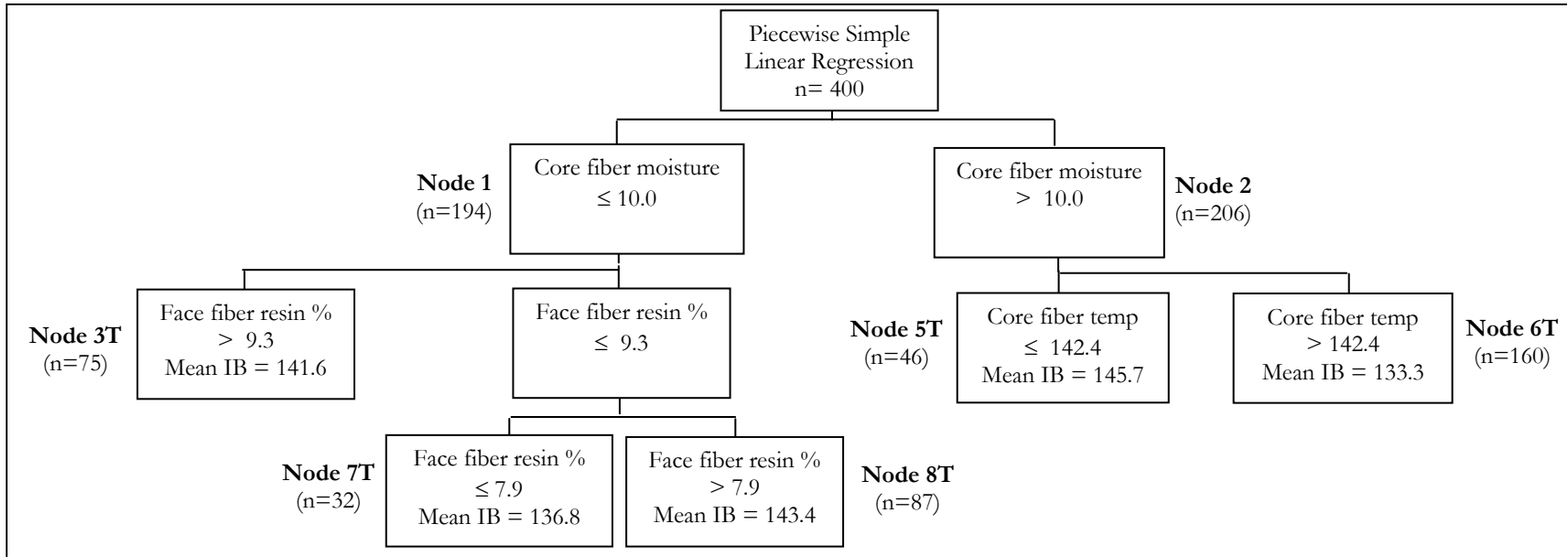


Figure 5.16. XY scatter plot of training (top) and validation data set (bottom) for the multiple linear quantile RT model with v-fold cross-validation node pruning for 0.625" MDF, n=300.



Node 3T:

Simple Linear Regression	Coefficient
Constant	-96.7
Press overall time set-point	1.20

Node 8T:

Simple Linear Regression	Coefficient
Constant	120.8
Core refiner feeder screw speed	0.036

Node 5T:

Simple Linear Regression	Coefficient
Constant	295.6
Core fiber temperature	-1.36

Node 6T:

Simple Linear Regression	Coefficient
Constant	118.5
Face refiner grinding steam flow	-0.005

Node 7T:

Simple Linear Regression	Coefficient
Constant	472.8
Swing digester pressure	-3.57

Note that all regressors in each terminal node are significant at an $\alpha < 0.01$

Figure 5.17. Piecewise simple linear model with v-fold cross-validation node pruning for 0.625" MDF, n=400.

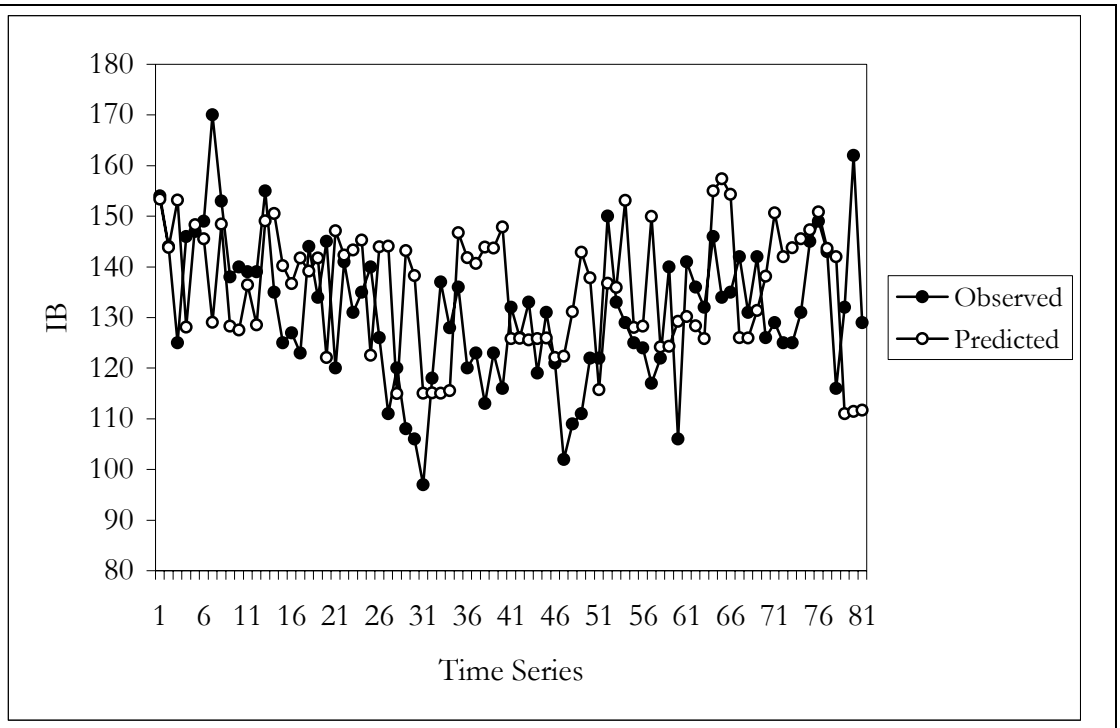


Figure 5.18. Time series graph of validation data set for the piecewise simple linear model with v-fold cross-validation node pruning for 0.625" MDF, n=400.

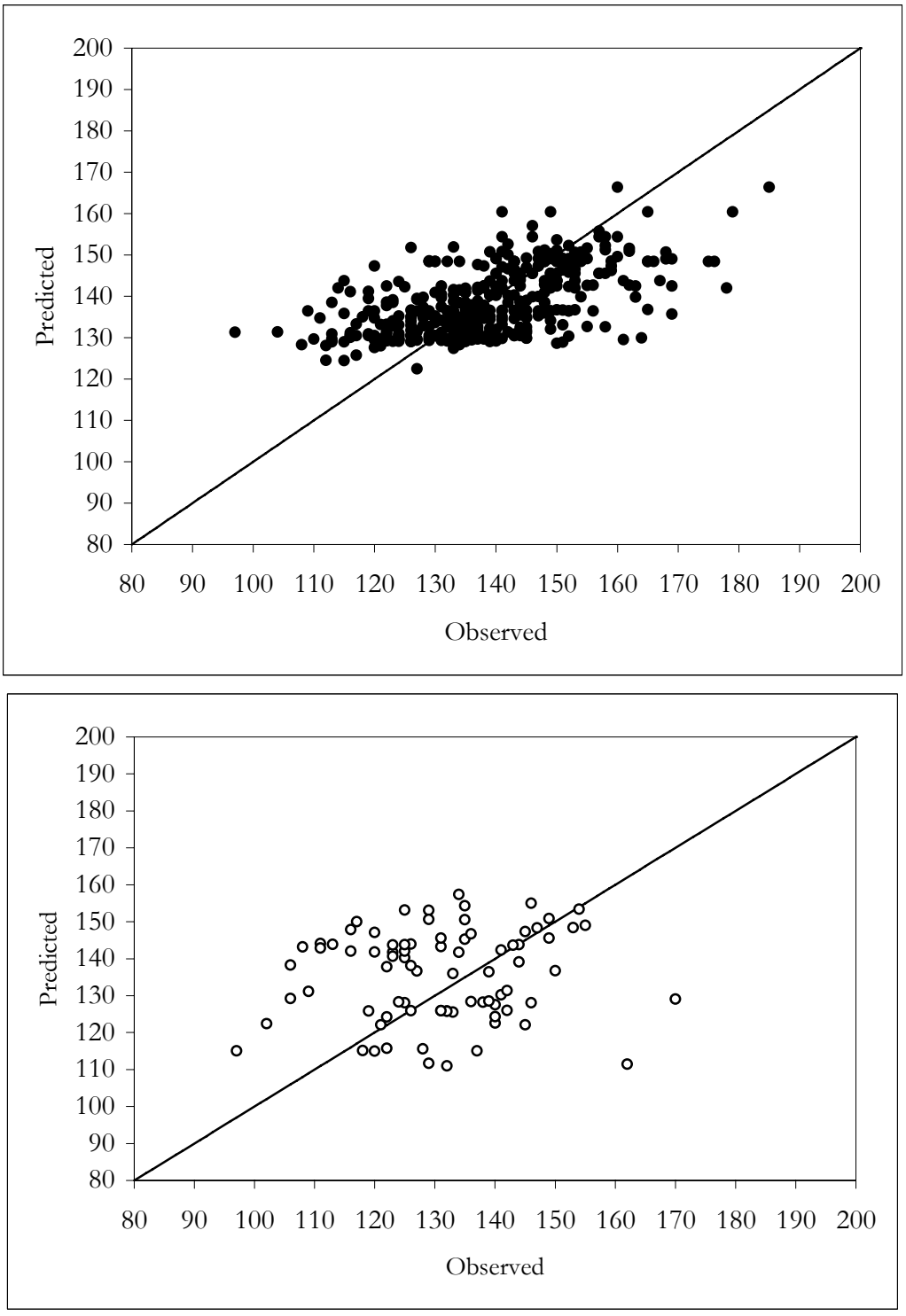


Figure 5.19. XY scatter plot of training data set (top) and validation data set (bottom) for the piecewise simple linear model with v-fold cross-validation node pruning for 0.625" MDF, n=400.

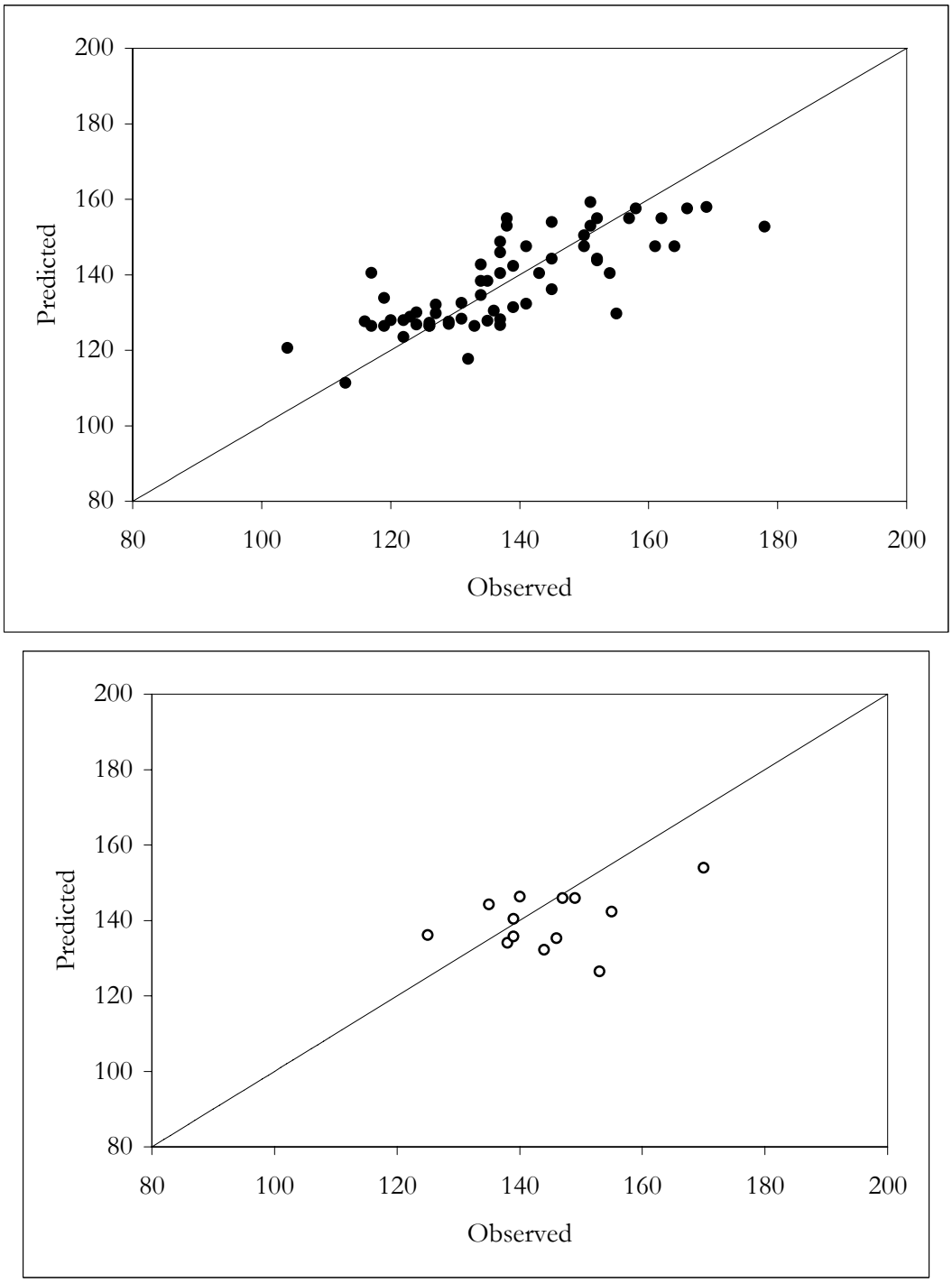


Figure 5.20. XY scatter plot of training data set (top) and validation data set (bottom) for the second-order RT model with v-fold cross-validation node pruning for 0.625" MDF, n=62.

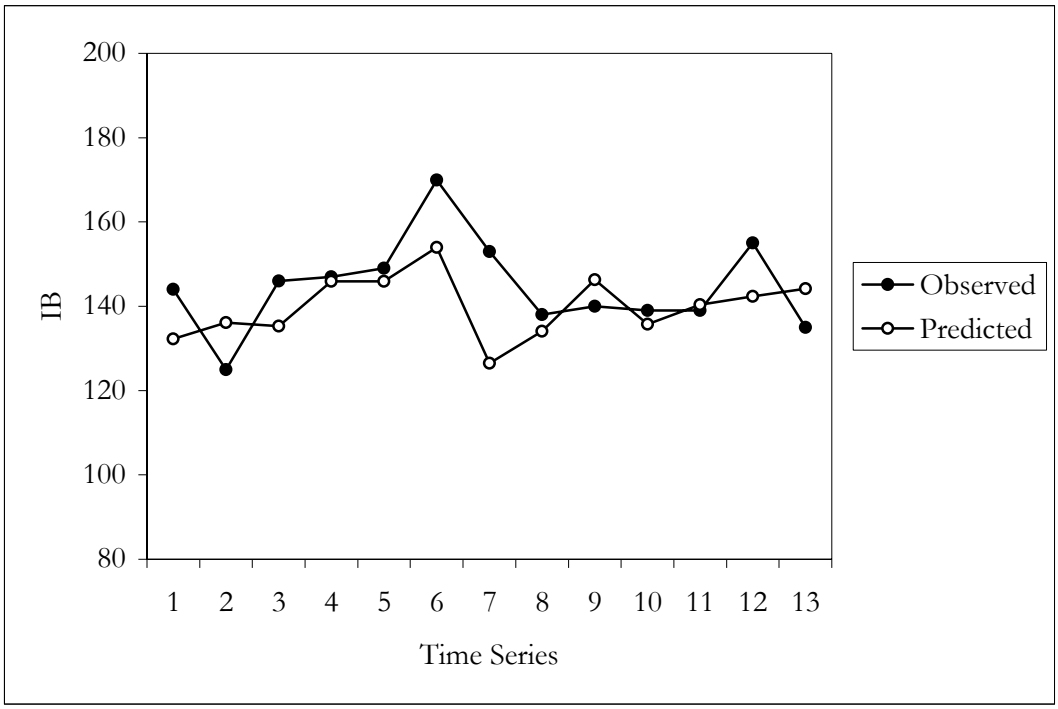


Figure 5.21. Time series graph of validation data set for the second-order RT model with v-fold cross-validation node pruning for 0.625" MDF, n=13.

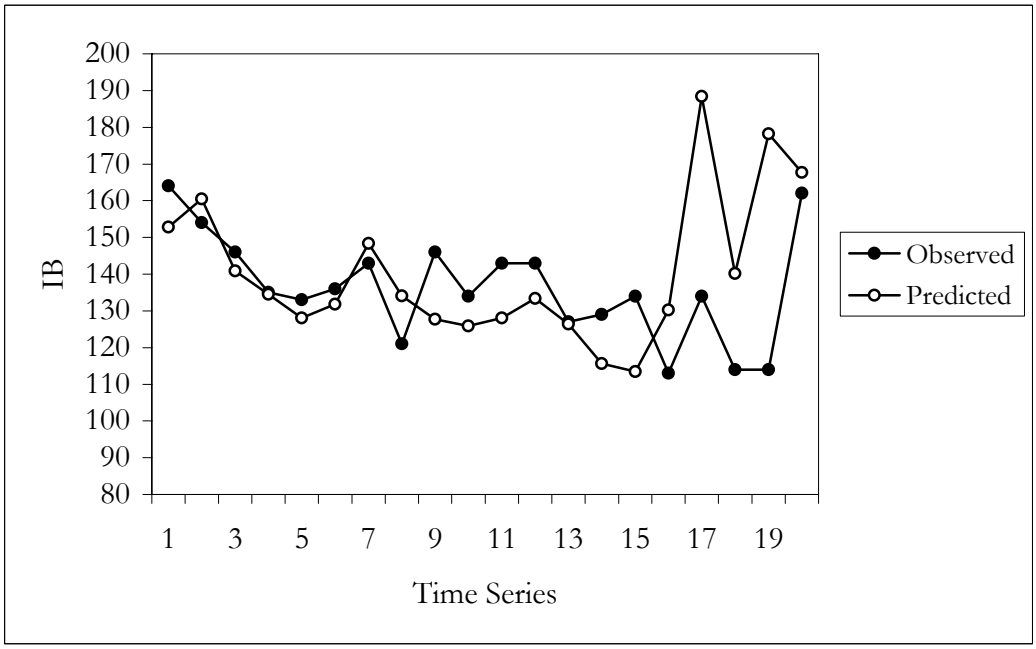
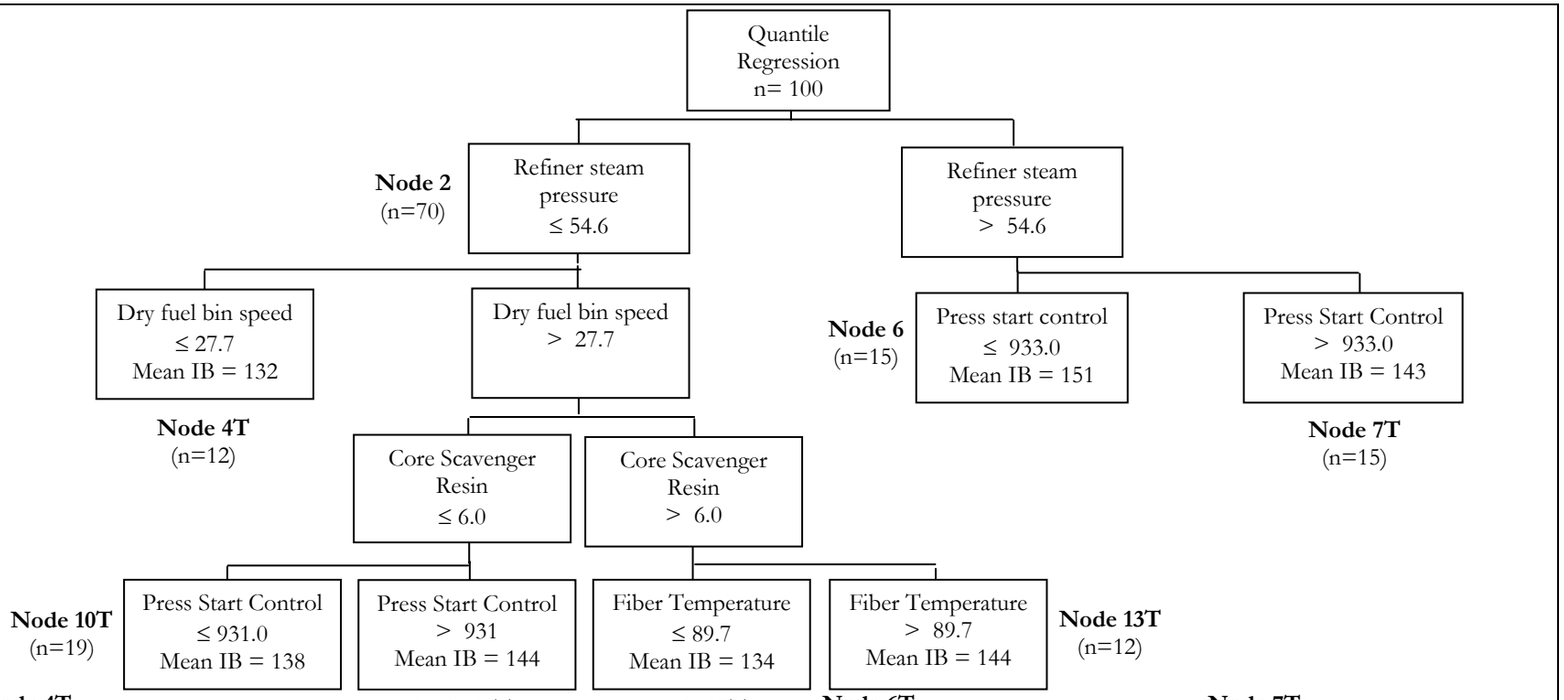


Figure 5.22. Time series graph of validation data set for the multiple linear quantile RT model without node pruning for 0.750" MDF, n=100.



Node 4T:

Quantile Regressor	Coefficient
Constant	22938.0
Refiner steam pressure	-2.61
Press start control	1.48
Core fiber scavenger resin rate	-4010.2
Dry fuel bin rate 236	-21.24
Dry fuel bin rate 237	21.54
Fiber temperature	-0.13

Node 6T:

Quantile Regressor	Coefficient
Constant	13750.0
Refiner steam pressure	-1.08
Press start control	-0.22
Core fiber scavenger resin rate	-2211.4
Dry fuel bin rate 236	-1.63
Fiber temperature	-0.21

Node 7T:

Quantile Regressor	Coefficient
Constant	-16418.0
Refiner steam pressure	1.39
Press start control	1.34
Core fiber scavenger resin rate	2561.8
Dry fuel bin rate 236	-1.02
Dry fuel bin rate 237	-5.78
Fiber temperature	-1.24

Figure 5.23. Multiple linear quantile RT model without node pruning for 0.750" MDF, n=100.

Node 10T:

Quantile Regressor	Coefficient
Constant	-2734.4
Refiner steam pressure	0.85
Press start control	0.84
Core fiber scavenger resin rate	34.27
Dry fuel bin rate 236	-2.63
Fiber temperature	0.86

Node 11T:

Quantile Regressor	Coefficient
Constant	4020.8
Refiner steam pressure	-0.27
Press start control	0.15
Core fiber scavenger resin rate	-667.37
Dry fuel bin rate 236	-154.25
eM2237Spd	154.49
Fiber temperature	-0.01

Node 12T:

Quantile Regressor	Coefficient
Constant	-41307.0
Refiner steam pressure	1.21
Press start control	-1.83
Core fiber scavenger resin rate	7094.60
Dry fuel bin rate 236	7.85
Fiber temperature	3.18

Node 13T:

Quantile Regressor	Coefficient
Constant	21874.0
Refiner steam pressure	-0.97
Press start control	-0.59
Core fiber scavenger resin rate	-3519.70
Dry fuel bin rate 236	0.89
Fiber temperature	-4.59

Figure 5.23 (continued). Multiple linear quantile RT model without node pruning for 0.750" MDF, n=100.

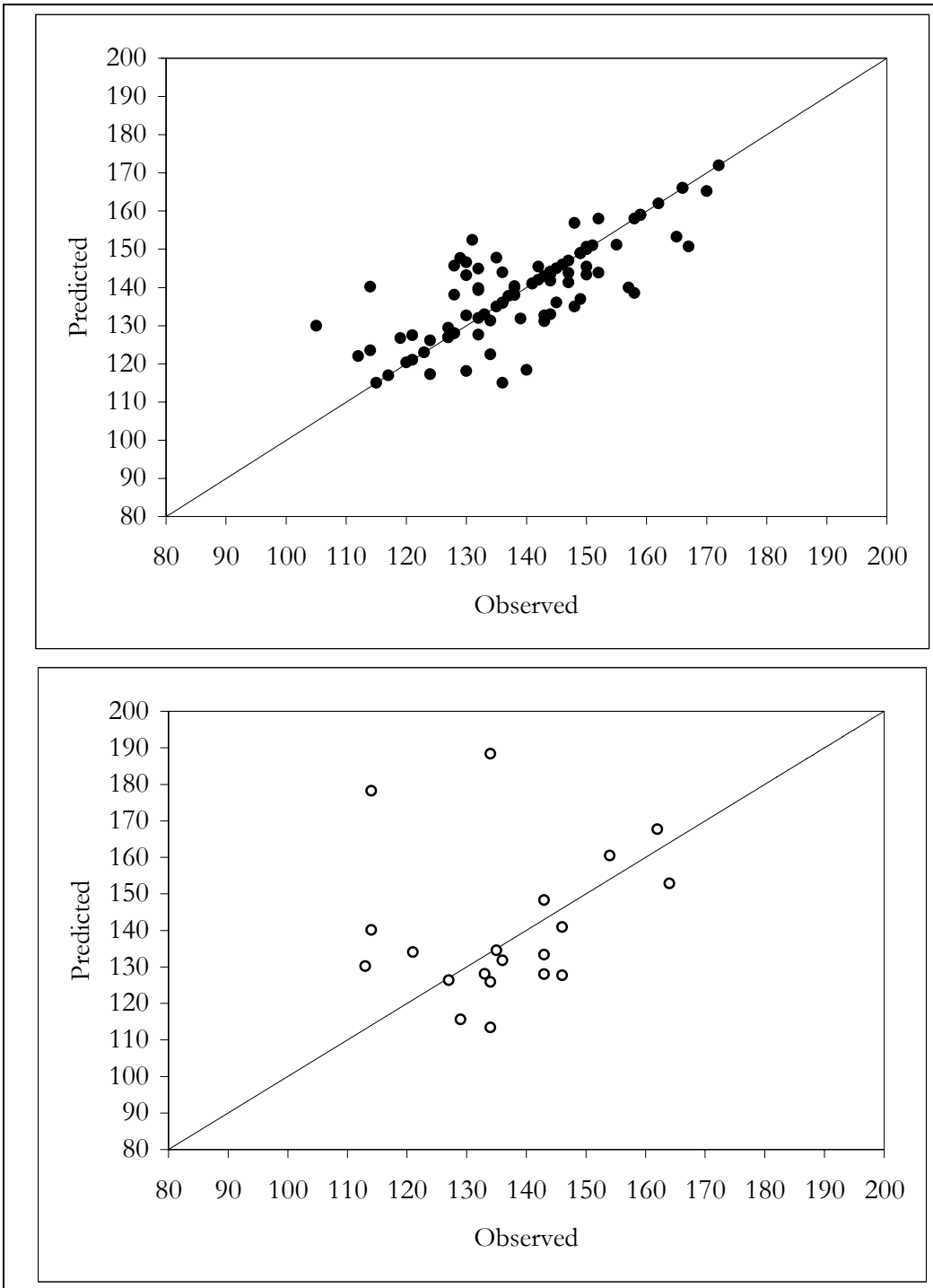


Figure 5.24. XY scatter plot of training (top) and validation data sets (bottom) for the multiple linear quantile RT model without node pruning for 0.750" MDF, n=100.

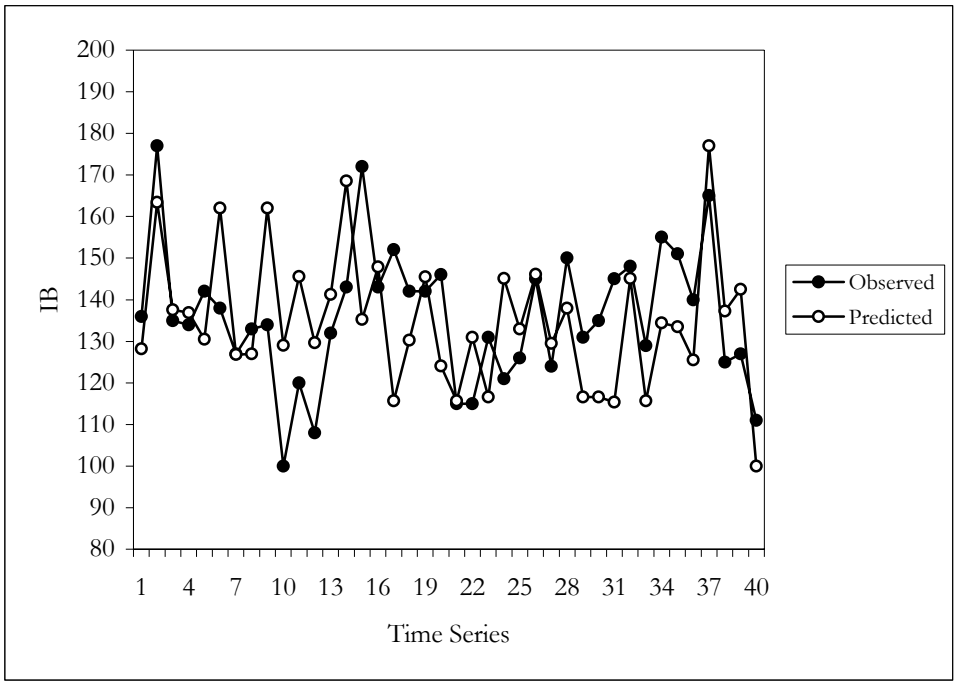


Figure 5.25. Time series graph of validation data set for the piecewise simple linear RT model without node pruning for 0.750" MDF, n=200.

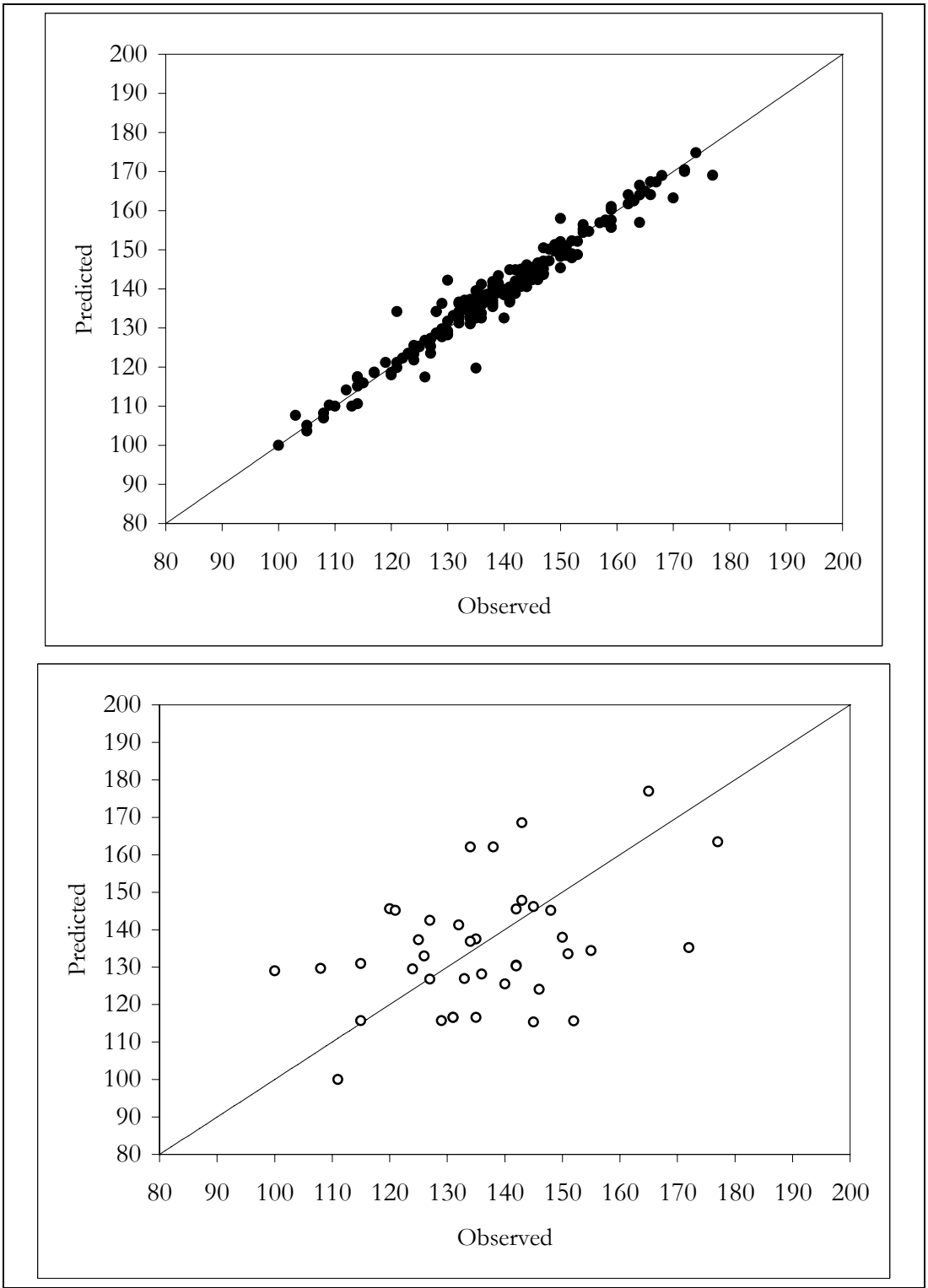


Figure 5.26. XY scatter plot of training (top) and validation data set (bottom) for the piecewise simple linear RT model without node pruning for 0.750" MDF, n=200.

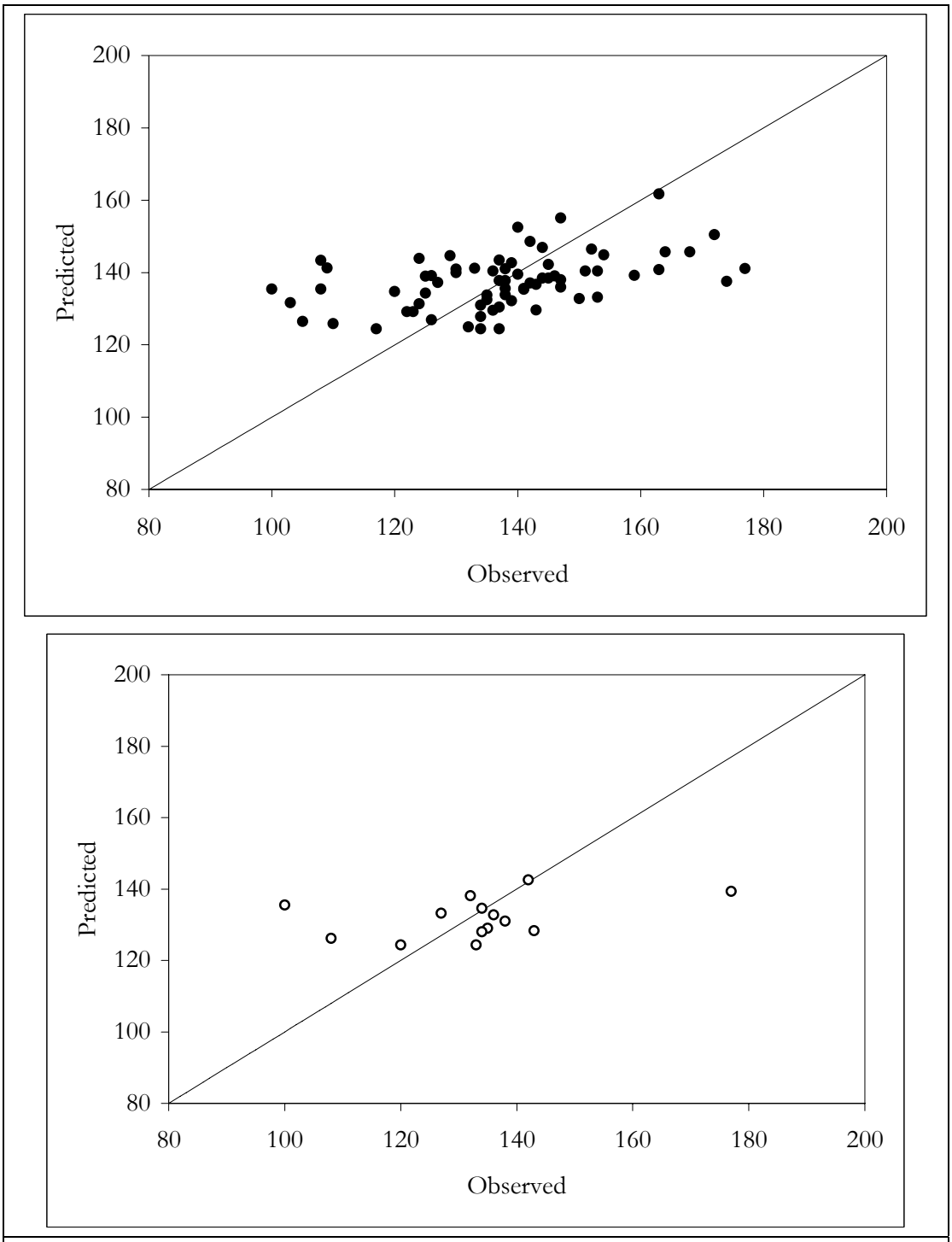


Figure 5.27. XY scatter plot of training (top) and validation data set (bottom) for the second-order RT model with v-fold cross-validation for 0.750" MDF, n=70.

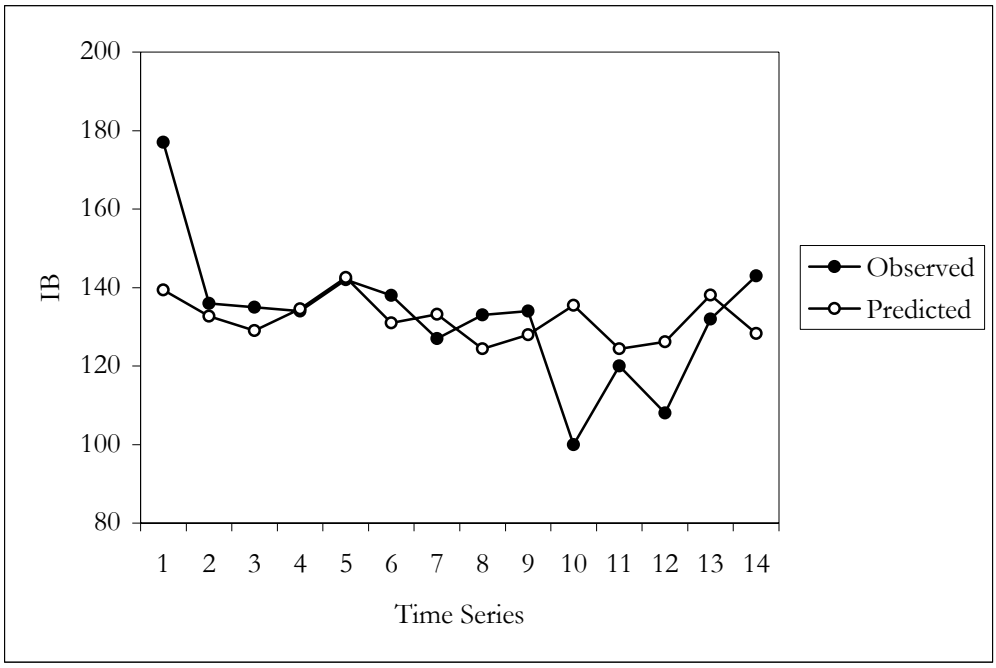


Figure 5.28. Time series graph of validation data set for the second-order RT model with v-fold cross-validation for 0.750" MDF, n=70.

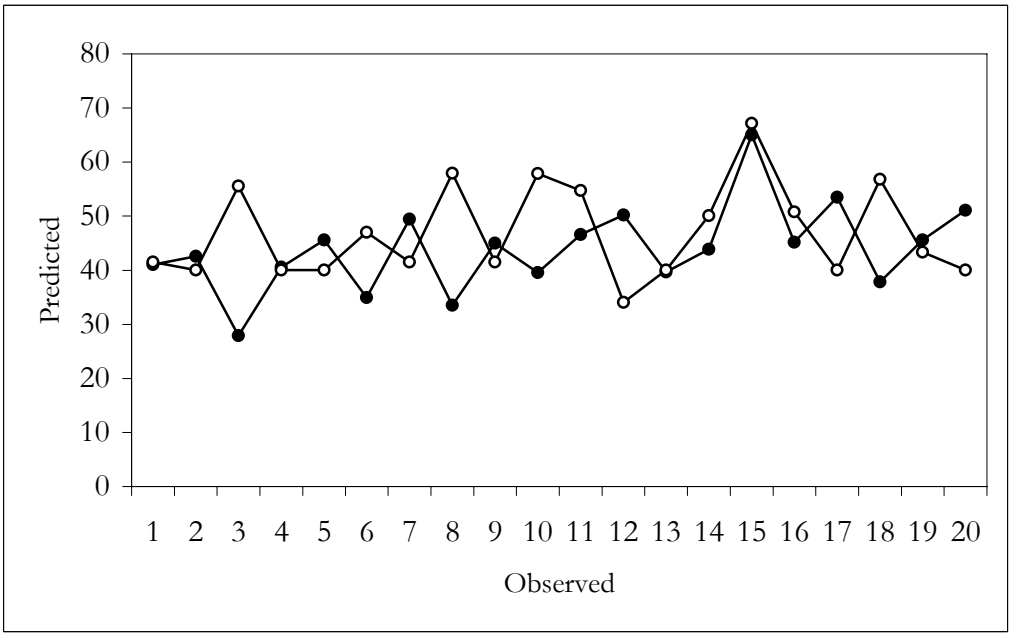


Figure 5.29. Time series graph of validation data set for the mixed stepwise all possible subsets RT model without node pruning for the IB of OSB, n=100.

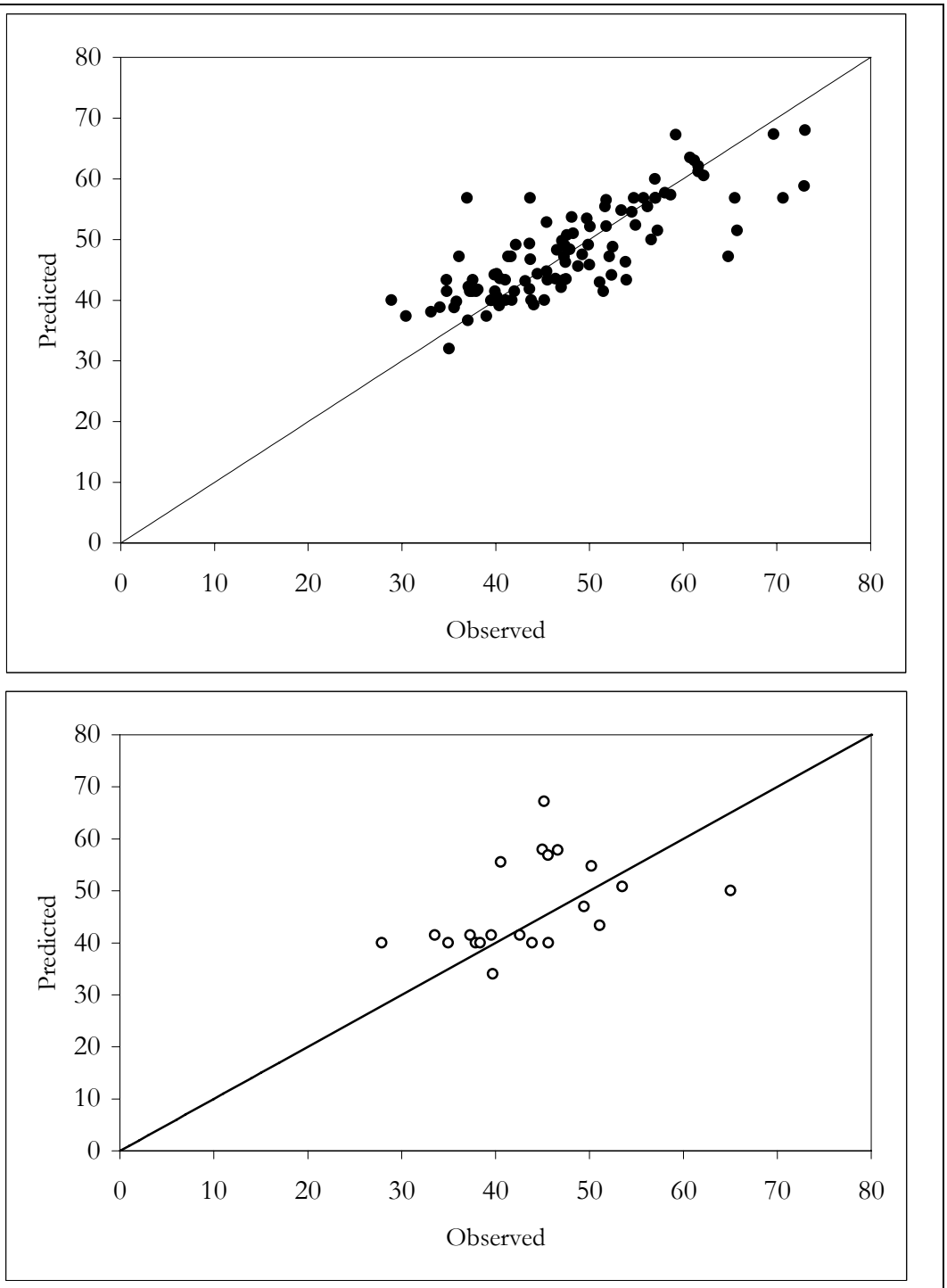


Figure 5.30. XY scatter plot of training (top) and validation data sets (bottom) for the mixed stepwise all possible subsets RT model without node pruning for the IB of OSB, n=100.

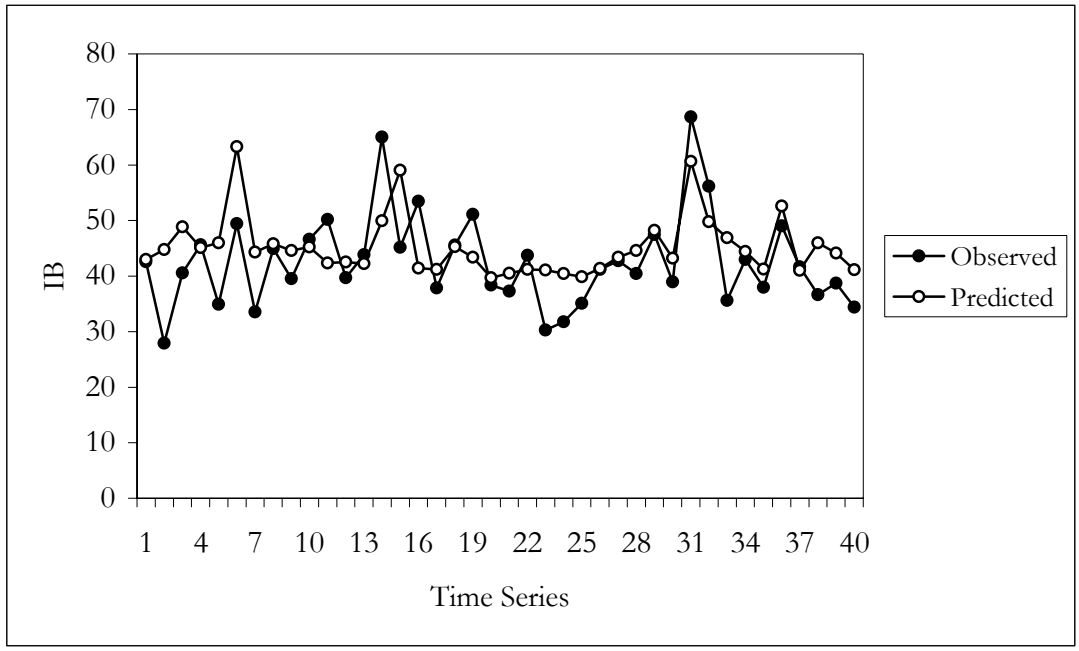


Figure 5.31. Time series graph of validation data set for the mixed stepwise all possible subsets RT model without node pruning for the IB of OSB, n=200.

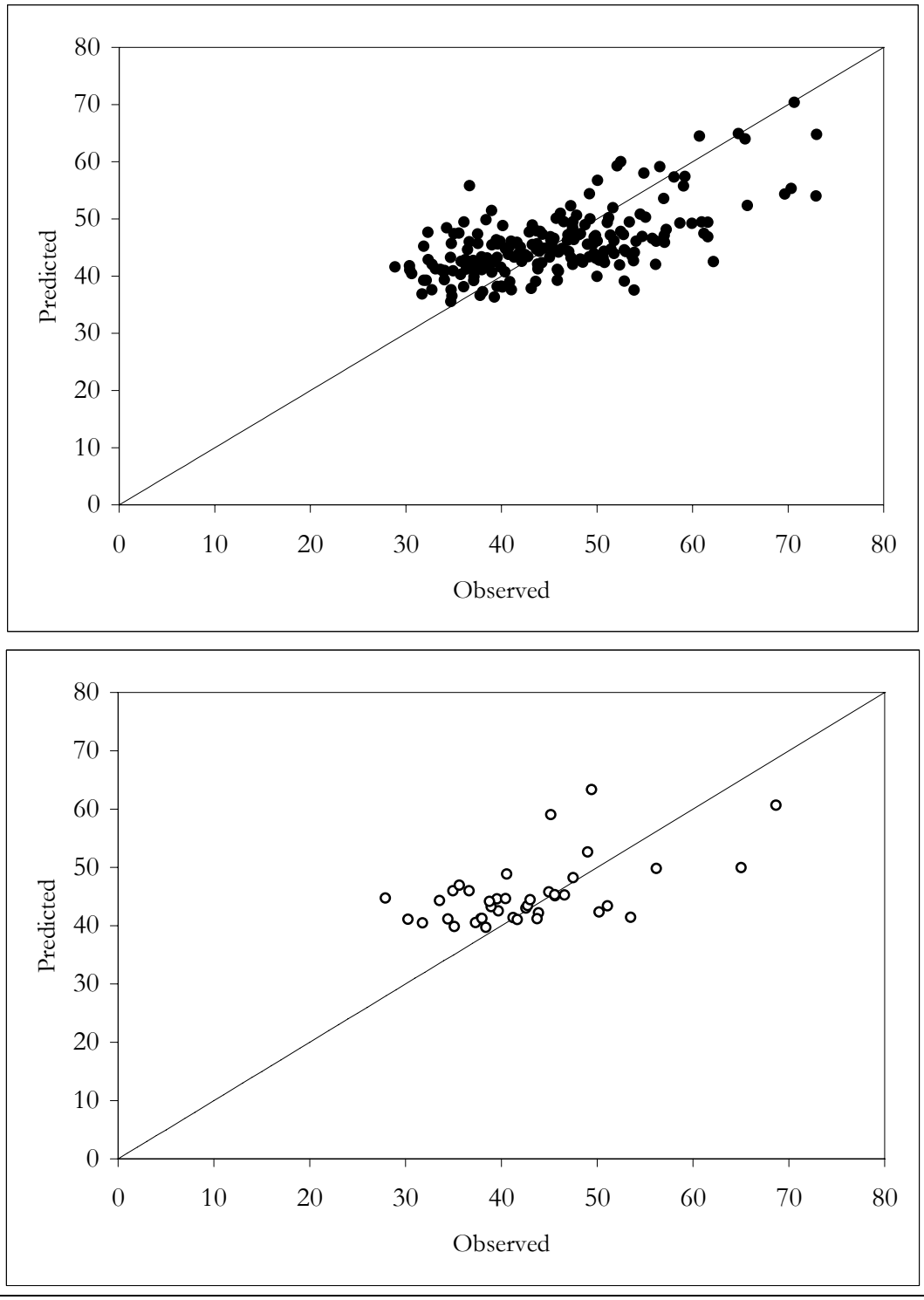


Figure 5.32. XY scatter plot of training (top) and validation data sets (bottom) for the mixed stepwise all possible subsets RT model without node pruning for the IB of OSB, n=200.

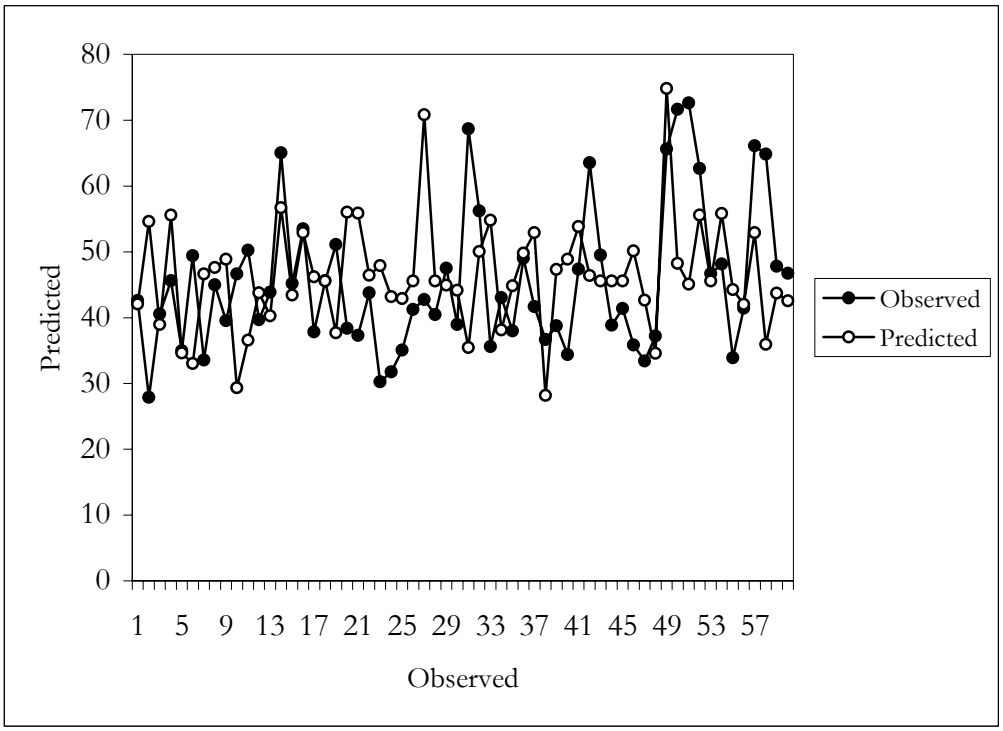


Figure 5.33. Time series graph of validation data set for the mixed stepwise all possible subsets RT model without node pruning for the IB of OSB, n=300.

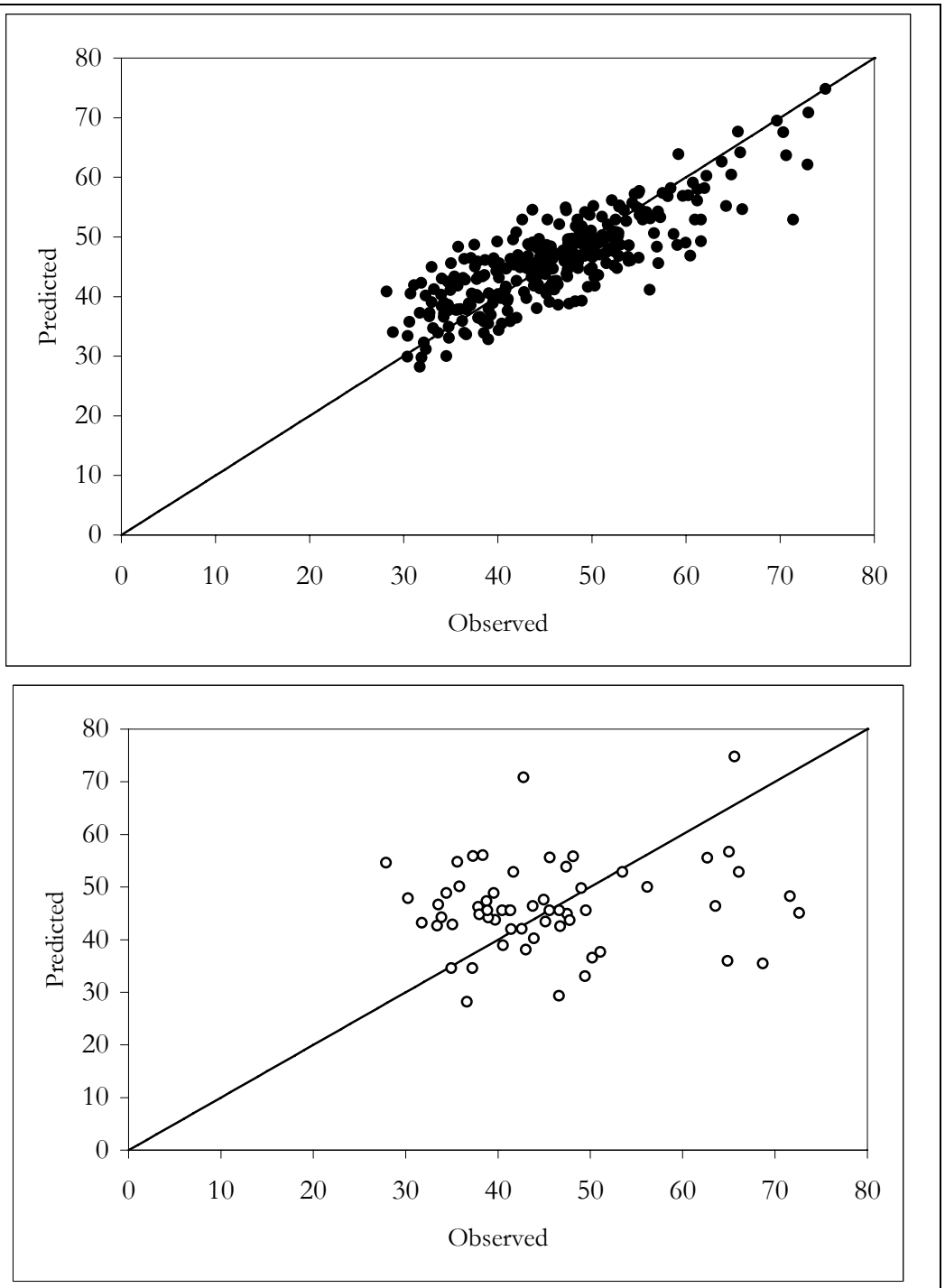


Figure 5.34. XY scatter plot of training (top) and validation data sets (bottom) for the mixed stepwise all possible subsets RT model without node pruning for the IB of OSB, n=300.

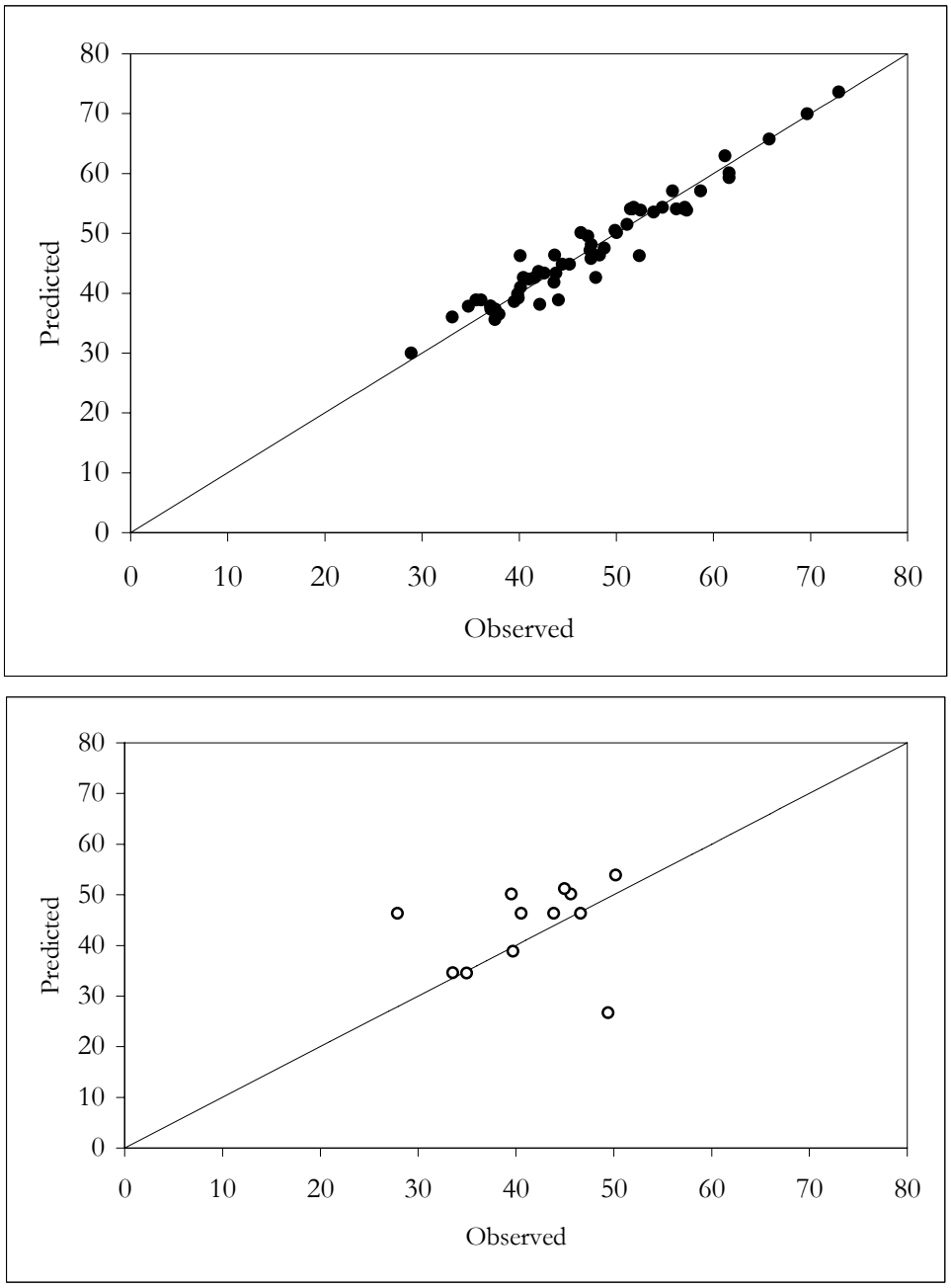


Figure 5.35. XY scatter plot of training (top) and validation data sets (bottom) for the second-order RT model without node pruning for the IB of OSB, n=59.

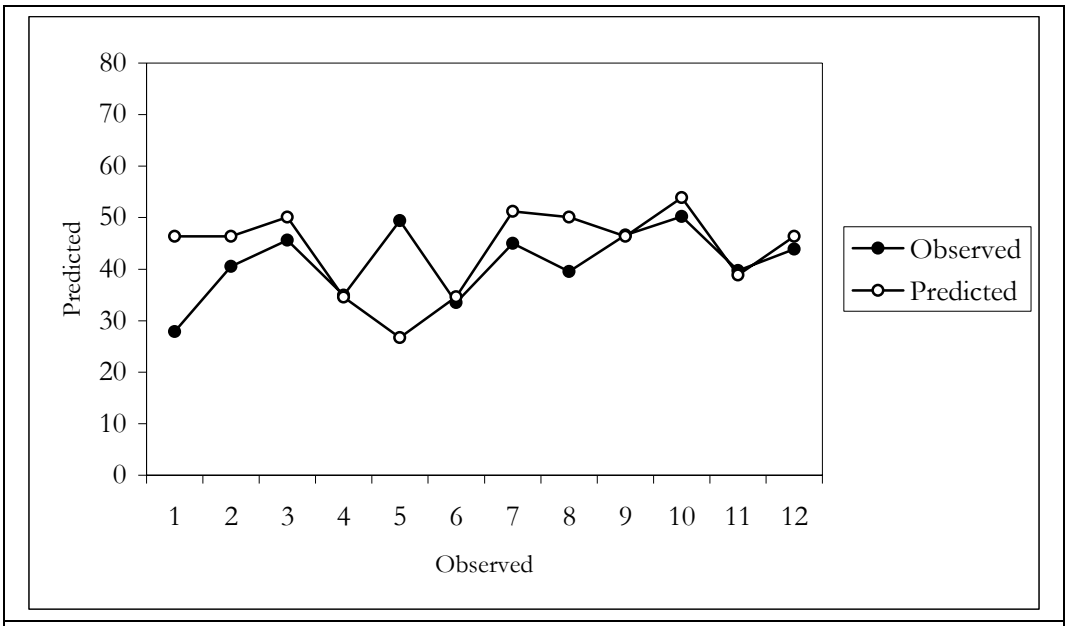


Figure 5.36. Time series graph of validation data set for the second-order RT model without node pruning for the IB of OSB, n=12.

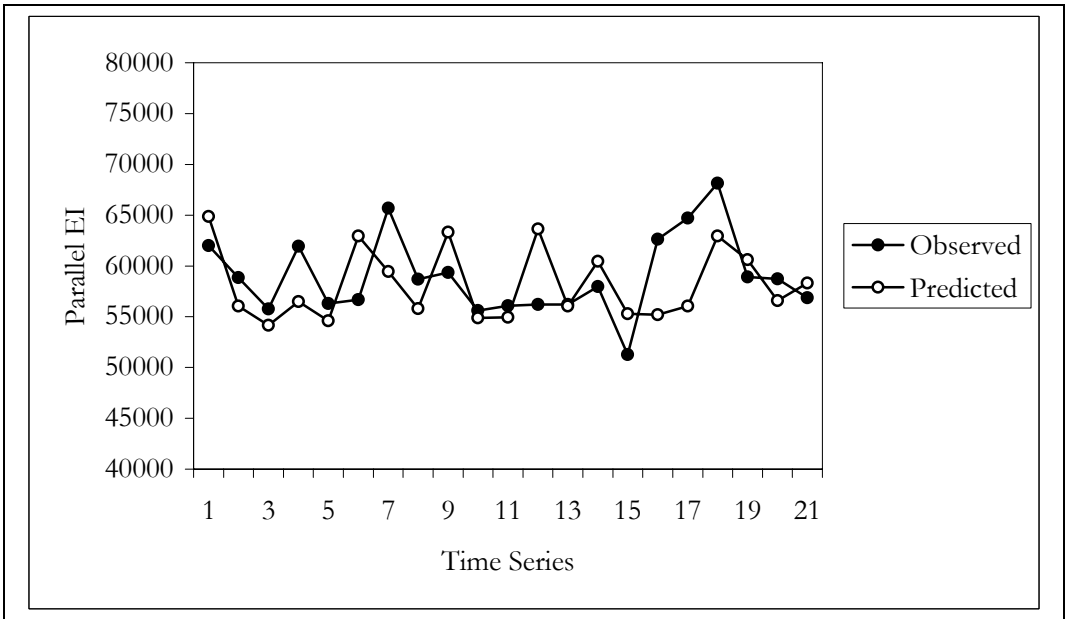


Figure 5.37. Time series graph of validation data set (lagged one time period) for the third-order RT model without node pruning for the Parallel EI of OSB, n=21.

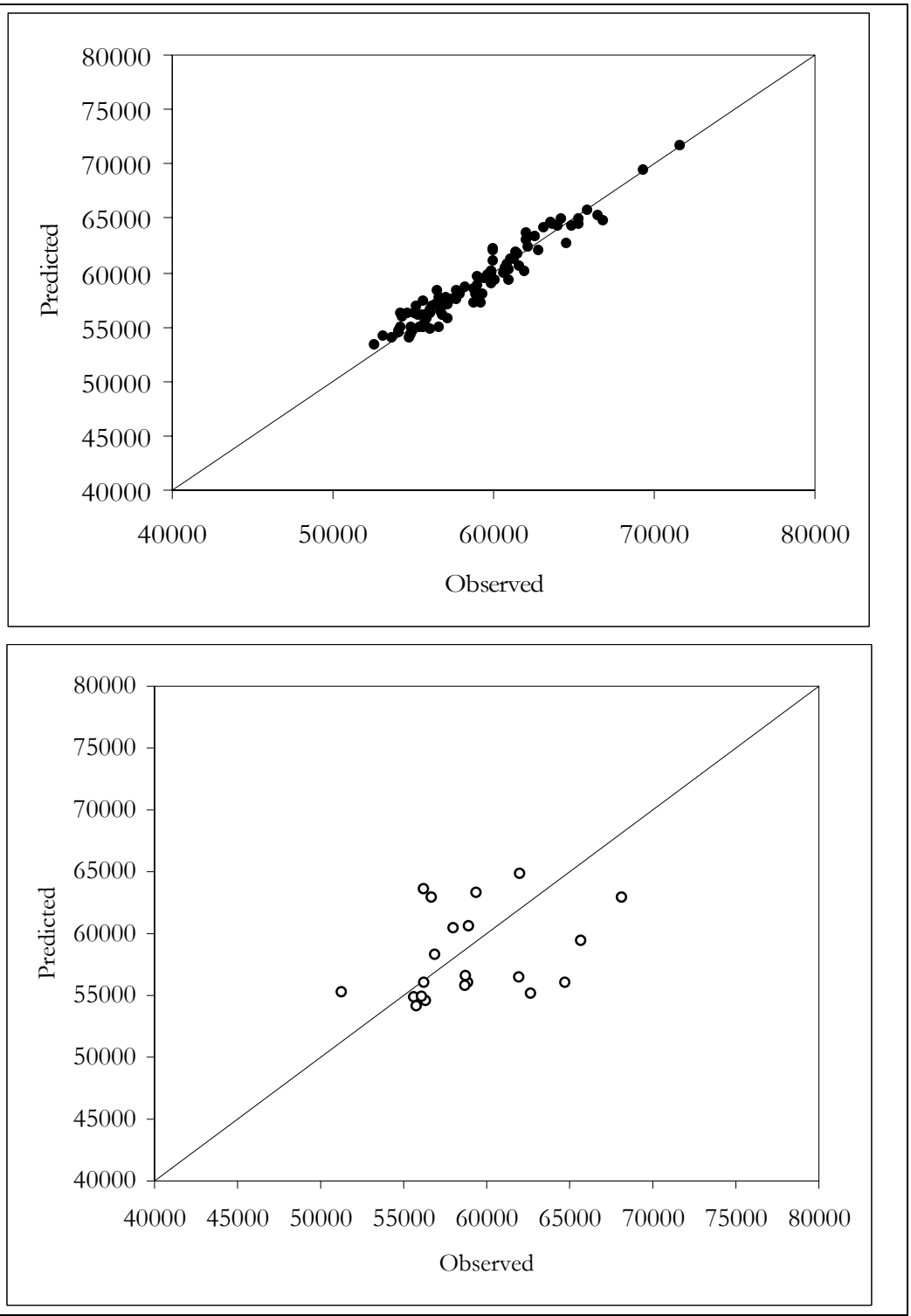


Figure 5.38. XY scatter plot of training (top) and validation data sets (bottom) for the third-order RT model without node pruning for the Parallel EI of OSB, n=100.

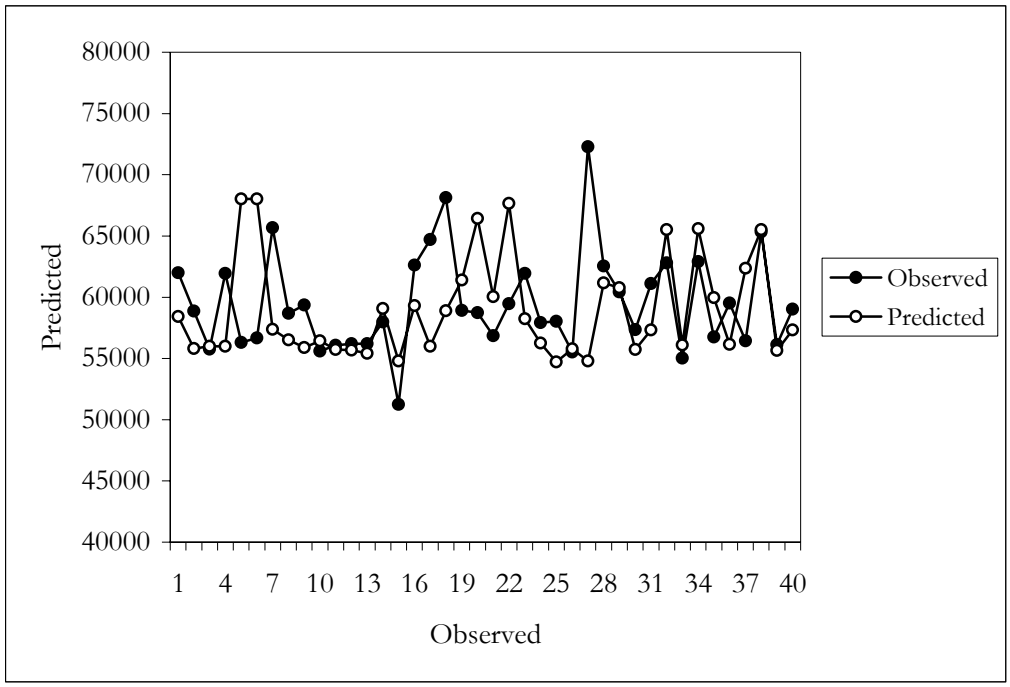


Figure 5.39. Time series graph of validation data set for the third-order RT model without node pruning for the Parallel EI of OSB, n=200.

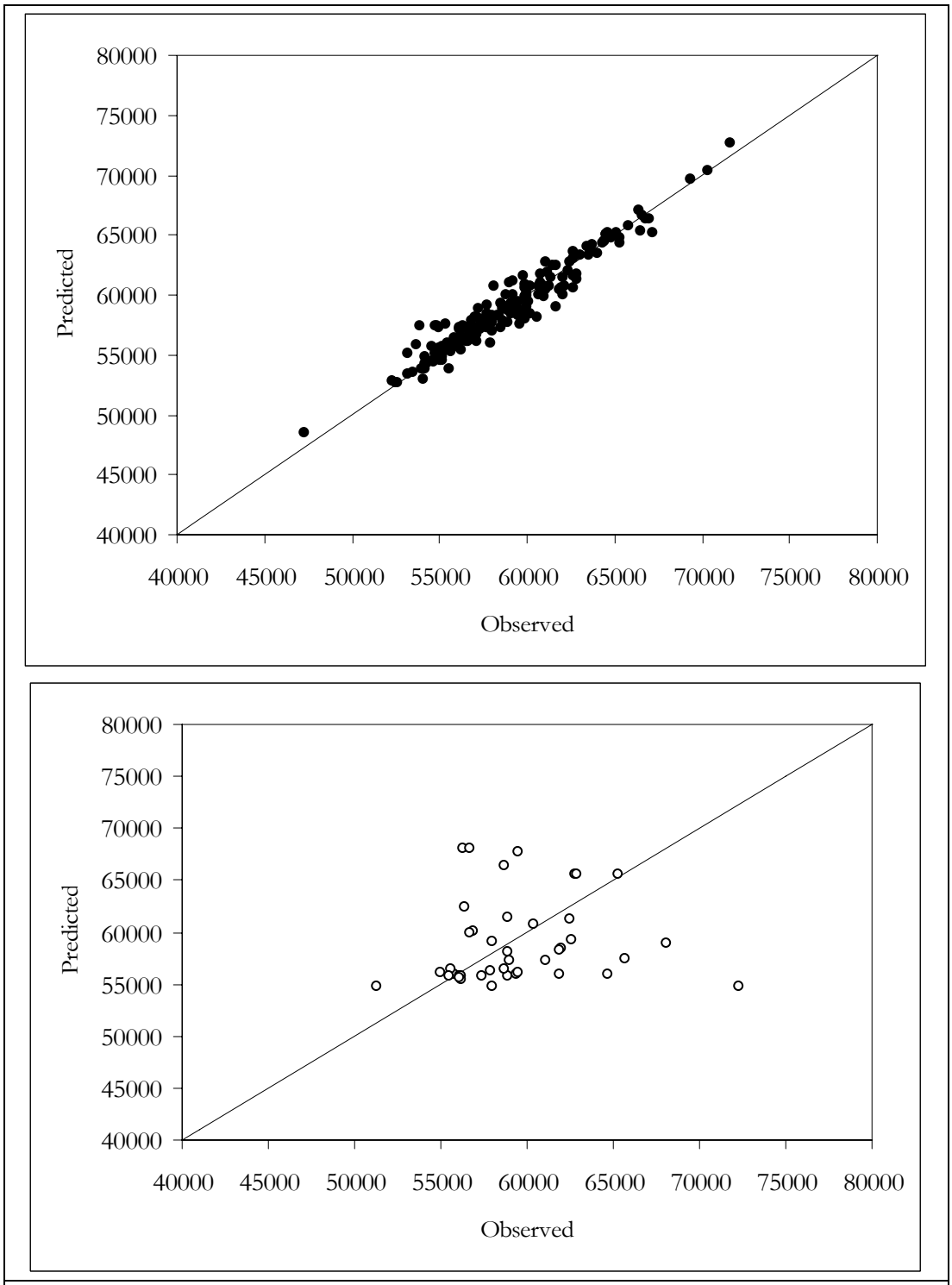


Figure 5.40. XY scatter plot of training (top) and validation data sets (bottom) for the third-order RT model without node pruning for the Parallel EI of OSB, n=200.

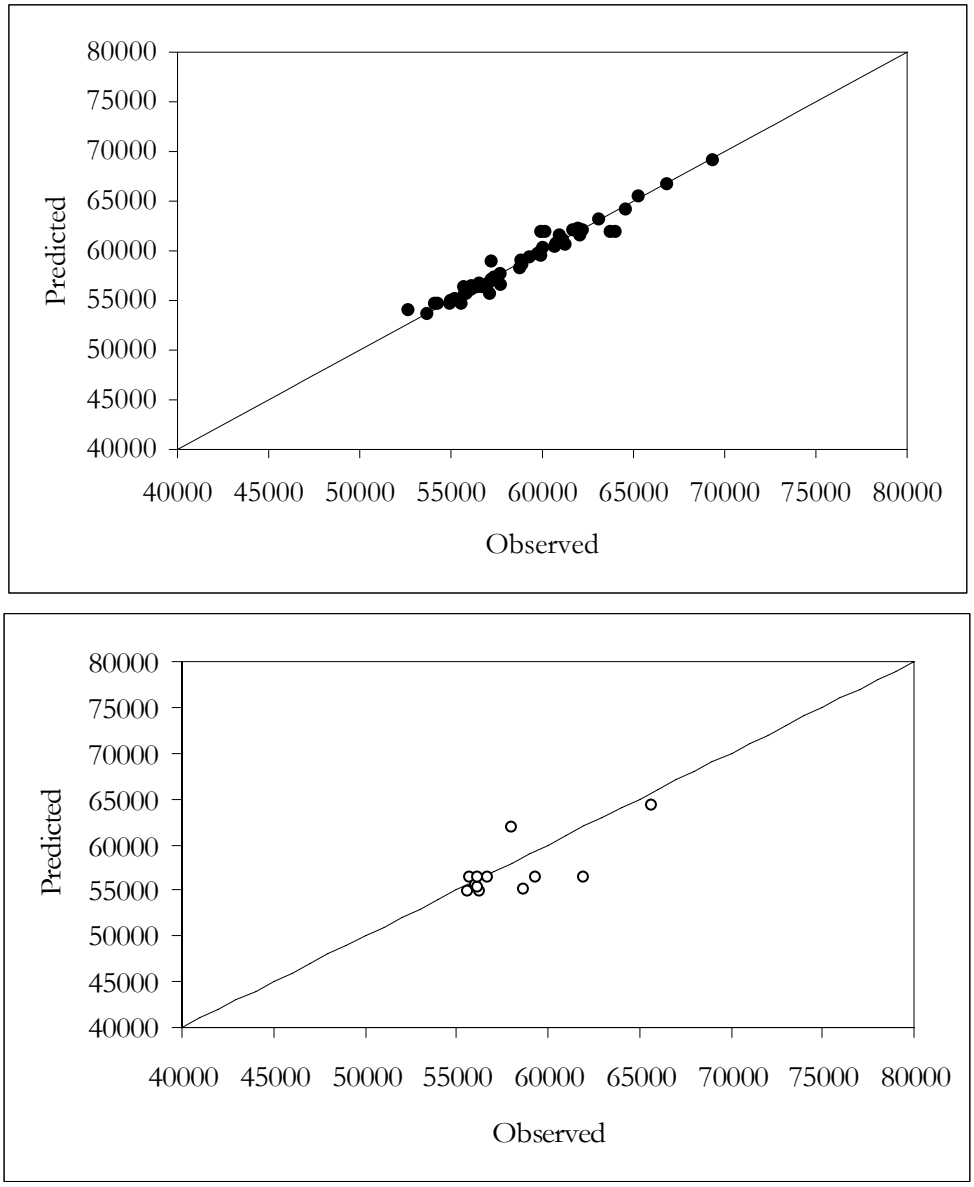


Figure 5.41. XY scatter plot of training (top) and validation data sets (bottom) for the second-order RT model without node pruning for the Parallel EI of OSB, n=58.

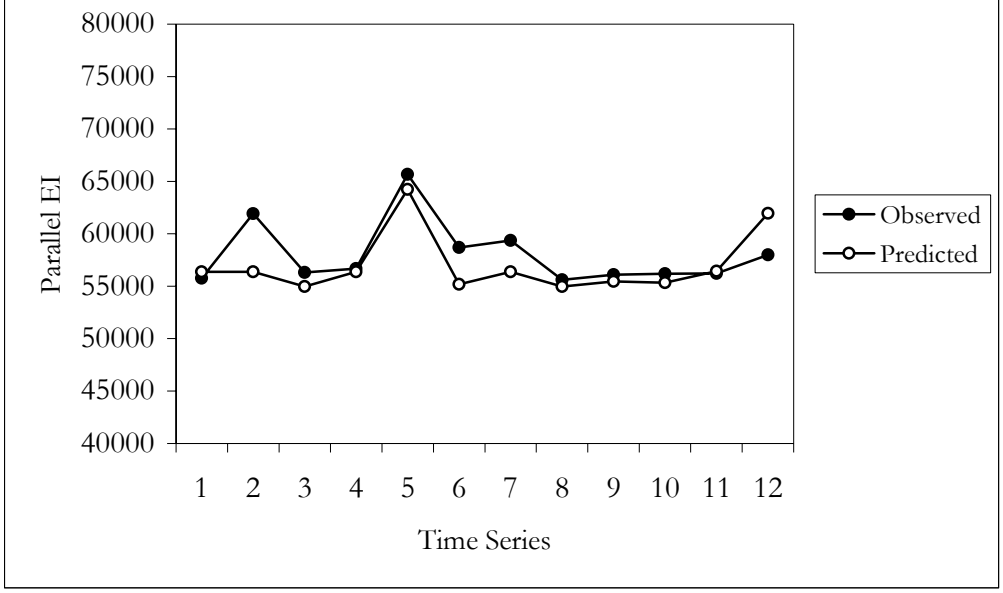
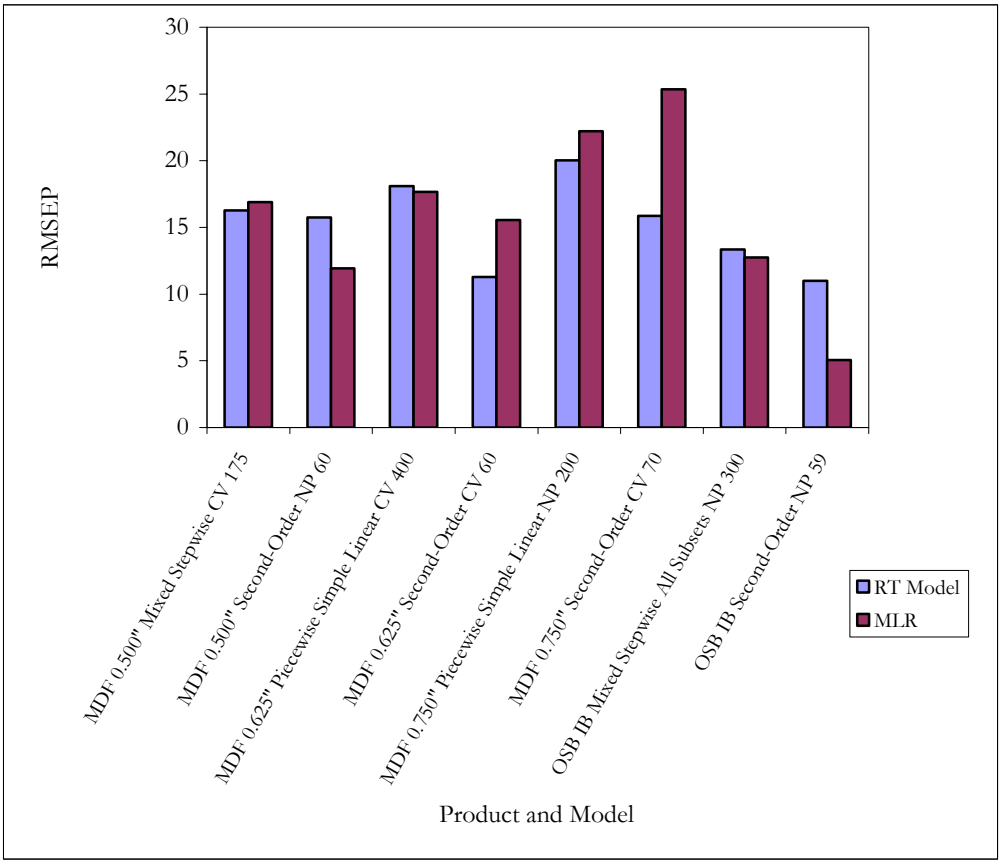


Figure 5.42. Time series graph of validation data set (lagged one time period) for the second-order RT model without node pruning for the Parallel EI of OSB, n=58.



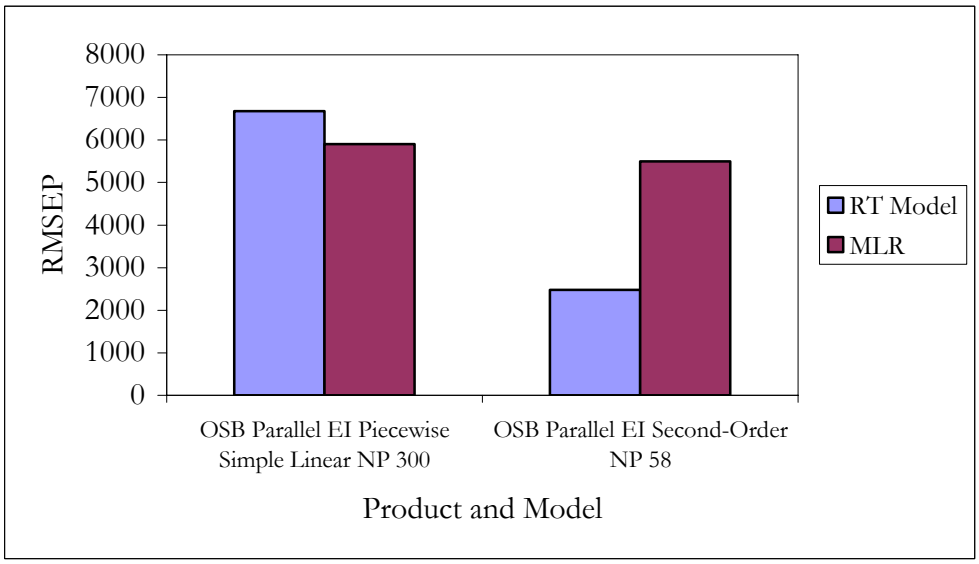


Figure 5.43. RMSEP for RT and MLR models discussed in Chapters IV and V.

CHAPTER VI.

REGRESSION TREE MODELS WITH BOX COX TRANSFORMATIONS OF OSB STRENGTH PROPERTIES WITH CONSIDERATIONS FOR QUANTILE REGRESSION

Results are presented in this chapter for the third objective of the dissertation. This chapter investigates regression trees (RT) with Box Cox transformations of *Y* for the internal bond (IB) of oriented strand board (OSB). Recall from the Chapter III Methods the description of the Box Cox transforms of the response *Y*, equation [3.5], Figure 3.1. The results of the chapter represent a synthesis of the results from Chapters III and IV. Preliminary studies of RT models with Box Cox transformations of IB for the MDF product types and the Parallel EI of OSB did not reveal any opportunities for RT model enhancement. Recall from Chapter III the departure from normality of the OSB strength properties relative to the approximate normal distributions of MDF strength properties, i.e., Box Cox transforms may be more helpful in modeling when the response *Y* departs from normality (Box and Cox 1964).

The ranking of regressors using GUIDE, version 5.2 (Loh 2006) indicate a difference in significant regressors with and without the Box Cox transform of the IB of OSB (Table 6.1). Rankings without the Box Cox transform indicate that the regressors "top and bottom core layer moisture contents" are highly ranked regressors. Regressors highly ranked with the Box Cox transform of IB are "top core layer forming spreader arm speed," "press overall step movement time" and "press overall time." The regressors related to "top and bottom core layer moisture contents" are not ranked in the presence of the Box Cox transform. In cases where the response *Y* departs from normality (e.g., IB of OSB), it may

be important when using RT methods to compare important regressors with and without the Box Cox transform. These types of transforms may improve the accuracy of detecting key sources of process variation that influence product strength. Accurate detection of sources of variation may optimize the use of critical company resources for programs targeted at improving product quality, product safety and lowering costs.

RT Models of OSB with Box Cox Transforms of IB

RT models with Box Cox transforms of IB are examined for four training data set record lengths, i.e., 59, 100, 200 and 300. These record lengths are consistent with those used in previous chapters. Recall the shortest record length of 59 is an outcome of the MLR results of Chapters IV and V. The time-ordering of the records of the training data set are identical to those used in previous chapters, e.g., a record length of 59 is a contiguous set of the most current records not including the validation data set. The record length of the validation data set is approximately 20 percent of the record length of the training data set records which are more recent than the training data set.

In the results, the predicted and observed IBs are converted back from the Box Cox transformations of IB for ease of comparison and interpretation. The two Box Cox transforms of IB identified in Chapter IV are:

$$Y^{\lambda} = \frac{Y^{1.8} - 1}{1.8\dot{Y}^{1.8-1}} = \frac{Y^{1.8} - 1}{38.7754}, \qquad [6.1]$$

and

$$Y^{\lambda} = \frac{Y^{0.2} - 1}{0.2\dot{Y}^{0.2-1}} = \frac{Y^{0.2} - 1}{0.0095}. \qquad [6.2]$$

OSB IB n=59

Eight parametric and four non-parametric RT models are investigated for this record length (Figure 6.1). The best candidate RT model with the Box Cox transform using equation [6.1] is a second-order model without node pruning. The RT model has 15 total nodes with eight terminal nodes (Figure 6.2). Key metrics of this model are a RMSE of 2.35 p.s.i., RMSEP of 9.21 p.s.i., R^2 of 0.90 and a homogenous residual pattern (Figure 6.3). Predicted IB for the validation data set approximate the observed IB in the time-ordered validation data set (Figures 6.4 and 6.5). Development of a quantile RT model of the median IB increases the RMSE to 5.75 p.s.i. but reduces the RMSEP to 7.91 p.s.i. with heteroscedasticity in the residuals. The predicted IB for the quantile model of the validation data set is comparable to the outcomes for the second-order RT model without node pruning developed in Chapter V (Figure 6.6).

The subgroup with the highest mean IB (55.6 p.s.i.) for this model occurs for a "main forming line total weight" > 24.54, "secondary dryer #3 outlet temperature ≤ 246.2 and "press time for step 701" ≤ 1023 (Figure 6.2). The simple linear equation with significant ($\alpha < 0.01$) regressors in the sub-tree is:

$$\text{IB} = 61.0 + 0.25(\text{main dryer \#5 outlet temperature}). \qquad [6.3]$$

The subgroup with the lowest mean IB (39.4 p.s.i.) for this model occurs for "main forming line total weight" > 24.54, "secondary dryer #3 outlet temperature > 246.2, "press time for step 742" ≤ 2730 and "main dryer #3 outlet temperature" > 261.8. The second-order model with significant ($\alpha < 0.01$) regressors in the sub-tree is:

$$\text{IB} = -5212 + 16.7(\text{core layer resin rate}) - 0.013(\text{core layer resin rate})^2. \qquad [6.4]$$

The statistical relationship between "main forming line total weight" and IB in the RT model (node 1) is illustrated in Figure 6.7. It may be important for the OSB manufacture to note "main forming line total weight" variation which falls below and above 25.54 and its influence on IB.

There is evidence from the predictions of IB in the validation data set that the RT model for this record length with the Box Cox transform of IB using equation [6.2] is superior to the second-order RT model without the Box Cox transform of IB developed in Chapter V. As Loh (2006) enhances GUIDE software, the addition of Box Cox transforms as a user-defined input may be an important consideration.

OSB IB n=100

Eight parametric and four non-parametric RT models are investigated for this record length (Figure 6.1). The best candidate RT model with the Box Cox transform is a mixed stepwise regression using all possible subsets without node pruning. The RT model has 21 total nodes with 11 terminal nodes. Key metrics of this model are a RMSE of 5.46 p.s.i., RMSEP of 9.08 p.s.i., R^2 of 0.64 and a homogenous residual pattern (Figure 6.8). Predicted IB for the validation data set approximate the observed IB in the time-ordered validation data set (Figures 6.9 and 6.10). Development of a quantile RT model of the median IB reduces the RMSE to 4.25 p.s.i. but increases the RMSEP to 14.58 p.s.i. with a homogeneous pattern in the residuals. The predicted IB for the quantile model of the validation data set does not perform as well as mixed stepwise all possible subsets RT model without node pruning.

There is evidence that the RT model for this record length with the Box Cox transform of IB using equation [6.2] is superior to the RT model with mixed stepwise

regression using all possible subsets without node pruning developed in Chapter V (Table 5.8, Appendix to Chapter V). Even though the RMSEP of 9.08 p.s.i. is higher and the R^2 is lower (0.64) relative to the RT model without transforms (RMSEP of 7.44 p.s.i., R^2 of 0.73), the strength of the RT model with Box Cox transforms is its better predictability of IB.

OSB IB n=200

Eight parametric and four non-parametric RT models are investigated for this record length (Figure 6.1). The best candidate RT model with the Box Cox transform using equation [6.1] is a piecewise simple linear RT model with v-fold cross-validation node pruning. The RT model has 7 total nodes with 4 terminal nodes. Key metrics of this model are a RMSE of 8.37 p.s.i., RMSEP of 9.60 p.s.i., R^2 of 0.37 and a homogenous residual pattern (Figure 6.11). Predicted IB does not approximate the observed IB in the time-ordered validation data set (Figures 6.12 and 6.13). The best candidate quantile RT model of the median IB reduces the RMSE to 4.47 p.s.i. but increases the RMSEP to 21.60 p.s.i. with a homogeneous pattern in the residuals. The predicted IB for the quantile model in the validation data set does not approximate observed IB.

There is no evidence that the RT model for this record length with the Box Cox transform of IB is superior to the RT model with mixed stepwise regression using all possible subsets without node pruning developed in Chapter V (Table 5.8, Appendix to Chapter V). Both the RMSE and RMSEP are higher than those of the RT model without the Box Cox transform and predictions of IB in the validation data set are worse.

OSB IB n=300

Eight parametric and four non-parametric RT models are also investigated for this record length (Figure 6.1). The best candidate RT model with the Box Cox transform using

equation [6.1] is a mixed stepwise regression using all possible subsets without node pruning. The RT model has 73 total nodes with 37 terminal nodes. Key metrics of this model are a RMSE of 6.06 p.s.i., RMSEP of 9.90 p.s.i., R^2 of 0.51 and a homogenous residual pattern (Figure 6.14). Predicted IB does not approximate observed IB in the time-ordered validation data set (Figures 6.15 and 6.16). The best candidate quantile RT model of the median IB increases the RMSE to 6.16 p.s.i. and increases the RMSEP to 16.82 p.s.i. with a homogeneous pattern in the residuals. The predicted IB for the quantile model of the validation data set does not approximate observed IB.

There is no evidence that the RT model for this record length with the Box Cox transform of IB is superior to the RT model without the Box Cox transform developed in Chapter V (Table 5.8, Appendix to Chapter V). The RMSE of 6.06 p.s.i. is higher than the RT model without transforms (RMSE of 5.07 p.s.i.), but the RMSEP of 9.90 p.s.i. is lower than that of the RT model without transforms (RMSEP of 13.35). Both models perform poorly in predicting IB in the time-ordered validation data set. It is interesting to note that the RT model with Box Cox transforms performs slightly better for the first 33 predictions of IB relative to the RT model without transforms, but the model predicts IB erratically in the later half of the validation data set.

Chapter VI Summary

The investigation of 48 RT models of the IB of OSB using Box Cox transforms of *Y* indicate that such transforms improve predictability of IB for the shorter record lengths of 59 and 100. There is no evidence that Box Cox transforms for longer records lengths of 200 and 300 improve predictability of the IB of OSB. There is also no evidence that quantile RT models with Box Cox transforms improve predictability of IB. The RT models with Box

Cox transforms for the shorter record lengths have strong explanatory value of IB e.g., tree structure in Figure 6.2).

Of the 48 models investigated in this chapter, 12 RT models have lower RMSEPs with Box Cox transforms. The mixed stepwise all subsets RT model for a record length of 300 has a lower RMSEP of 6.06 p.s.i. when compared to all of the other MLR models with and without Box Cox transforms of Chapter IV (Table 4.13). However, the RMSEP of 9.21 p.s.i. for a record length of 59 is greater that the RMSEPs for all MLR models of Chapter IV of the same record length.

The ranking of regressors using GUIDE (Loh 2006) indicate a difference in significant regressors with and without the Box Cox transform of Y using equation [6.1]. Regressors highly ranked with the Box Cox transform of equation [6.1] are "top core layer forming spreader arm speed," "press overall step movement time" and "press overall time." The regressors related to "top and bottom core layer moisture contents" which are ranked as being important without Box Cox transforms are not ranked. This difference in rankings may be important when investigating response variables that exhibit departures from normality as is the case with the IB of OSB (recall Chapter III). The possible incorrect ranking of significant regressors of the process may result in misallocation of scarce resources of a manufacturer, i.e., investigation and quality improvement programs targeted at process variables that do not have strong influences on strength properties. The results of this chapter highlight the importance of investigating models with transforms which may be especially important when response variables depart from normality.

Appendix to Chapter VI

Table 6.1. Comparison of ten most important independent variables for OSB IB using GUIDE scoring with and without Box Cox transform.

Without Box Cox Transform		**With Box Cox Transform**	
Score	Description	Score	Description
100.00	Top core layer moisture content	100.00	Top core layer forming spreader arm speed
95.65	Bottom core layer moisture content	88.72	Press overall step movement time
72.85	Press position time	82.93	Press overall time
63.15	Press overall time step movement time	80.96	Top core layer forming speed
62.78	Press overall time	73.68	Press time for step 741B
56.89	Bottom core layer density set-point	73.68	Press time for step 741A set-point
54.94	Top core layer forming speed	73.37	Main flaker speed
53.80	Blender bottom surface layer wood total	71.93	Press position time
52.32	Total time of press open to press close	70.47	Bottom surface layer forming spreader arm speed
70.78	Blender bottom surface layer resin total	69.97	Press time for step 741C

*Process variables highlighted in blue are common across all product types.

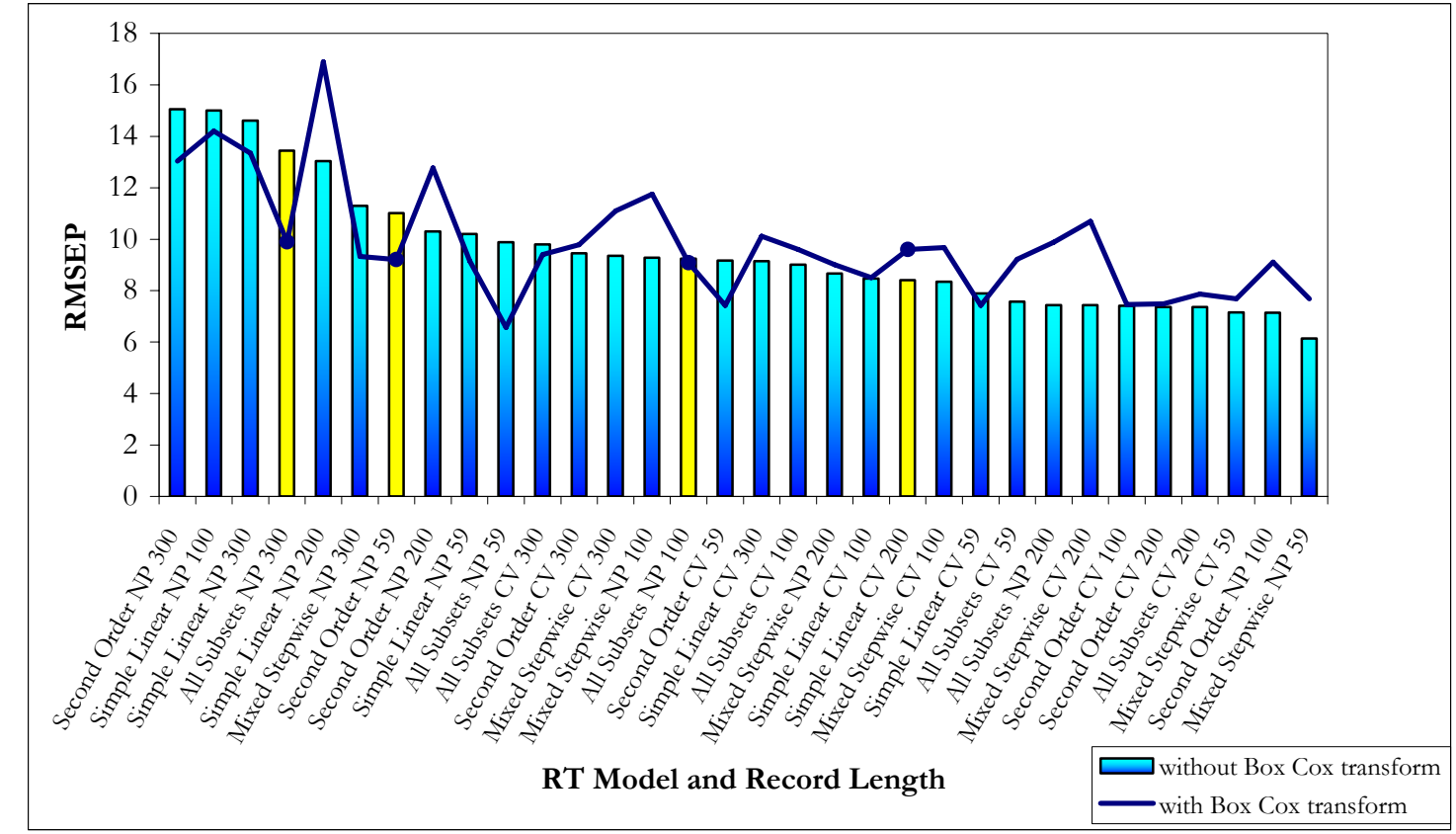


Figure 6.1. Comparison of RMSEP with and without Box Cox transform for all RT models analyzed for OSB IB (yellow bars indicate best candidate models discussed in this chapter).

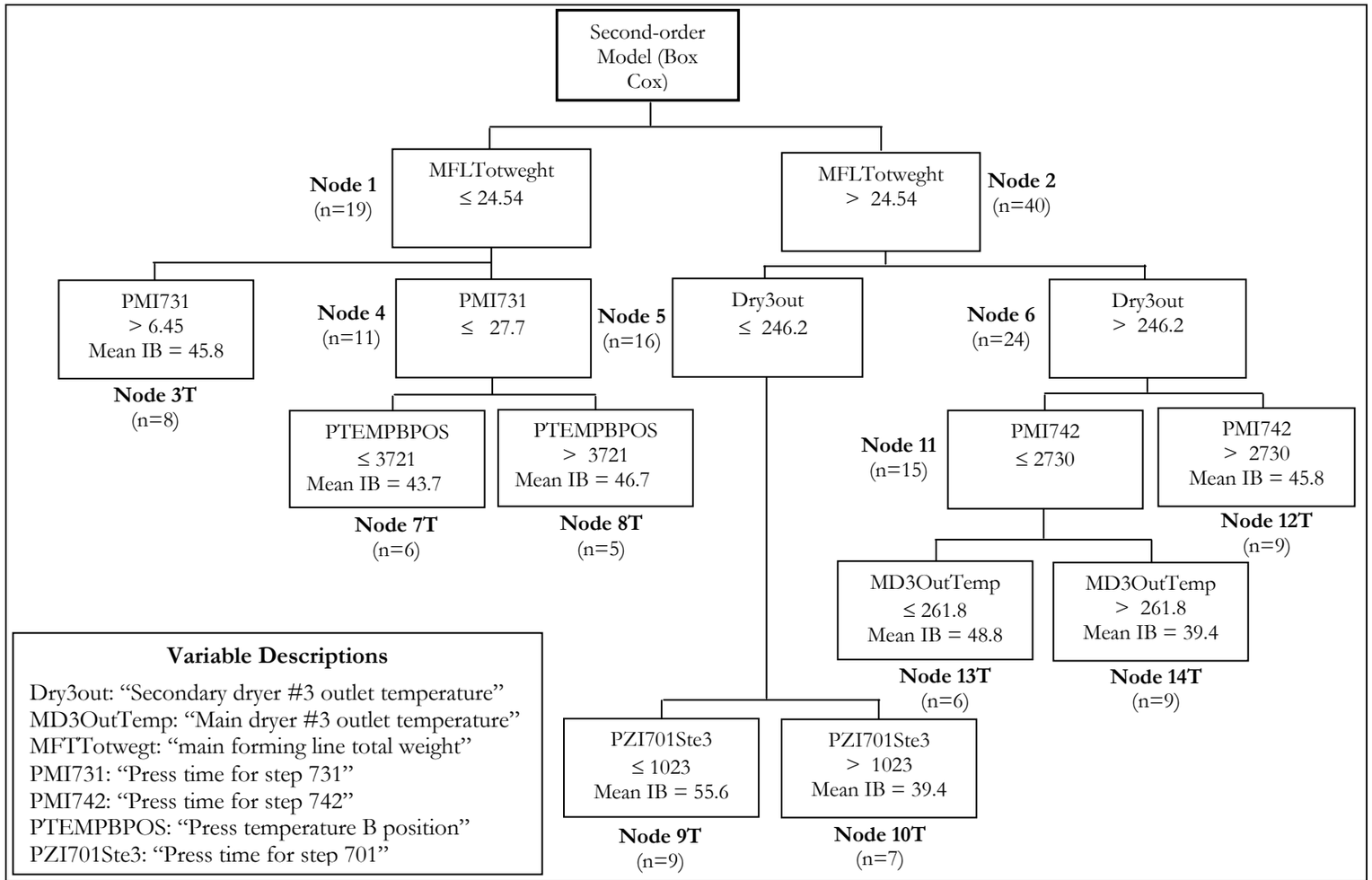
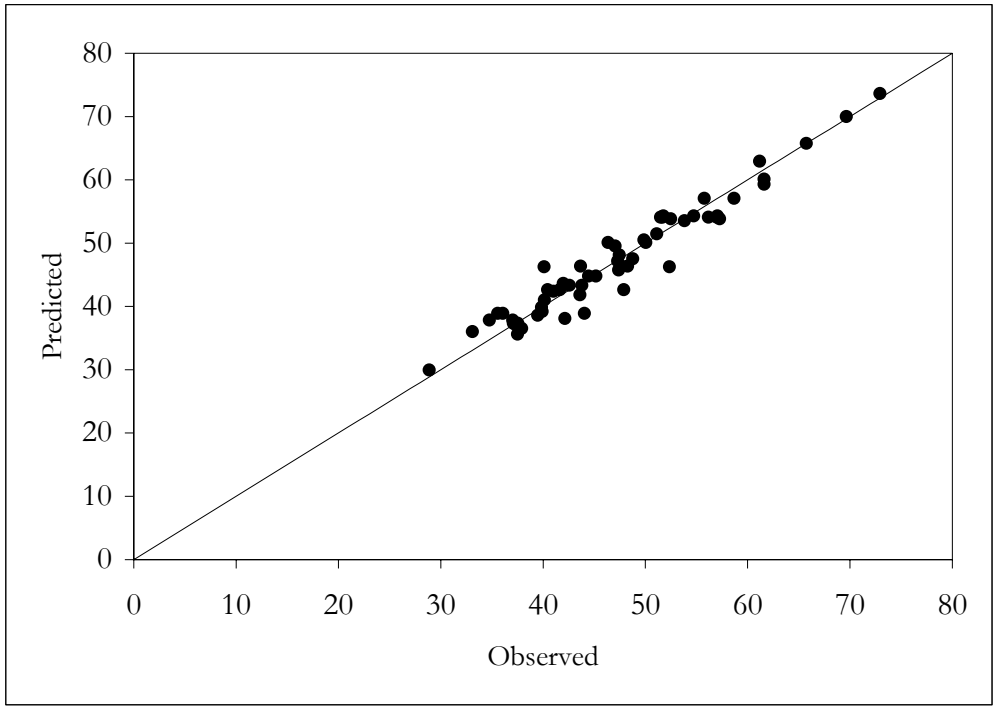


Figure 6.2. Second-order model for OSB IB (n=59) with Box Cox transformation.



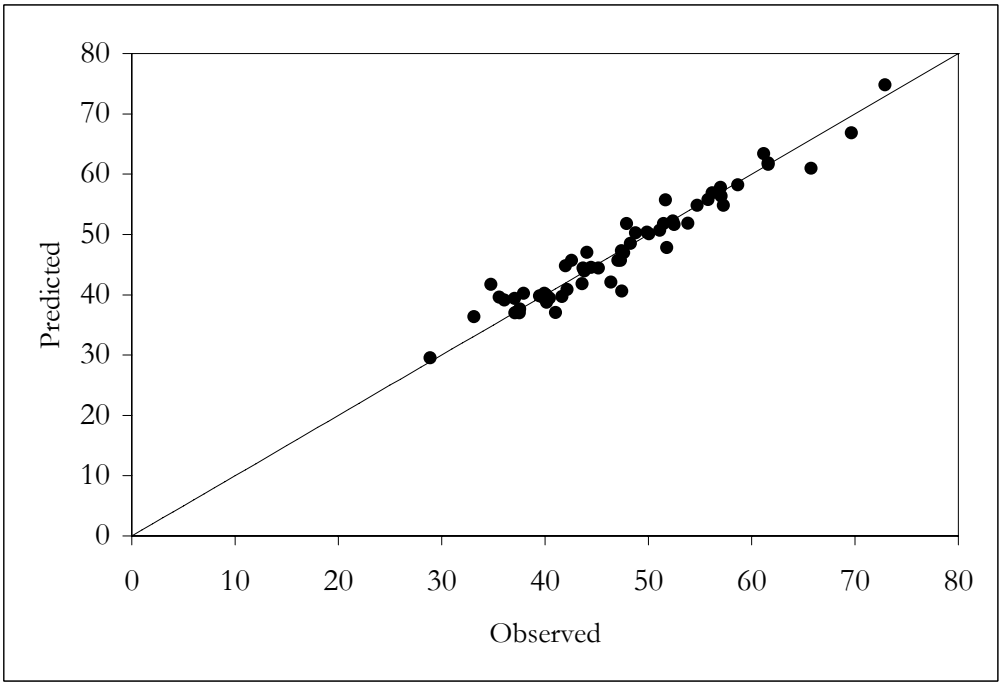


Figure 6.3. XY scatter plots of predicted and observed IB for OSB IB (n=59) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

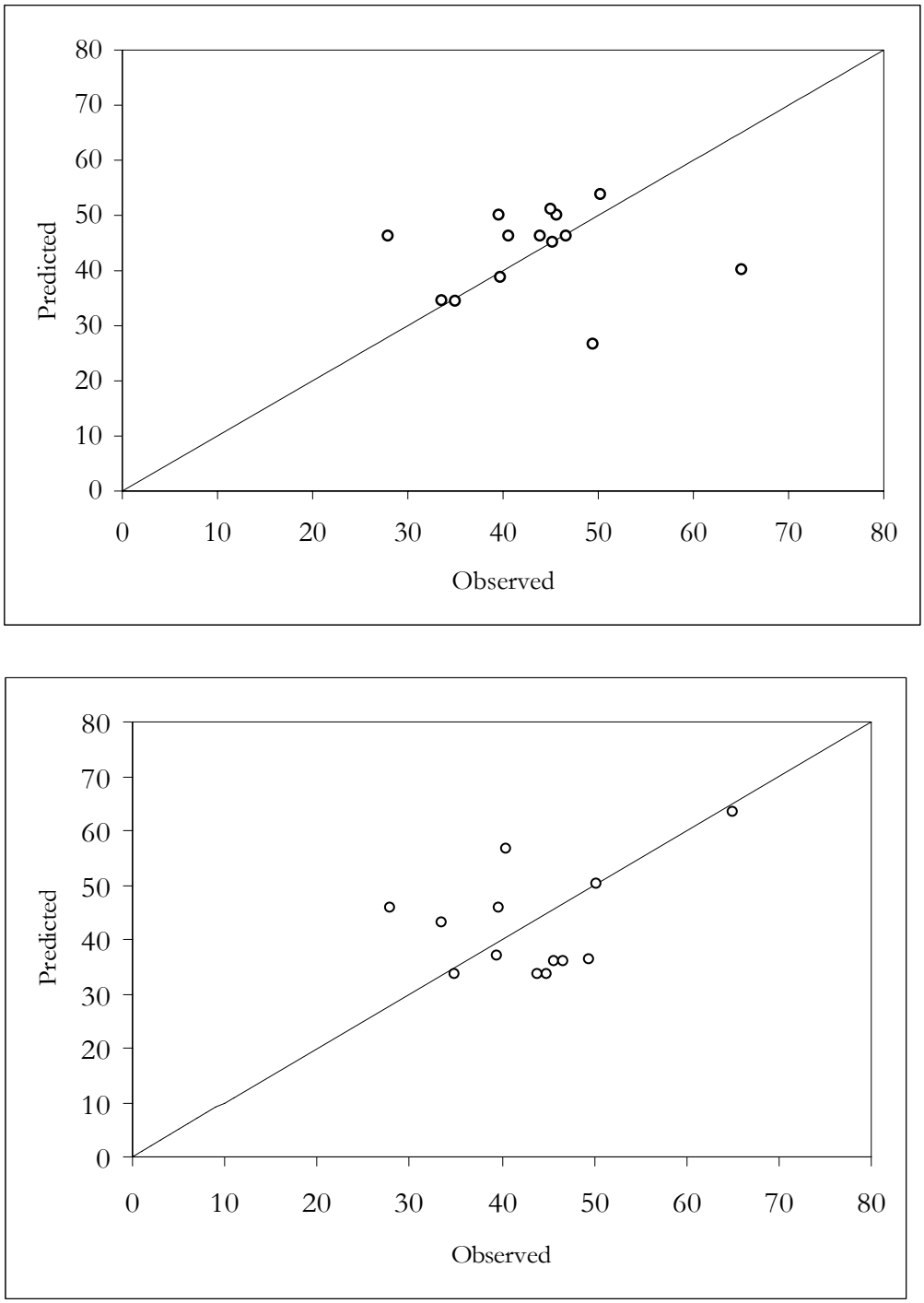


Figure 6.4. XY scatter plot of validation data sets for OSB IB (n=59) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

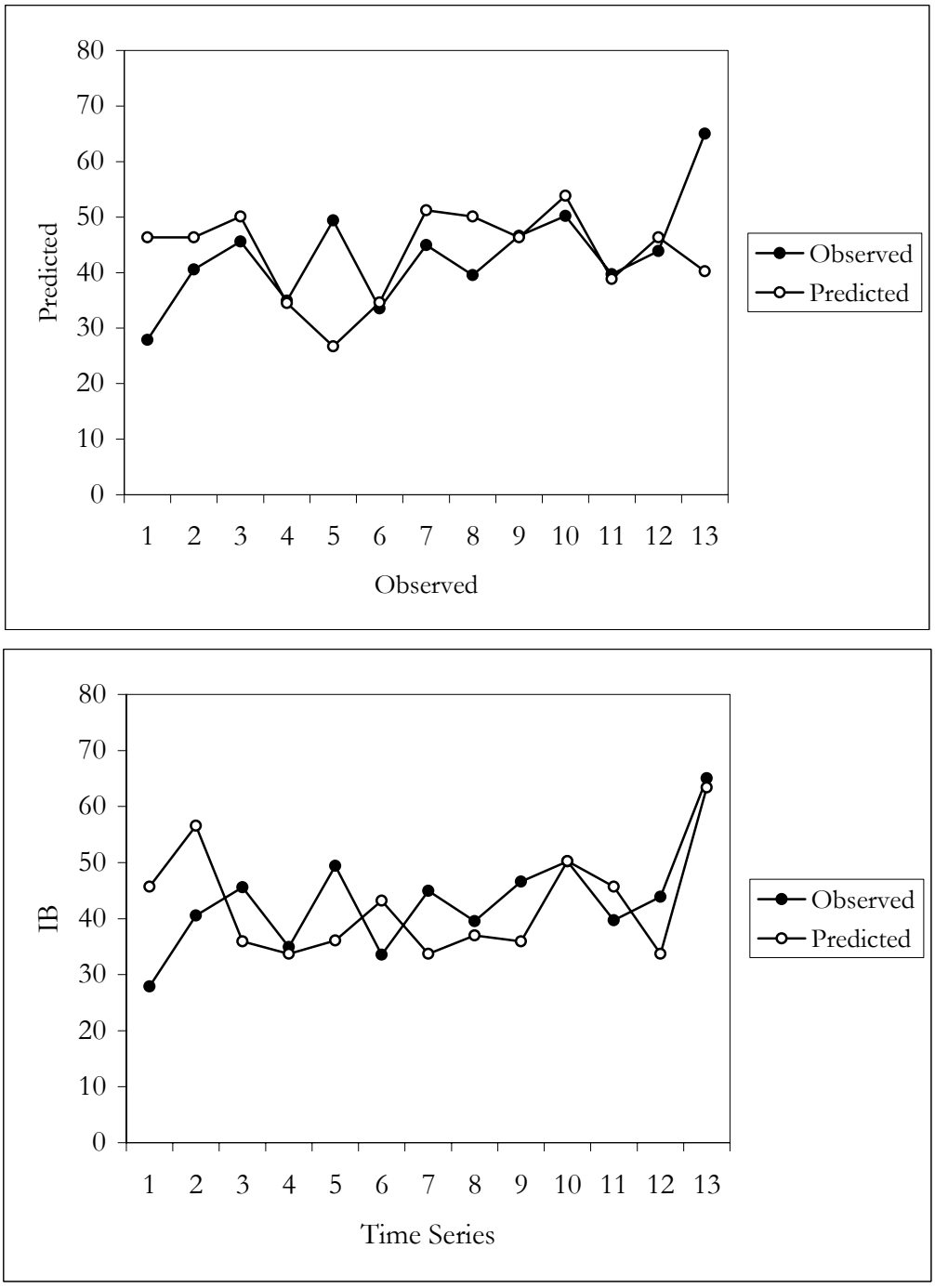


Figure 6.5. Time series graph of validation data sets for OSB IB (n=59) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

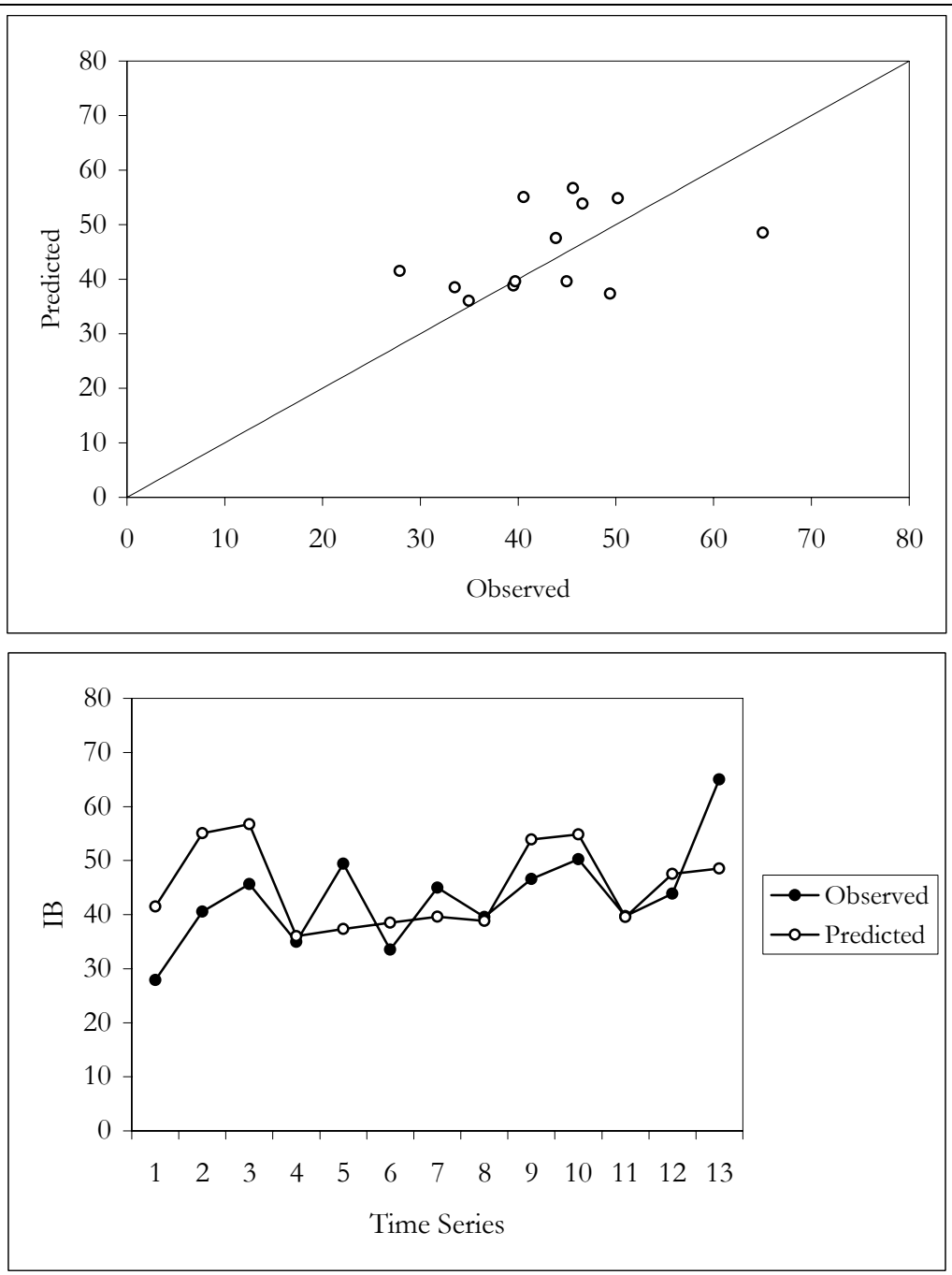
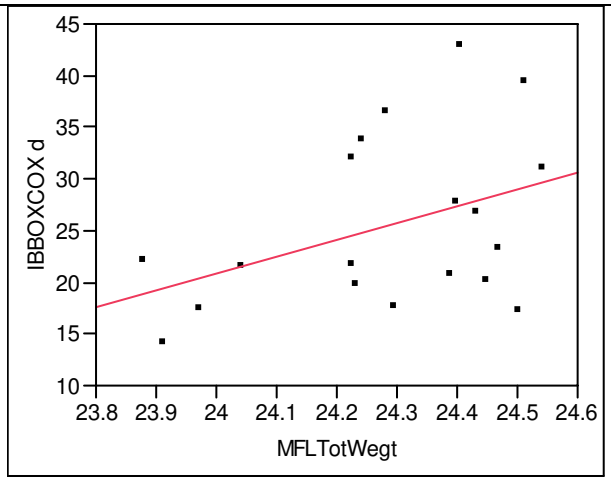


Figure 6.6. XY scatter plot (top graph) and time series graph (bottom graph) of predicted and observed IB for quantile RT model for OSB IB, n=59.

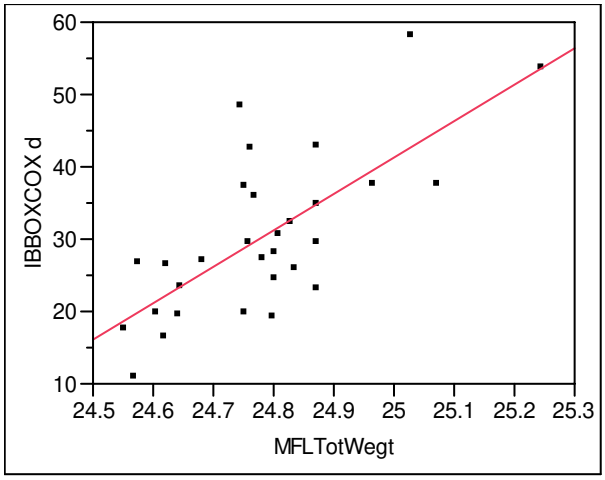


Linear Fit

IB transform = -368.4208 + 16.222431*MFLTotWegt

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-368.4208	216.0693	-1.71	0.1064
MFLTotWegt	16.222431	8.897779	1.82	0.0859



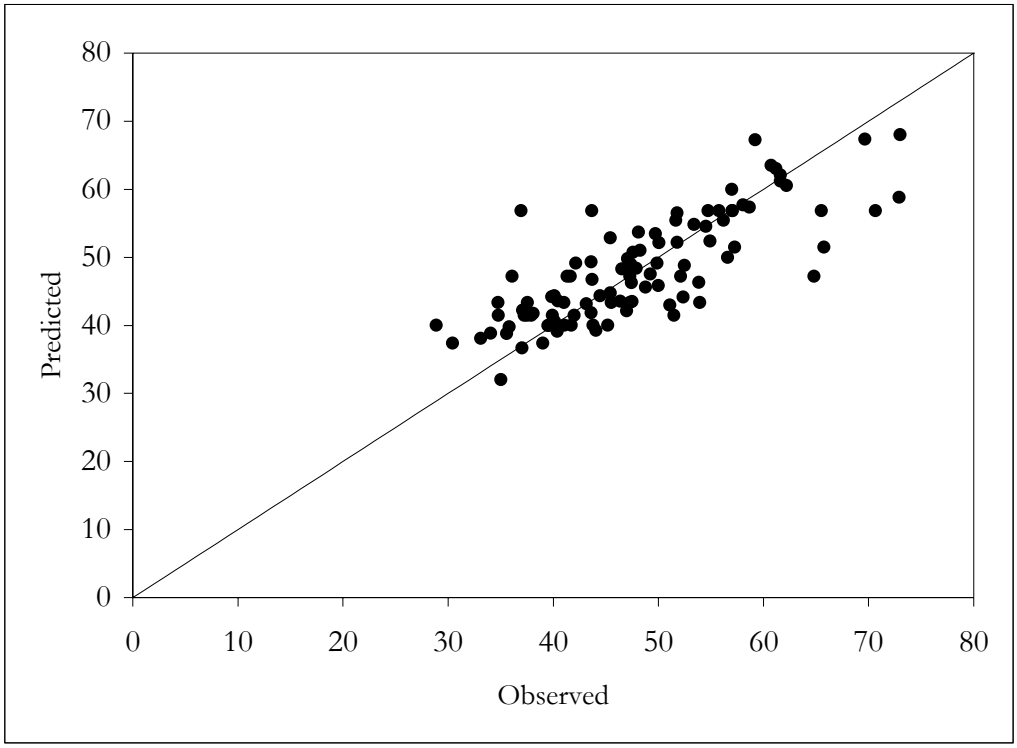
Linear Fit

IB transform = -1219.316 + 50.420183*MFLTotWegt

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1219.316	231.2688	-5.27	<.0001
MFLTotWegt	50.420183	9.332103	5.40	<.0001

Figure 6.7. Illustration of regression model differences for "main forming line total weight" in node one of RT model for OSB IB (n=59).



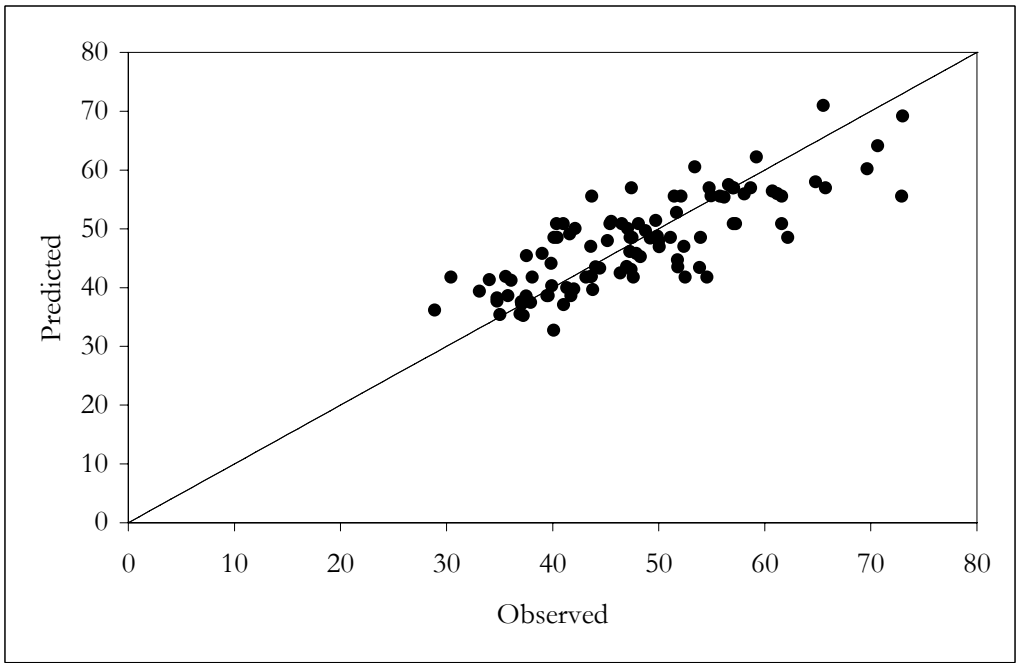


Figure 6.8. XY scatter plots of predicted and observed IB for OSB IB (n=100) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

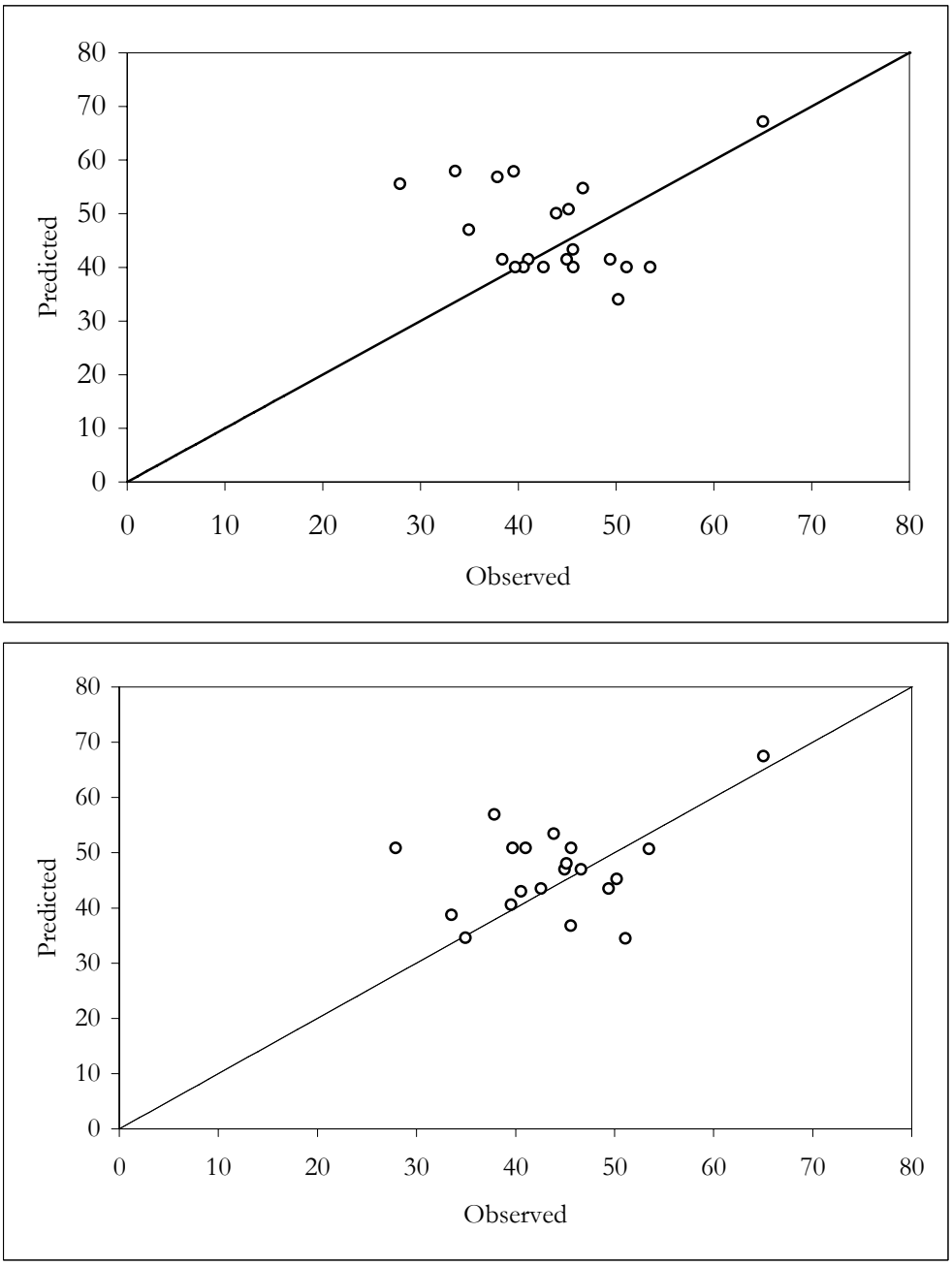


Figure 6.9. XY scatter plot of validation data sets for OSB IB (n=100) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

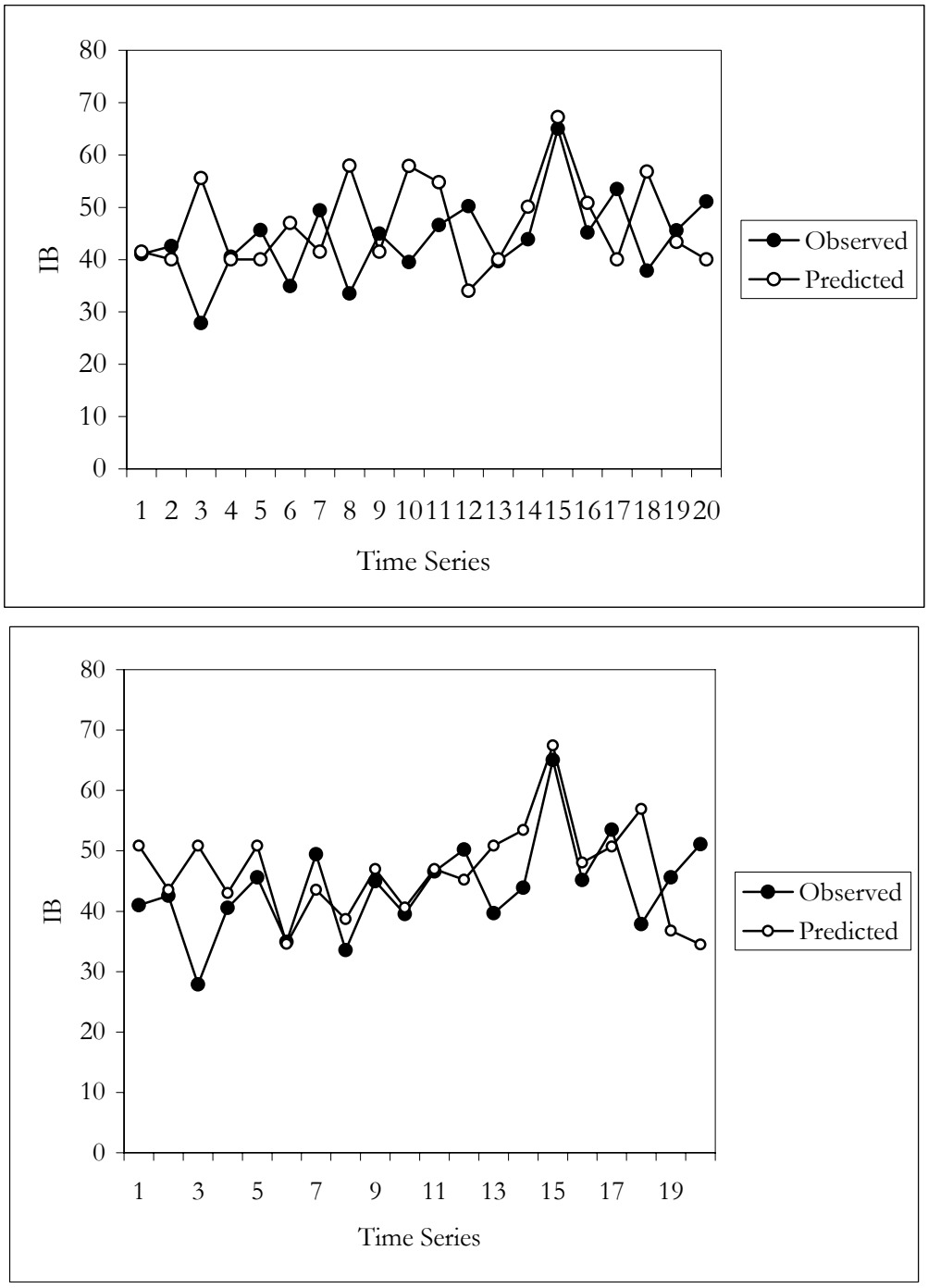
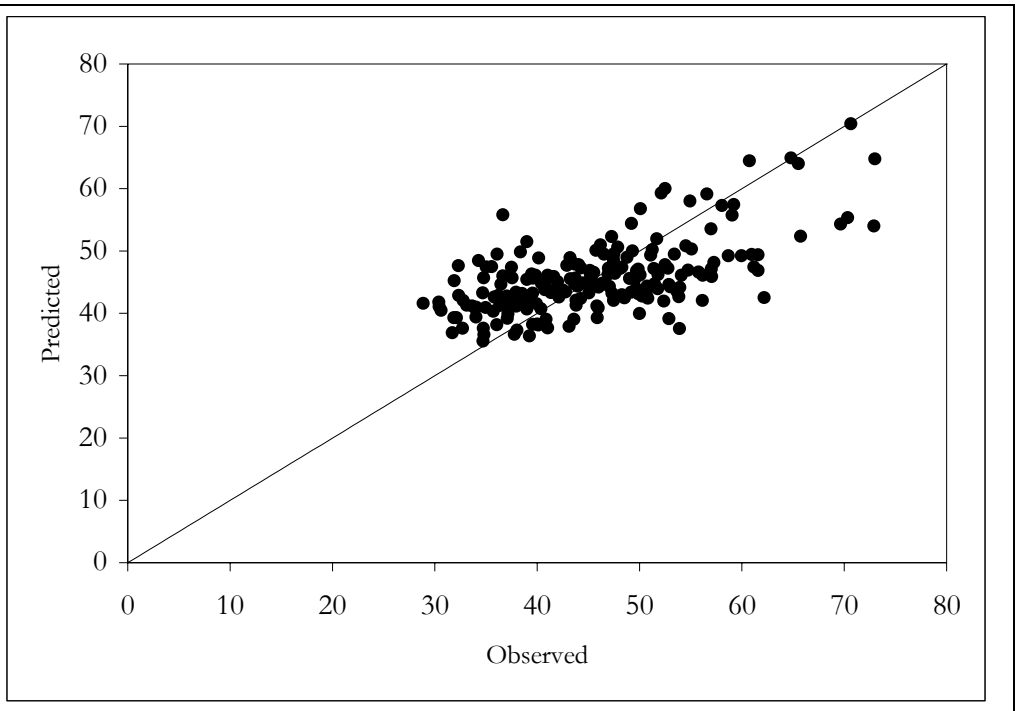


Figure 6.10. Time series graph of validation data sets for OSB IB (n=100) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).



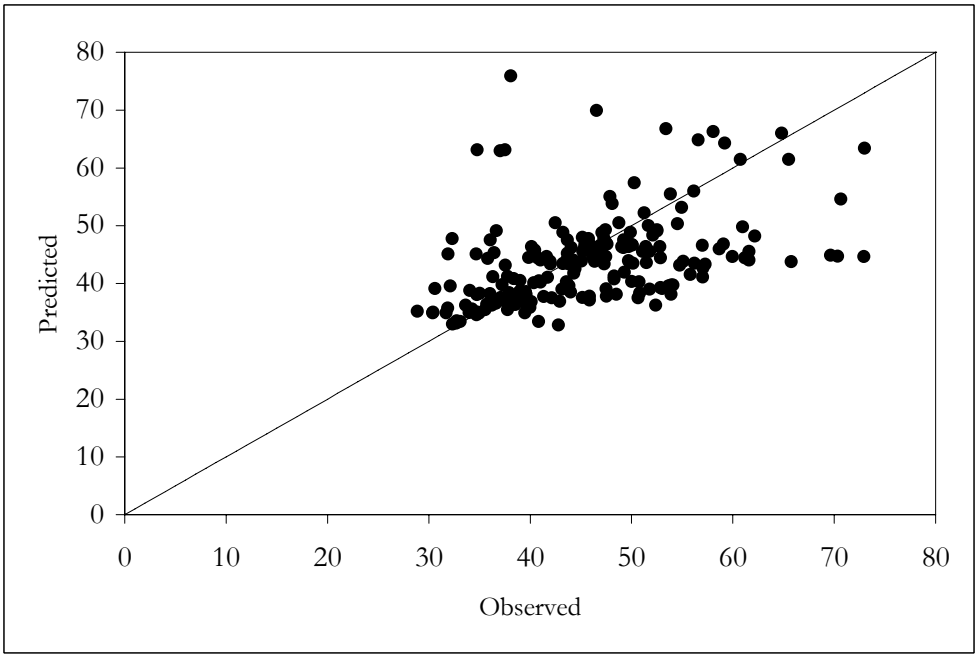


Figure 6.11. XY scatter plots of predicted and observed IB for OSB IB (n=200) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

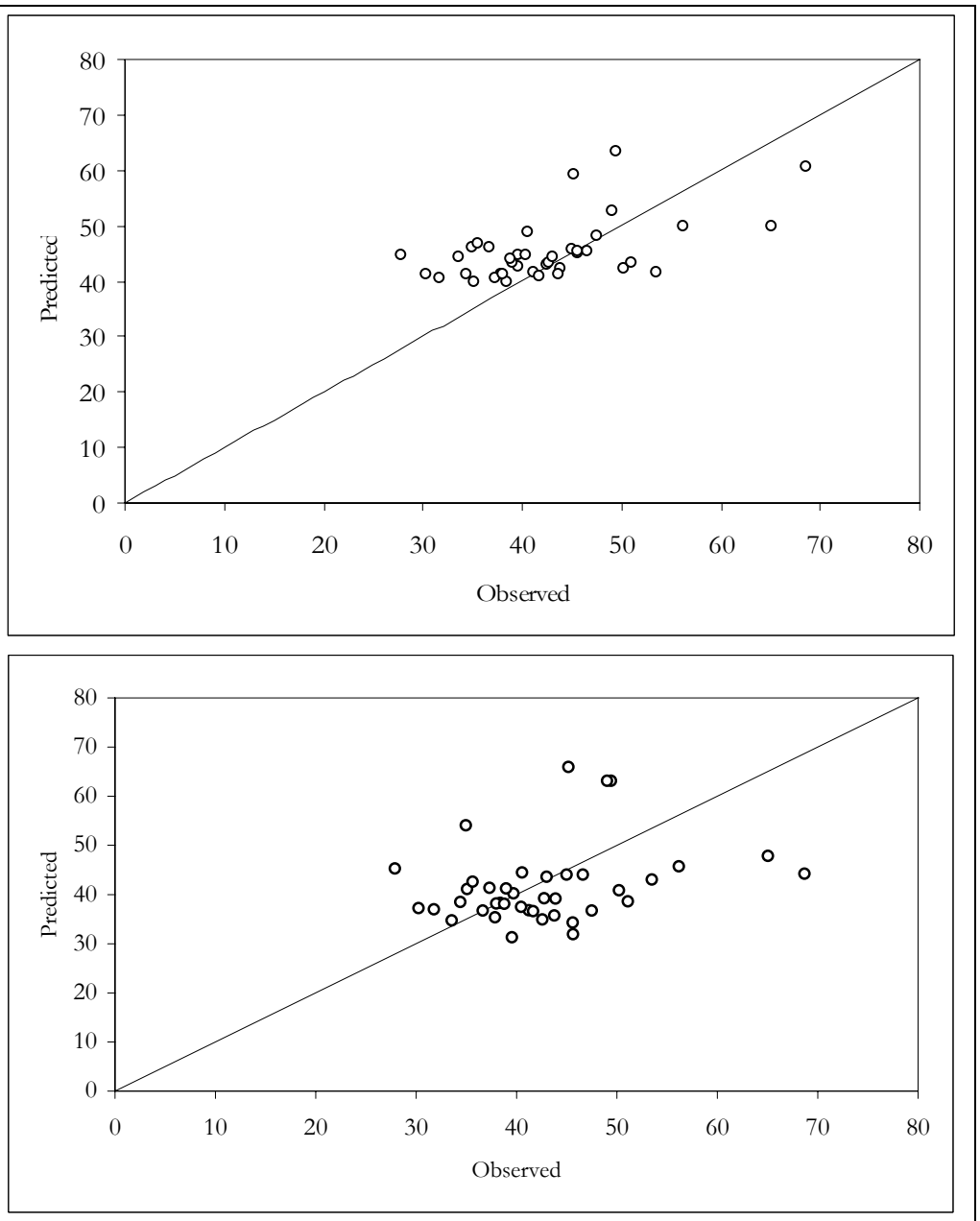


Figure 6.12. XY scatter plot of validation data sets for OSB IB (n=200) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

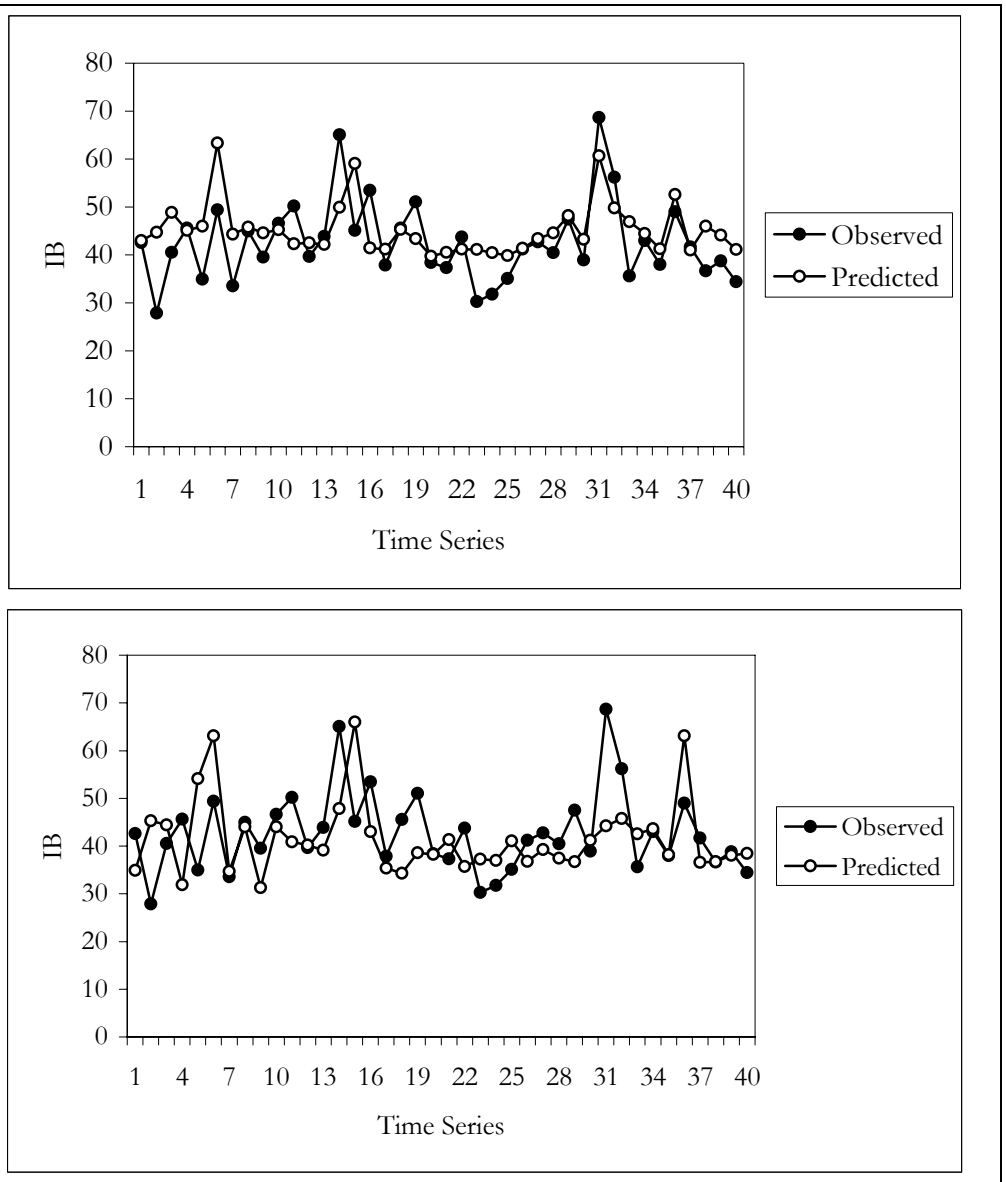


Figure 6.13. Time series graph of validation data sets for OSB IB (n=200) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

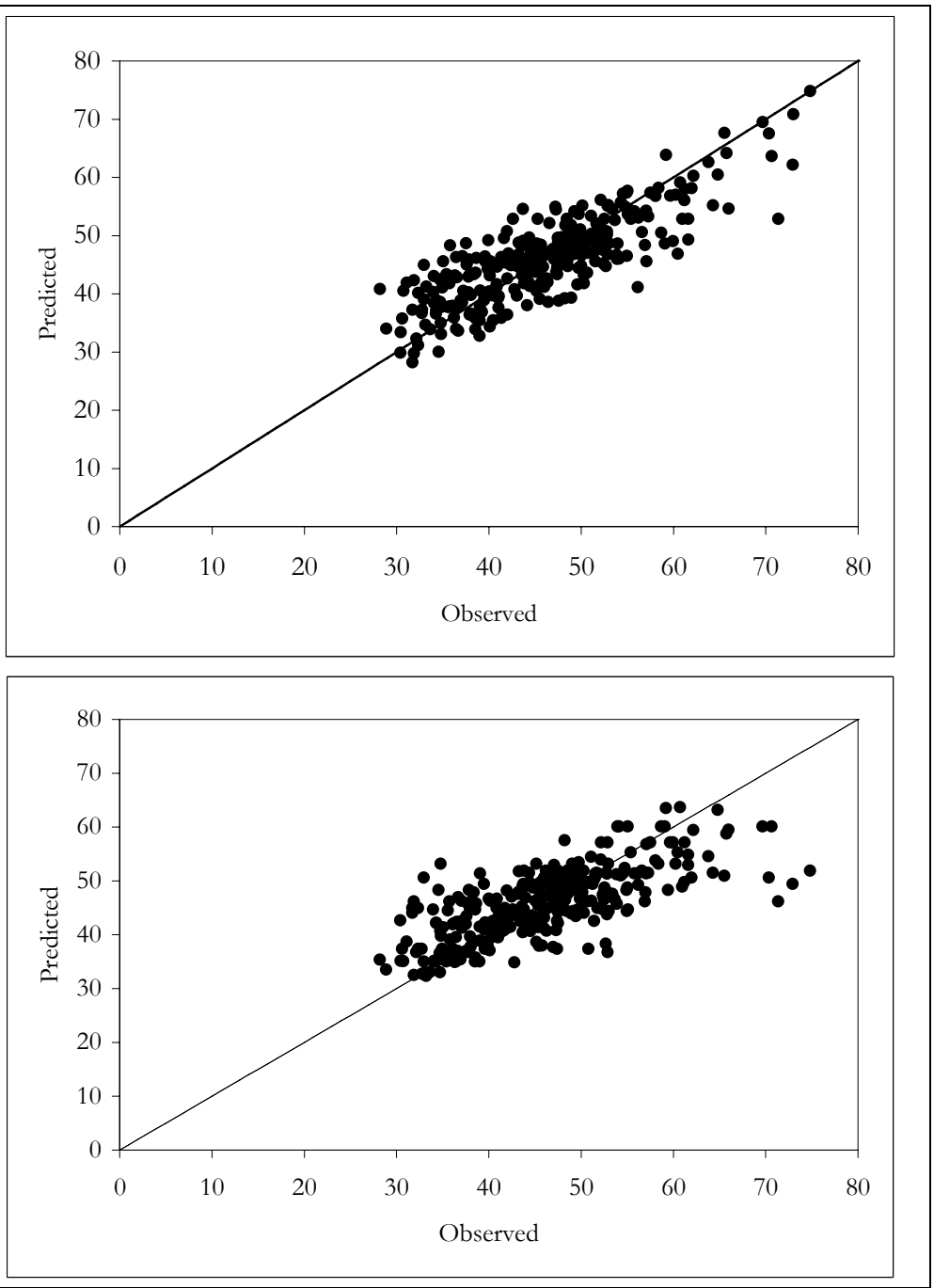


Figure 6.14. XY scatter plots of predicted and observed IB for OSB IB (n=300) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

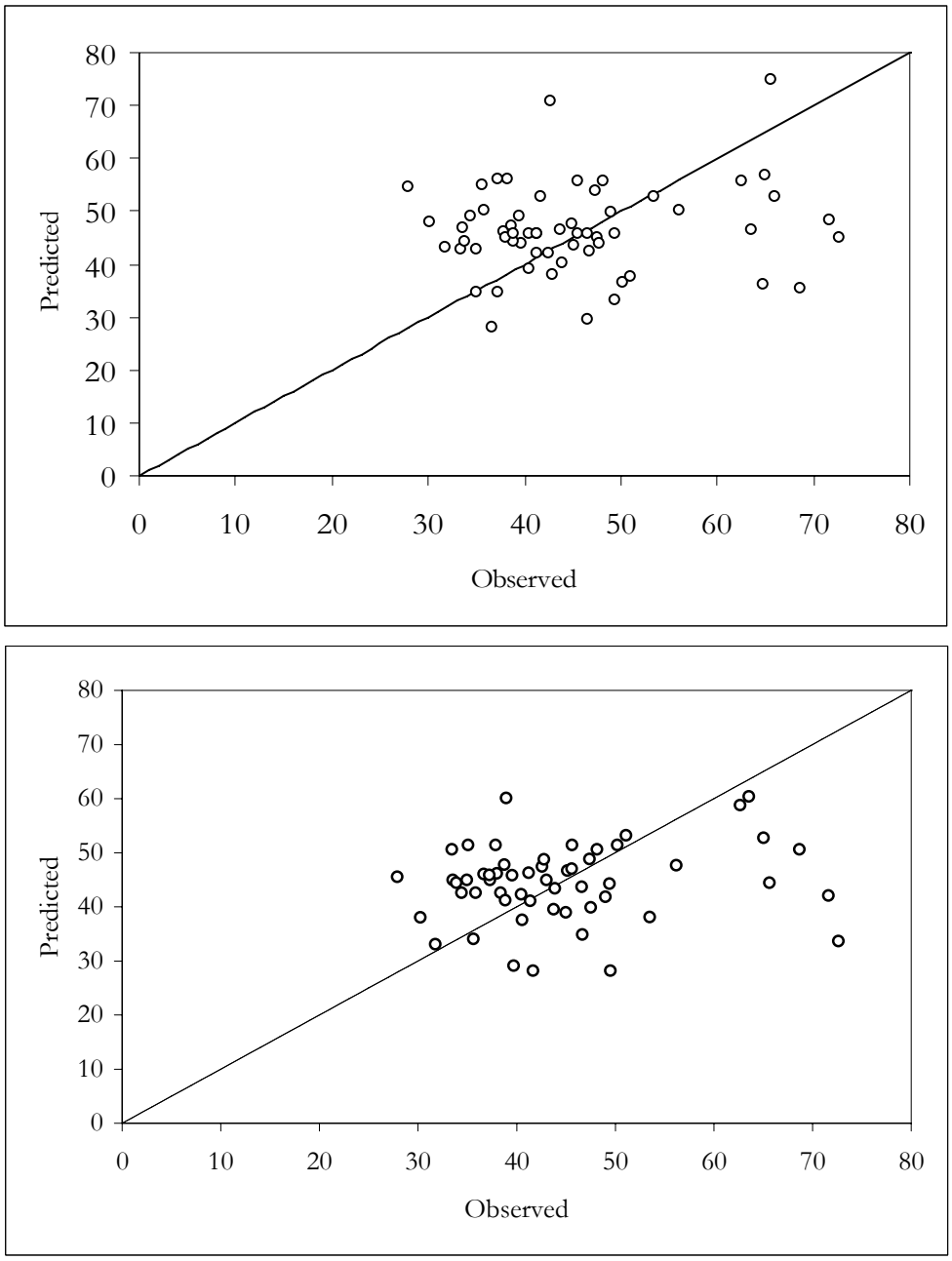


Figure 6.15. XY scatter plot of validation data sets for OSB IB (n=300) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

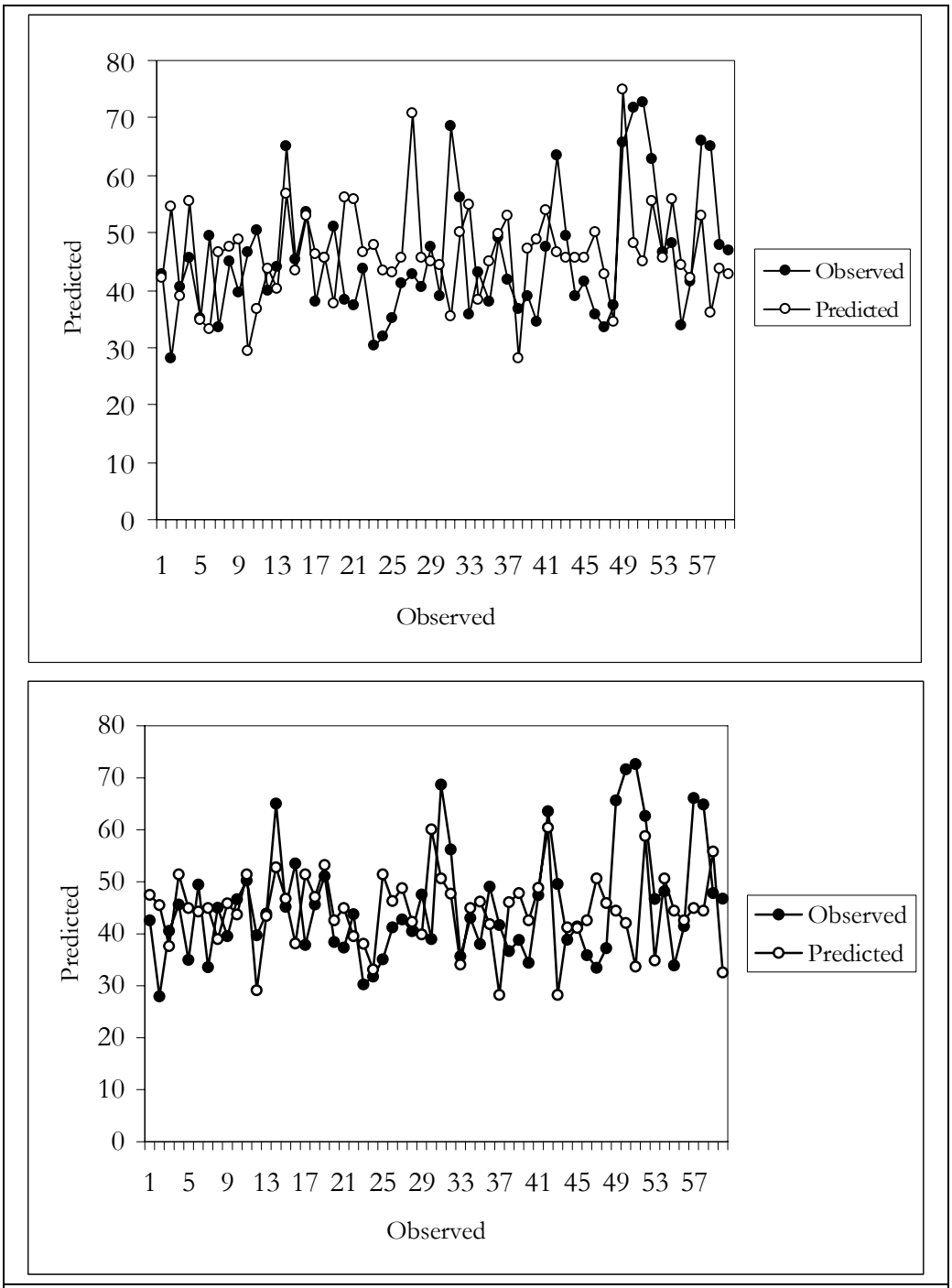


Figure 6.16. Time series graph of validation data sets for OSB IB (n=300) without Box Cox transform (top graph) and with Box Cox transform (bottom graph).

CHAPTER VII.

SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

A quote cited earlier summarizes the theme and spirit of this dissertation, "Knowledge is the most valuable asset of a manufacturing enterprise, as it enables a business to differentiate itself from competitors and to compete efficiently and effectively to the best of its ability." This statement is very appropriate for the wood composites industry, or any other industry, which exists in a business era of unprecedented international competition, increasing raw material costs, increasing energy costs and declining product prices. The results of this dissertation may directly benefit the wood composites industry in improving product quality, improving product safety, improving production efficiency and lowering costs.

This dissertation is aligned with the philosophy of data mining (DM). DM is directly related to the science of Decision Theory which is on the contemporary edge of the sciences of Artificial Intelligence, Machine Learning, Pattern Recognition and Data Visualization. Decision trees (DTs) are one of the most popular predictive learning methods used in DM. DTs were developed largely in response to the limitations of kernel methods, i.e., no matter how high the dimensionality of the predictor variable space, or how many variables are actually used for prediction (splits), the entire model is represented by a two-dimensional graphic, which can be plotted and easily interpreted. DTs also have an advantage of being very resistant to irrelevant predictor or regressor variables.

This dissertation is a comprehensive review of multiple linear regression (MLR) and regression tree (RT) models of the strength properties of medium density fiberboard (MDF)

and oriented strand board (OSB). Parametric and non-parametric (quantile regression) RT models are investigated with additional considerations for the Box Cox transform of *Y*. Process variables influencing the tensile strength properties of the internal bond (IB) for MDF and OSB are investigated. The flexure strength property of Parallel EI (EI) is also investigated for OSB.

Fused data are used for model development of destructive test data of strength properties and real-time sensor data from the production line. Appropriate time-laggings of sensor data are used in the fused database. There are 184 independent variables or regressors for MDF and 234 regressors for OSB. Three MDF product types are examined, 0.500", 0.625" and 0.750" with corresponding record lengths of 209, 245 and 517. One OSB product type of 7/16" roof-sheathing (RS) of record length 393 is examined. The data sets are from a southeastern U.S. MDF producer and a southeastern U.S. OSB producer. The distribution of the IB of MDF from the data exhibits normality. The log normal distribution is the closest fit for the distribution of the IB of OSB and the largest extreme value distribution is the closest fit for the Parallel EI of OSB.

The null research hypothesis of this dissertation is that there is no significant difference in the explanatory or predictive capabilities of three modeling methods: first-, second- and third-order MLR models with interaction terms; parametric regression trees; and non-parametric (quantile) regression trees. Given this hypothesis the dissertation has three objectives:

- Investigate the explanatory or predictive capabilities of first-, second- and third-order MLR models with interactions;
- Investigate parametric and non-parametric (quantile) regression trees models using GUDE software, version 5.2 (Loh 2006);

- Compare the explanatory and predictive capabilities of the models developed in the first and second objectives.

To satisfy these objectives a total of 1,335 models are developed and analyzed.

Acceptable models using traditional MLR are more feasible for MDF than for OSB. MLR models for the IB of MDF are more acceptable as thickness increases and record length decreases. Box Cox transforms of *Y* (IB) for MDF did not improve MLR model quality or predictive capability. Acceptable MLR models for MDF are a second-order model for 0.500" thickness for a small record length (n=60) and a first-order model for 0.625" thickness for a small record length (n=62). Significant regressors ($\alpha < 0.01$) for these MLR models are related to overall pressing time and press pre-position time settings.

One acceptable second-order MLR model is developed of the IB of OSB for a short record length (n=58) of the IB of OSB. No acceptable MLR models are feasible for the Parallel EI for OSB. Common among all statistically significant regressors for all MLR models of OSB Parallel EI are process variables related to mat forming speed which are negatively correlated with IB. The explanatory value of MLR models for the IB and Parallel EI of OSB may be more valuable to the wood composites practitioner than the predictive capability of such models. However, the practitioner should note that models with poor predictability have weak explanatory value in the regressors.

RT models have stronger explanatory value from their two-dimensional structure than MLR models. In most instances RT models outperform MLR in predictability in the validation data sets. Of the 160 RT models investigated without Box Cox transforms, 15 models have strong explanatory value and four of these 15 models have accurate predictability of the validation data sets. For MDF, process variables related to overall pressing time (negative correlation), press position times (negative correlation) and core fiber

moisture (positive correlation) are highly significant ($\alpha < 0.01$) in influencing IB. Surprisingly, the RT analysis identified a variable related to crew that is negatively correlated with the IB of MDF. Significant ($\alpha < 0.01$) process variables related to the IB and Parallel EI of OSB from RT models without Box Cox transforms are top and bottom core layers moisture content (positive correlation). Core layer forming and other forming-related process variables (negative correlation) are highly significant ($\alpha < 0.01$) in influencing the variation of the Parallel EI of OSB.

The investigation of RT models of the IB of OSB using Box Cox transforms of *Y* indicate that such transforms improve predictability for the record lengths less than 100. There is no evidence that Box Cox transforms for records lengths greater than 100 improve predictability of the IB of OSB. There is also no evidence that quantile RT models with Box Cox transforms improve predictability of OSB IB.

Significant regressors of the IB of OSB are different with and without the Box Cox transform. "Top core layer forming spreader arm speed," "press overall step movement time" and "press overall time" are highly ranked regressors with the Box Cox transform. Regressors related to "top and bottom core layer moisture contents" which are ranked as being highly important without Box Cox transforms are not ranked in the presence of the transform (recall the departure from normality in the IB of OSB). This may highlight the importance of examining the distribution of a dependent variable and the effect of the distribution on model results.

The evidence from the investigation of 1,335 models of the strength properties of MDF and OSB support the alternative research hypothesis, i.e., MLR models and RT models with and without Box Cox transforms yield different conclusions. Given that

industrial data may be highly heterogeneous and depart from normality the results indicate that MLR methods should be used with great caution. The results further indicate that RT models have high explanatory value with their tree structure and are at least as predictive as black box deterministic methods that have low explanatory value.

Results from this dissertation will hopefully advance the industrial engineering and statistical sciences as applied to wood composites manufacture. It is also hoped that the results of this dissertation will be valuable to the wood composites practitioner.

Future research should explore the development and use of regression trees in real-time wood composites manufacture. Pre-conditioning regression trees with deterministic algorithms such as genetic algorithms and neural networks needs further exploration. The combination of regression trees with principal components analysis and partial least squares methods should be studied. NIR spectroscopy methods offer capabilities of direct measurement of furnish; modeling industrial processes using NIR spectroscopy data fused with production-line sensor data in combination with RT analysis may offer fruitful research opportunities and benefits the industry.

From a scientific perspective can benefit the fields of decision theory and decision trees, and lead to advancements in decision tree software. From a practical perspective my hope is that the results will lead to improvements in MDF and OSB product quality, product safety and manufacturing costs. In the present business era, the wood composites industry cannot afford to ignore the potential benefits from the use of data mining and regression tree methods.

BIBLIOGRAPHY

Acuna, M.A. and G.E. Murphy. 2006. Use of near infrared spectroscopy and multivariate analysis to predict wood density of Douglas-fir from chain saw chips. Forest Products Journal. 56(11/12): 67-72.

Adair, C. 2005. Market outlook for the OSB and structural panel market. APA – Engineered Wood Association. Tacoma, WA.

Ahn, H. and W.-Y. Loh. 1994. Tree-structured proportional hazards regression modeling. Biometrics. 50: 471-485.

Akaike, H. 1974. Factor Analysis and AIC. Pschychometrika. 52: 317-332.

Alexander, W. P. and S. D. Grimshaw. 1996. Treed regression. Journal of Computational and Graphical Statistics. 5: 156-175.

André, N**.,** T.M. Young and T.G. Rials. 2006. On-line monitoring of the buffer capacity of particleboard furnish by near-infrared spectroscopy. Applied Spectroscopy. 60(10): 1204-1209.

André, N., H.W. Cho, S.H. Baek, M.K. Jeong and T.M. Young. 2007. Enhanced prediction of internal bond strength in a medium density fiberboard process using supervised probabilistic regression and feature selection. Wood Science and Technology. *In Press.*

Antti, H., M. Sjöström and L. Wallbäcks. 1996. Multivariate calibration models using NIR spectroscopy on pulp and paper industrial applications. Journal of Chemometrics. 10: 591-603.

Barnes, D. 2001. A model of the effect of strand length and strand thickness on the strength properties of oriented wood composites. Forest Products Journal. 51(9): 36-46.

Bassett, G. and H.-L Chen. 2001. Quantile style: return-based attribution using regression quantiles. Empirical Economics. 26(1): 293-305.

Becker, R., J. Chambers and A. Wilks. 1988. The new S language: a programming environment for data analysis and graphics. Wadsworth. Pacific Grove, CA.

Berget, I. and T. Naes. 2002. Sorting of raw materials with focus on multiple end-product properties. Journal of Chemometrics. 16: 263-273.

Bernardy and Scherff. 1998. Saving costs with process control, engineering and statistical process optimization. Proceedings 2nd European Panel Products Symposium (EPPS). Llandudno, Wales.

Bernardy and Scherff. 1999. Process modeling provides on-line quality control and process optimization in particle and fiberboard production. ATR Industrie-Elektronik GmbH&Co.KG – TextilstraBe. D-41751 Viersen Germany.

Blackman, S.S. and T.J. Broida. 1990. Multiple sensor data association and fusion in aerospace applications. Journal of Robotic Systems. 7(3): 445-485.

Box, G.E.P. 1979. Robustness in the strategy of scientific model building in Robustness in Statistics. Launter, R.L., and Wilkonsin, G.N. (eds). Academic Press. New York, NY. P. 201-236.

Box, G. 1993. Quality improvement - the new industrial revolution. International Statistical Review. 61(1): 3-19.

Box, G.E.P. and D.R. Cox. 1964. An analysis of transformations. Journal of the Royal Statistical Society. B26: 211–243.

Box, G.E.P. and A. Luceno. 1997. Statistical control by monitoring and feedback adjustment. John Wiley and Sons, Inc. New York, NY.

Bozdogan, H. 2000. Akaike's information criterion and recent developments in information complexity. Journal of Mathematical Psychology. 44(1): 62-91.

Breiman, L. 1996. Bagging predictors. Machine learning. 24: 123-140.

Breiman, L. and J.H. Friedman. 1988. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association. 80: 580-597.

Breiman, L. and J.H. Friedman. 1997. Predicting multivariate responses in multiple linear regression. Journal of the Royal Statistical Society. 59(1):3-54.

Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone. 1984. Classification and regression trees. Wadsworth, Belmont, CA.

Buchinsky, M. 1994. Changes in the U.S. wage structure 1963-1987: application of quantile regression. Econometrica. 62: 405-458.

Chamberlain, G. 1994. Quantile regression, censoring and the structure of wages. Proceedings Advances in Econometrics. Christopher Sims, ed. New York: Elsevier, p. 171–209. San Diego, Department of Economics Working Paper 99/20.

Chaudhuri, P. 1991a. Global nonparametric estimation of conditional quantile functions and their derivatives. Journal of Multivariate Analysis. 39: 246-269.

Chaudhuri, P. 1991b. Nonparametric estimates of regression quantiles and their local Bahadur representation. Annals of Statistics. 19: 760-777.

Chaudhuri, P. 2000. Asymptotic consistency of median regression trees. Journal of Statistical Planning and Inference. 91: 229-238.

Chaudhuri, P. and W.-Y. Loh. 1998. Quantile regression trees. Technical Report 994. Department of Statistics, University of Wisconsin, Madison.

Chaudhuri, P. and W.-Y. Loh. 2002. Nonparametric estimation of conditional quantiles using quantile regression trees. Bernoulli. 8(5): 561-576.

Chaudhuri, P., M.C. Huang, W.Y. Loh and R. Yao. 1994. Piecewise- regression trees. Statistica Sinica. 4: 143-167.

Chaudhuri, P., W.-D. Lo, W.-Y. Loh and C.-C. Yang. 1995. Generalized regression trees. Statistica Sinica. 5: 641-666.

Chaudhuri, P., K. Doksum and A. Samarov. 1997. On average derivative quantile regression. Annals of Statistics. 25: 715-744.

Chen, W.-W. 2005. A reliability case study on estimating extremely small percentiles of strength data for the continuous improvement of medium density fiberboard product quality. M.S. Thesis. The University of Tennessee, Knoxville.

Chen, W., R.V. León, T.M. Young and F.M. Guess. 2006. Applying a forced censoring technique with accelerated modeling for improving estimation of extremely small percentiles of strengths. International Journal of Reliability and Application. 7(1):27-39.

Cheng, K.F. 1983. Nonparametric estimators for percentile regression functions. Communications in Statistics, Theory & Methods. 12: 681-692.

Cheng, K.F. 1984. Nonparametric estimation of regression function using linear combinations of sample quantile regression function. Sankhy Series A. 46: 287-302.

Chernick, M.R. 1999. Bootstrap methods: a practitioner's guide. John Wiley and Sons, Inc. New York, NY.

Chipman, H., E.I. George and R.E. McCulloch. 1998. Bayesian CART model search. Journal of the American Statistical Association. 93: 935-960.

Cho, K. and W.-Y. Loh. 2005. Bias and convergence rate of the coverage probability of prediction intervals in Box-Cox transformed linear models. Journal of Statistical Planning and Inference. 136: 3614-3624.

Ciampi, A., S.A. Hogg, S. McKinney, and J. Thifault. 1988. RECPAM: A computer program for recursive partition and amalgamation for censored survival data and other situations frequently occurring in biostatistics, I: methods and program features. Computer Methods and Programs in Biomedicine. 26: 239-256.

Ciampi, A., Z. Lou, Q. Lin, and A. Negassa. 1991. Recursive partition and amalgamation with the exponential family: theory and applications. Applied Stochastic Models and Data Analysis. 7: 121-137.

Clapp, N.E., Jr., T.M. Young and F.M. Guess. 2007. Predictive modeling the internal bond of medium density fiberboard using principal component analysis. Forest Products Journal. *In Press.*

Composite Panel Association. 2004. 2004 North American Capacity Report. Gaithersburg, MD. 14p.

Cook, D.F. and C.C. Chiu. 1997. Predicting the internal bond strength of particleboard, utilizing a radial basis function neural network. Engineering Applications of Artificial Intelligence. 2: 171-177.

Cook, D.F. and M.L. Wolfe. 1991. Genetic algorithm approach to a lumber cutting optimization problem. Cybernetics and Systems: An International Journal. 22: 357-365.

Dabrowska, D. 1992. Nonparametric quantile regression with censored data. Sankhy Series A. 54: 252-259.

Davidson, A.C. and D.V. Hinkley. 1996. Bootstrap methods and their applications. Cambridge University Press. New York, NY.

Dawson, C., J. Allen, Jr. and T.M. Young. 2006. MIGANN ver 1.0 – machine intelligence genetic algorithm neural network system for the engineered wood industry. USDA SBIR Phase-I Final Report. 111p. with video DVD of system. Washington, D.C.

de Mast, J. and A. Trip. 2007. Exploratory data analysis in quality-improvement projects. Journal of Quality Technology. 39(4): 301-311.

Deming, W.E. 1986. Out of the crisis. Massachusetts Institute of Technology, Center for Advanced Engineering Study. Cambridge, MA.

Deming, W.E. 1993. The new economics. Massachusetts Institute of Technology, Center for Advanced Engineering Study. Cambridge, MA.

Denison D.G., B.K. Mallick and A.F.M. Smith. 1998. A Bayesian CART algorithm. Biometrika. 85: 363-377.

DiCiccio, T.J. and B. Enfron. 1996. Bootstrap confidence intervals. Statistical Science. 11: 189-228.

Dolezel-Horwath, E. T. Hutter, R. Kessler and R. Wimmer. 2005. Feedback and feedforward control of wet-processed hardboard production process using spectroscopy and chemometric modeling. Analytica Chimica Acta. 544: 47-59.

Doyle, P. 1973. The use of automatic interaction detector and similar search procedures. Operational Research Quarterly. 24: 465-467.

Draper, N.R. and H. Smith. 1981. Applied regression analysis, 2nd Ed. John Wiley and Sons, Inc. New York, NY. 709p.

Dudley, R. M. 1978. Central limit theorems for empirical measures. Annals of Probability. 6: 899-929; Corr: 7: 909-911.

Edgeworth, F. 1888. On a new method of reducing observations related to several quantities. Philosophical Magazine. 25: 184-191.

Edwards, D.J. 2004. An applied statistical reliability analysis of the internal bond of medium density fiberboard. M.S. Thesis. The University of Tennessee, Knoxville.

Edwards, D.J., F.M. Guess and T.M. Young. 2007. Bootstrap confidence intervals for percentiles of reliability of modern engineered wood. IIE Transactions. *Submitted.*

Efron, B. 1979. Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics 7: 1–26.

Efron, B. 1987. Better Bootstrap Confidence Intervals. Journal of the American Statistical Association. 82(397): 171-185.

Efron, B. 2003. Second Thoughts on the Bootstrap. Statistical Science. 18(2): 135-140.

Efron, B. and G. Gong. 1983. A leisurely look at the bootstrap, the jack-knife and cross-validation. American Statistician. 37: 36-48.

Efron, B. and R. Tibshirani. 1994. An Introduction to the Bootstrap. Chapman and Hall. New York, NY.

Efroymson, M.A. 1960. Multiple regression analysis. In Ralston, A. and Wilf, HS, editors, Mathematical Methods for Digital Computers. John Wiley and Sons, Inc. New York, NY.

Ellingwood, B.R. 1997. Probability-based LRFD for engineered wood construction. Structural Safety. 19:53-65.

Engineered Wood Association. 2004. Forecast 2004 – structural wood panels and engineered wood. http://www.apawood.org/level_b.cfm?content=sty_feature_2.

Erilsson, L., P. Hagberg, E. Johansson, S. Rannar, O. Whelehan, A. Astrom and T. Lindgren. 2000. Multivariate process monitoring of a newsprint mill. Application to modeling and predicting COD load resulting from de-inking of recycled paper. Journal of Chemometrics. 15: 337-352.

Estévez, P.A., C.A. Perez and E. Goles. 2003. Genetic input classification to a neural classifier for defect classification of radiata pine boards. Forest Products Journal. 53(7/8): 87-94.

Fan, J. and I. Gijbels. 1996. Local Modeling and Its Applications. Chapman and Hall, London.

Fan, J., T.C. Hu and Y.K. Truong. 1994. Robust nonparametric function estimation. Scandinavian Journal of Statistics 21: 433-446.

Feigenbaum, A.V. 1991. Total quality control. McGraw-Hill. New York, NY.

Fielding, A. 1977. Binary segmentation: the automatic detector and related techniques for exploring data structure. In C. A. O'Muircheartaigh and C. Payne, editors, The Analysis of Survey Data, Volume I, Exploring Data Structures. John Wiley and Sons, Inc. . New York.

Friedman, J.H. 1987. Exploratory projection pursuit. Journal of the American Statistiscal Association. 82: 249-266.

Friedman, J.H. 1989. Regularized discriminant analysis. Journal of the American Statistiscal Association. 84: 165-185.

Friedman, J.H. 1991. Multivariate adaptive regression splines. Annals of Statistics. 19: 1-141.

Friedman, J.H. 1994. An overview of computational learning and function Approximation. Proceedings from Statistics to Neural Networks. Theory and Pattern Recognition Applications. (Cherkassy, Friedman, Wechsler, eds.) Springer-Verlag 1.

Friedman, J.H. 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics. 29: 1189-1232.

Friedman, J.H. and J.W. Tukey. 1974. A projection pursuit algorithm for exploratory data analysis. IEEE Trans. Computers. C-23: 881-890.

Friedman, J.H. and W. Stuetzle. 1981. Projection pursuit regression. Journal of the American Statistiscal Association. 76: 817-823.

Friedman, J.H. and J.J. Meulman. 2003. Multiple additive regression trees with application in epidemiology. Statistics in Medicine. 22(9): 1365-1381.

Friedman, J.H. and M. Wall. 2005. Graphical views of suppression and multicollinearity in multiple linear regression. American Statistician. 59: 127-136.

Friedman, J.H., J. Bentely and R.A. Finkel. 1977. An algorithm for finding best matches in logarithmic expected time. ACM Trans. Math. Software. 3: 209-226.

Gleser, L.J. 1996. Comment on "bootstrap confidence intervals" by T.J. DiCiccio and B. Enfron. Statistical Science. 11: 219-221.

Goldratt, E.M. 1997. Critical chain. North River Press. Great Barrington, MA.

Goldratt, E.M. and J. Cox. 1984. The Goal. North River Press. Great Barrington, MA.

Goodman, I.R., R.P.S. Mahler and H.T. Nguyen. 1997. Mathematics of data fusion. Kluwer Academic Publishers. Norwell, MA.

Grant, R.M., R. Shani and R. Krishan. 1994. TQM's challenge to management theory and practice. Sloan Management Review. 35(2): 25-35.

Gruebel, D. 1999. Practical experiences with a process simulation model in particleboard and MDF production. Proceedings 2nd European Wood-based Panel Symposium. Hanover, Germany.

Guess, F.M., D.J. Edwards, T.M. Pickerell and T.M. Young. 2003. Exploring graphically and statistically the reliability of medium density fiberboard. International Journal of Reliability and Application. 4(4): 97-109.

Guess, F.M., R.V. Leon, W. Chen and T.M. Young. 2004. Forcing a closer fit in the lower tails of a distribution for better estimating extremely small percentiles of strength. International Journal of Reliability and Application. 6(4): 79-85.

Guess, F.M., X. Zhang, T.M. Young and R.V. Leon. 2005. Using mean residual life functions for unique insights into strengths of materials data. International Journal of Reliability and Application. 6(4): 79-85.

Guess, F.M., J.C. Steele, T.M. Young and R.V. León. 2006. Applying novel mean residual life confidence intervals. International Journal of Reliability and Application. 7(2): 27-39.

Hahn, G.J and W.Q. Meeker, Jr. 1983. Chemtech 13(5): 282-284.

Hall, D. 1992. Mathematical techniques in multisensor data fusion. Artech House. Norwood, MA.

Han, A. 1987. A nonparametric analysis of transformations. Journal of Econometrics. 35: 191-209.

Harding, J.A., M. Shahbaz, S. Srinivas and A. Kusiak. 2006. Data mining in manufacturing: a review. Journal of Manufacturing Science and Engineering. 128: 969-976.

Hardle, W. and T. Stoker. 1989. Investigating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association. 84: 986-995.

Harry, M.J. 1997. The vision of six sigma: a roadmap for breakthrough. Tri Star Publishing. Phoenix, AZ.

Hawkins, D.M. 1997. FIRM: Formal inference-based recursive modeling, PC version, Release 2.1. Technical Report 546, School of Statistics. University of Minnesota, MN.

Hendricks, W. and R. Koenker. 1992. Hierarchical spline model for conditional quantiles and the demand for electricity. Journal of the American Statistical Association. 87: 58-68.

Hogg, R.V. 1975. Estimates of percentile regression line using salary data. Journal of the American Statistical Association. 70: 56-59.

Hollander, M. and D.A. Wolfe. 1999. Nonparametric statistical methods. John Wiley and Sons, Inc. New York, NY.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology. 24: 417-441.

Hovanessian, S. 1980. Introduction to synthetic array and imaging radars. Artech House. Norwood, MA.

Humphrey, P.E. and H. Thoemen. 2000. The continuous pressing of wood-based composites: a simulation model, input data and typical results. In Proceedings Pacific Rim Bio-Based Composites Conference. Australian National University Press. Canberra, Australia. p. 303-311.

Inmon, W.H. and R.D. Hackathorn. 1994. Using the data warehouse. John Wiley and Sons. New York, NY.

Ishikawa, K. 1976. Guide to quality control. JUSE Press Ltd. Tokyo, Japan.

Jackson, J.E. 1991. A user's guide to principal components. John Wiley and Sons, Inc. New York, NY.

Janssen, P. and N. Veraverbeke. 1987. On nonparametric regression estimators based on regression quantiles. Communications in Statistics, Theory and Methods. 16: 383-396.

Juran, J.M. and F.M. Gryna. 1951. Juran's quality control handbook. McGraw-Hill Book Company. New York, NY.

Juran, J.M. and F.M. Gryna. 1993. Quality planning and analysis. McGraw-Hill, Inc., New York, NY.

Kaplan, E.L. and P. Meier. 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 53: 457-481.

Kass, G.V. 1975. Significance testing in automatic interaction detection (A.I.D.). Applied Statistics. 24: 178-189.

Kim, H. and W.-Y. Loh. 2001. Classification trees with unbiased multiway splits. Journal of the American Statistical Association. 96: 589-604.

Kim, H., F.M. Guess and T.M. Young. 2007a. Using data mining tools of decision trees in reliability applications. IIE Transactions. *In Press.*

Kim, H., W.-Y. Loh, Y.-S. Shih and P. Chaudhuri. 2007b. Visualizable and interpretable regression models with good prediction power. IIE Transactions. 39: 565-579.

Koenker, R. 1995. Quantile regression software. Available at http://www.econ.uiuc.edu/~roger/my.html. referenced 10/5/07.

Koenker, R. 2005. Quantile regression. Cambridge University Press. New York, NY.

Koenker, R. and G. Bassett. 1978. Regression quantiles. Econometrica. 46(1): 33–50.

Koenker, R. and V. d'Orey. 1987. Computing regression quantiles. Applied Statistics. 36: 383-393.

Koenker, R., P. Ng and S. Portnoy. 1994. Quantile smoothing splines. Biometrika 81: 673-680.

Koenker, R. and J. Machado. 1999. Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association. 94(448): 1296–310.

Koenker, R. and K. Hallock. 2000. Quantile regression: an introduction. Available at http://www.econ.uiuc.edu/~roger/my.html. referenced 10/5/07.

Koenker, R. and Y. Bilias. 2001. Quantile regression for duration data: a reappraisal of the Pennsylvania reemployment bonus experiments. Empirical Economics. 26(1): 199-220.

Koenker, R. and K.F. Hallock. 2001. Quantile regression. Journal of Economic Perspective. 15(4): 143-156.

Kruse, K. and C. Dai and A. Pielasch. 2000. An analysis of strand and horizontal density distributions in oriented strand board (OSB). Holz Als Roh-Und Werkstoff. 58(4): 270-277.

Kutner, M.H., C.J. Nachtsheim and J. Neter. 2004. Applied linear regression models. 4th Ed. McGraw-Hill Irwin, Inc. Boston, MA. 699p.

Lavine, B.K., C.E. Davidson, A.J. Moores and P.R. Griffiths. 2001. Raman spectroscopy and genetic algorithms for classification of wood types. Applied Spectroscopy. 55(8): 960-966.

Lee, M.J., A.S. Hanna and W.-Y. Loh. 2003. Decision tree approach to classify and quantify cumulative impact of change orders on productivity. Journal of Computing in Civil Engineering. 18: 132-144.

Lejeune, M.G. and P. Sarda. 1988. Quantile regression: a nonparametric approach. Computational Statistics and Data Analysis. 6: 229-239.

Levene, H. 1960. Robust tests for equality of variances. Proceedings I. Olkin, S. G. Ghurye, W. Hoefding, W. G. Madow, and H. B. Mann, editors, Contributions to Probability and Statistics, pages 278-292. Stanford University Press, Palo Alto, CA.

Li, K.-C., H.-H. Lue and C.-H. Chen. 2000. Interactive tree-structured regression via principal Hessian directions. Journal of the American Statistical Association. 95: 547-560.

Lim, T.-S., W.-Y. Loh and Y.-S. Shih. 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine Learning Journal. 40: 203-228.

Loh, W.-Y. 1984a. Bounds on AREs for restricted classes of distributions defined via tail-orderings. Annals of Statistics. 12: 685-701.

Loh, W.-Y. 1984b. Estimating an endpoint of a distribution with resampling methods. Annals of Statistics. 12: 1543-1550.

Loh, W.-Y. 1984c. Strong unimodality and scale mixtures. Annals of the Institute of Statistical Mathematics. 36: 441-449.

Loh, W.-Y. 1985. A new method for testing separate families of hypotheses. Journal of the American Statistical Association. 80: 362-368.

Loh, W.-Y. 1986a. Testing multivariate normality by simulation. Journal of Statistical Computation and Simulation. 26: 243-252.

Loh, W.-Y. 1986b. Improved estimators for ratios of variance components. Journal of the American Statistical Association. 81: 699-702.

Loh, W.-Y. 1987a. Calibrating confidence coefficients. Journal of the American Statistical Association. 82: 155-162.

Loh, W.-Y. 1987b. Does the correlation coefficient really measure the degree of clustering around a line? Journal of Educational Statistics. 12: 235-239.

Loh, W.-Y. 1989. Bounds on the size of the chi-square-test of independence in a contingency table. Annals of Statistics. 17: 1709-1722.

Loh, W.Y. 2002. Regression trees with unbiased variable selection and interaction detection. Statistica Sinica. 12: 361–386.

Loh, W.-Y. 2006a. Box-Cox transformations - II, Encyclopedia of Statistical Sciences, 2nd Edition. John Wiley and Sons, Inc. New York, NY. p. 637-644.

Loh, W.-Y. 2006b. Regression tree models for designed experiments. The Second Erich L. Lehmann Symposium - Optimality. Lecture Notes-Monograph Series. 49: 210-228. Institute of Mathematical Statistics.

Loh, W.-Y. 2007a. Regression by parts: fitting visually interpretable models with GUIDE. Handbook of Computational Statistics (Volume III) Data Visualization. Springer. *In Press*.

Loh, W.-Y. 2007b. Classification and regression tree methods. Encyclopedia of Statistics in Quality and Reliability, in press. John Wiley and Sons, Inc. New York, NY.

Loh, W.-Y. and N. Vanichsetakul. 1988. Tree-structured classification via generalized discriminant analysis. Journal of the American Statistical Association. 83: 715-728.

Loh, W.-Y. and Y.-S. Shih. 1997. Split selection methods for classification trees. Statistica Sinica. 7: 815-840.

Loh, W.-Y. and X. Yu. 1993. Bounds on the size of the likelihood ratio test of independence in a contingency table. Journal of Multivariate Analysis. 45: 291-304.

Loh, W.-Y., C.-C. Chen and W. Zheng. 2007. Extrapolation errors in linear models. ACM Transactions on Knowledge Discovery from Data. 1(2): *In Press*.

Mallows, C.L. 1973. Some comments on Cp. Technometrics. 15: 661-675

Maloney, T.M. 1993. Modern particleboard and dry-process fiberboard manufacturing. Miller Freeman, Inc. San Francisco, CA.

Manly, B.F.J. 1997. Randomization, bootstrap and Monte Carlo methods, 2nd Ed. Chapman and Hall. London, UK.

Marshall, R.J. 1995. A program to implement a search method for identification of clinical subgroups. Statistics in Medicine. 14: 2645-2659.

Meder, R., A. Thumm and D. Marston. 2003. Sawmill trial of at-line prediction of recovered lumber stiffness by NIR spectroscopy of Pinus radiata cants. Journal of Near Infrared Spectroscopy. 11: 137-143.

Meeker, W.Q. and L.A. Escobar. 1998. Statistical methods for reliability data. John Wiley and Sons, Inc. New York, NY.

Morgan, J.N. and J.A. Sunquist. 1963. Problems in the analysis of survey data and a proposal. Journal of the American Statistical Association. 58: 415-434.

Mosteller, F. 1946. On some useful 'inefficient' statistics. Annals of Mathematical Statistics. 17:377-408.

Mosteller, F. and J. Tukey. 1977. Data analysis and regression: a second course in statistics. Addison-Wesley. Reading, MA.

Myers, R.H. 1990. Classical and modern regression with applications. PWS-Kent Publishing Company. Boston, MA.

Neave, H.R. 1990. The Deming dimension. SPC Press, Inc. Knoxville, TN.

Neter, J., M.H. Kutner, C.J. Nachtsheim and W. Wasserman. 1996. Applied linear regression models. 3rd Ed. Irwin, Inc. Chicago, IL.

Newey, W.K. and T.M. Stoker. 1993. Efficiency of weighted average derivative estimators and index models. Econometrica. 61: 1199-1223.

Nishimura, T. and M.P. Ansell. 2002. Monitoring fiber orientation in OSB during production using filtered image analysis. Wood Science and Technology. 36(3): 229-239.

Perhac, D.G. 2007. An applied statistical reliability analysis of the modulus of elasticity and modulus of rupture for wood-plastic composites. M.S. Thesis. The University of Tennessee, Knoxville.

Perhac, D.G., T.M. Young, F.M. Guess and R.V. León. 2007. Exploring reliability of wood-plastic composites: stiffness and flexural strengths. International Journal of Reliability and Application. *Submitted.*

Plackett, R.L. 1960. Regression analysis. Clarendon Press. Oxford, England.

Powell, J., J. Stock and T. Stoker. 1989. Semiparametric estimation of index coefficients. Econometrica. 57: 1403-1430.

Quinlan, J. R. 1992. Learning with continuous classes. Proceedings 5th Australian Joint Conference on Artificial Intelligence. p 343–348.

Rials, T.G., S.S. Kelley and C.L. So. 2002. Use of advanced spectroscopic techniques for predicting the mechanical properties of wood composites. Wood and Fiber Science. 34(3): 398-407.

Rosowsky, D. and B. Ellingwood. 1992. Reliability of wood systems subjected to stochastic live loads. Wood and Fiber Science. 24(1):47-59.

Rosowsky, D. and P. Line. 2005. Reliability-based code calibration for design of wood members using load resistance factor design. Journal of Structural Engineering-ACSE. 131(2): 338-344.

Samarov, A. 1993. Exploring regression structure using nonparametric functional estimation. Journal of the American Statistical Association. 88: 836-849.

SAS Institute, Inc. 2007. JMP 7.0 statistics and graphics guide. SAS Institute, Inc. Cary, NC.

Scherkenbach, W.W. 1991. Deming's road to continual improvement. SPC Press, Inc. Knoxville, TN.

Senge, P.M. 1990. The fifth discipline. Doubleday. New York, NY.

Shaffer, L.B. 2007. Examining regression analysis beyond the mean of the distribution using quantile regression. M.S. Thesis. The University of Tennessee, Knoxville.

Shaffer, L.B., H. Bensmail, F.M. Guess and T.M. Young. 2006. Using R software for statistical and reliability analysis of the internal bond of medium density fiberboard. Presented at 60th International Forest Products Society Convention. Newport Beach, CA. (contact tmyoung1@utk.edu).

Shaffer, L.B., T.M. Young, F.M. Guess, H. Bensmail and R.V. León. 2007. Using R software for reliability data analysis. International Journal of Reliability and Application. *Personal communication.*

Sharp, J.G. 2004. Formaldehyde – the big issue. Proceedings 8th European Panel Products Symposium. Llandudno, UK. p. 2-1 to 2-25.

Shepherd, K.D. and M.G. Walsh. 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal. 66: 988-998.

Shepherd, K.D., C.A. Palm, C.N. Gachengo and B. Vanlauwe. 2003. Rapid characterization of organic resource quality for soil and livestock management in tropical agroecosystems using near-infrared spectroscopy. Journal of Agronomics. 65: 1314-1322.

Sherman, R. 1993. The limiting distribution of the maximum rank correlation estimator. Econometrica. 61: 123-137.

Shewhart, W.A. 1931. Economic control of quality of manufactured product. D. Van Nostrand Company. New York, NY.

Shorack, G. R. and J.A. Wellner. 1986. Empirical Processes with Applications to Statistics. John Wiley and Sons, Inc. New York, NY.

Shupe, T.F., C.Y. Price and E.W. Price. 2001. Flake orientation effects on physical and mechanical properties of sweetgum flakeboard. Forest Products Journal. 51(9): 38-43.

Sjöblom, E., B. Johnsson and H. Sundström. 2004. Optimization of particleboard production using NIR spectroscopy and multivariate techniques. Forest Products Journal. 54(6): 71-75.

So, C.-L., B.K. Via, L.H. Grooms, L.R. Scimleck, T.F. Shupe, S.S. Kelley and T.G. Rials. 2004. Near infrared spectroscopy in the forest products industry. Forest Products Journal. 54(3): 6-16.

Spake, A. 2007. Toxic trailers. Sun Herald Feb. 22, 2007 issue. Biloxi, MS.

Steele, J.C. 2006. Function domain sets confidence intervals for the mean residual life function with applications in production of medium density fiberboard product quality. M.S. Thesis. The University of Tennessee, Knoxville.

Stigler, S.M. 1986. The history of statistics – the measurement of uncertainty before 1900. The Belkamp Press of Harvard University Press. Cambridge, MA.

Suchsland, O. and G.E. Woodson. 1986. Fiberboard manufacturing practices in the United States. U.S. Department of Agriculture Forest Service. Agricultural Handbook No. 640. GPO: Washington D.C.

Taguchi, G. 1993. Taguchi on robust technology development. The American Society of Mechanical Engineers. American Society of Mechanical Engineers (ASME) Press. New York, NY.

Taylor, J. 1999. A quantile regression approach to estimating the distribution of multiperiod returns. Journal of Derivatives. 7(1): 64-78.

Taylor, S.E., D.A. Bender, D.E. Kline and K.S. Kline. 1992. Comparing length effect models for lumber tensile-strength. Forest Products Journal. 42(2): 23-30.

TECO. 2004. OSB, plywood and other structural engineered wood plants in the U.S. Personal e-mail. Madison, WI.

TECO. 2007. Formaldehyde emissions for wood based panels. Techtips. 2(6): 1.

Thorpe, J. 1998. The information paradox. McGraw-Hill. Toronto, Canada.

Toivanen, J., J.P. Hamalainen, K. Miettinen and P. Tarvainen. 2003. Designing paper machine headbox using GA. Materials and Manufacturing Processes. 18(3): 533-541.

Truong, Y.K.N. 1989. Asymptotic properties of kernel estimators based on local medians. Annals of Statistics. 17: 606-617.

Tukey, J.W. 1977. Exploratory data analysis. Addison-Wesley. Reading, MA.

U.S. Census Bureau. 2004. 2002 Economic census: table 1. Advance summary statistics for the United States 2002 NAICS basis. Washington, D.C. http://www.census.gov/econ/census02/advance/TABLE1.HTM. Referenced 10/5/07.

Van de Lindt, J.E. and D.V. Rosowsky. 2005. Strength-based reliability of wood shearwalls subject to wind load. Journal of Structural Engineering-ACSE. 131(2): 359-363.

Van de Lindt, J.W., J.N. Huart and D.V. Rosowsky. 2005. Strength-based seismic reliability of wood shear walls designed according to AF&PA/ASCE 16. Journal of Structural Engineering-ACSE. 131(8): 1307-1312.

Vapnik, V.N. and A.Y. Chervonenkis. 1971. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and Its Applications. 16: 264-280.

Venables, W.N. 2002. Modern applied statistics. S. Springer-Verlag. New York, NY.

Via, B.K. 2004. Near Infrared spectroscopy in the forest products industry. Forest Products Journal. 54(3): 27-36.

Wand, M.P. and M.C. Jones. 1995. Kernel Smoothing. Chapman and Hall, London.

Wang, Y. 2007. Reliability analysis of oriented strand board's strength with a simulation study of the median censored method for estimating of lower percentile strength. M.S. Thesis. The University of Tennessee, Knoxville.

Wang, Y., T.M. Young, F.M. Guess and R.V. León. 2007. Exploring reliability of oriented strand board's tensile and stiffness strengths. International Journal of Reliability and Application. 8(1): 111-124.

Ware, M., E. Frank, G. Holmes, M. Hall and I.H. Witten. 2001. Interactive machine learning – letting users build classifiers. International Journal of Human-Computer Studies. 55(3): 281-292.

Weibull, W. 1939. A statistical theory of the strength of materials. Ing. Vetenskaps Akad. Handl. 151(1):1-45.

Weibull, W. 1951. A statistical distribution function of wide applicability. Journal of Applied Mechanics. 18(1): 293-297.

Weibull, W. 1961. Fatigue Testing and Analysis of Results. Pergamon Press. London, UK.

Welsh, A.H. 1996. Robust estimation of smooth regression and spread functions and their derivatives. Statistica Sinica. 6: 347-366.

Wheeler, D.J. 1995. Advanced topics in statistical process control. SPC Press, Inc. Knoxville, TN.

White, C. 2002. Intelligent business strategies: real-time data warehousing heats up. DM Review Magazine (August). http://www.dmreview.com/article_sub.cfm?articleId=5570. referenced 10/5/07.

Williams, T.N. 2002. A modified six sigma quality approach to improving the quality of hardwood flooring. M.S. Thesis. The University of Tennessee, Knoxville.

Wilson, E.B. and M.M. Hilferty. 1931. The distribution of Chi-Squares. Proceedings of the National Academy of Sciences. 17: 684-688.

Woestheinrich, A. and J. Meier. 2001. Operating experiences – producing OSB with continuous presses. Proceedings 35th International Particleboard/Composite Materials Symposium. Washington State University. Pullman, WA. p. 103-108.

Wu, Q. and C. Piao. 1999. Thickness swelling and its relationship to internal bond strength loss of commercial oriented strandboard. Forest Products Journal. 49(7/8): 50-55.

Xu, W. 2000. Influence of percent alignment and shelling ratio on linear expansion of oriented strandboard: a model investigation. Forest Products Journal. 50(7/8): 88-98.

Young, T.M. 1997. Process improvement through "real-time" statistical process control in MDF manufacture. Proceedings Process and Business Technologies for the Forest Products Industry. Forest Products Society Proceedings No. 7281. p 50-51.

Young, T.M. and F.M. Guess. 1994. Reliability processes and structures. Microelectronics and Reliability. 34: 1107-1119.

Young, T.M. and F.M. Guess. 2002. Developing and mining higher quality information in automated relational databases for forest products manufacture. International Journal of Reliability and Application. 3(4): 155-164.

Young, T.M. and C.W. Huber. 2004. Predictive modeling of the physical properties of wood composites using genetic algorithms with considerations for distributed data fusion. Proceedings of the 38th International Particleboard/Composite Materials Symposium. Washington State University, Pullman, WA. p.145-153

Young, T.M., N. André and C.W. Huber. 2004. Predictive modeling of the internal bond of MDF using Genetic Algorithms with distributed data fusion. Proceedings of the 8th European Panel Products Symposium. Llandudno, UK. p. 45-59.

Young, T.M., D.G. Hodges and T.G. Rials. 2007a. The Forest Products Economy of Tennessee. Forest Products Journal. 57(4): 12-19.

Young, T.M., D.G. Perhac, F.M. Guess and R.V. León. 2007b. Bootstrap confidence intervals for percentiles of reliability data for wood plastic composites. Forest Products Journal. *Personal communication.*

Young, T.M., L.B. Shaffer, F.M. Guess, H. Bensmail and R.V. León. 2007c. A comparison of multiple linear regression and quantile regression for modeling the internal bond of medium density fiberboard. Forest Products Journal. *In Press.*

Zombori, B.G., F.A. Kamke and L.T. Watson. 2001. Simulation of the mat formation process. Wood and Fiber Science. 33(4):564-579.

GENERAL APPENDICES

APPENDIX A

Development of an automated real-time distributed data fusion system

An automated real-time distributed data fusion database was developed aligned the real-time process sensor data with the destructive testing data of the laboratory. Real-time process data were collected using Wonderware Industrial SQL 8.0 (http://www.wonderware.com/) and were combined with the laboratory test results by product type at the point in time when a panel was extracted from the production line for testing. The real-time sensor data were collected using a median of 100 sensor data records (note, sensor data varied in the rate of collection from 2 milliseconds to several seconds depending on type of sensor). Lag times, corresponding to the period of time required for the furnish to travel through the process from the point where a given parameter had an influence, to the point where the panel was extracted for destructive testing, were taken into account when collecting process data with Industrial SQL. A unique number (idnum) was generated when the panel was extracted from the process and was later used to match the process data with the lab results.

When the lab results were matched with the process data, the combined data were recorded into two tables that appear in a fused database, i.e., relational database of real-time sensor data and destructive test lab data. A Microsoft SQL table "InSQLPivot" table was created that contained the lab and process snapshot data. The "InSQLPivot_AV" table contained the lab data and median estimates of the last 100 records of real-time process data. Real-time process records are stored at rates of seconds and milliseconds. An exclusive feature of the "InSQLPivot_AV" was not only the automated alignment and fusion of real-time process data with lab data, but also the time-lagging of real-time process data. Real-time process data from on-line sensors were time-lagged according to the distance and time from the press. The time lag varied from zero seconds at the press, to several minutes at the forming and refining stage.

Data Fusion Process. -- Four tables were created within a Microsoft SQL database named "GANN." The tables in the "GANN" database were:

- InSQLPivot
- InSQLPivot_AV
- InSQLPivot_temp
- InSQLPivot_temp2.

"InSQLPivot" stored the lab and process snapshot data. "InSQLPivot" consisted of columns for both lab and process data. The field names for the lab data were preserved from the table where the lab data originated. The lab data were retrieved from "table_main" which was also located in the "GANN" database. The filed names for the process data were identical to the tag names for the respective process parameters that were stored in these fields. In addition to these columns, there also existed a "DateTime" column that is used to record the time at which the process data were collected. Process data are actually collected at three different times for a given entry into the "InSQLPivot" table ('gUnloaderDataSnapshot', 'gPressDataSnapshot' and 'gFormingDataSnapshot') the time that is recorded into the "DateTime" field corresponds to the latest of these three times

('gUnloaderDataSnapshot'). These sample tags associated with lab test time varied by test site.

"InSQLPivot" stored the lab data and the time-lagged process summary data. "InSQLPivot_AV" consisted of identical columns to the "InSQLPivot" table and accumulated data in the same manner. "InSQLPivot_temp" and "InSQLPivot_temp2" were used for temporary storage of data while the stored procedures were running. These tables should never be altered as the stored procedures will not run without them.

SQL Stored Procedures. -- Four stored procedures were created within the GANN SQL database:

- sp_Fill_InSQLPivot_From_Lab_DB
- sp_Fill_ InSQLPivot_AV_From_Lab_DB
- sp_Fill_From_eventsnapshot
- sp_Fill_From_summarydata.

The stored procedure "sp_Fill_InSQLPivot_From_Lab_DB" records a copy of the lab data into the "InSQLPivot" table. The Transact-SQL code for this procedure is available in Dawson et al. (2006). The "sp_Fill_ InSQLPivot_AV_From_Lab_DB" stored procedure records a copy of the lab data into the "InSQLPivot_AV" table. The stored procedure "sp_Fill_From_eventsnapshot" records data from the "v_eventsnapshot" view (generated by Industrial SQL) into the "InSQLPivot." The "v_eventsnapshot" data are matched to the lab data using a unique number (i.e., idnum). The stored procedure "sp_Fill_From_eventsnapshot" records data from the "v_summarydata" view (generated by Industrial SQL) into the "InSQLPivot." The "v_summarydata" data are matched to the lab data using a unique number (idnum).

SQL Data Transformation Services. -- Data transformation services (DTS) are used to output the "InSQLPivot" and "InSQLPivot_AV" tables into two comma separated variable (CSV) files ("InSQLPivot.csv" and "InSQLPivot_AV.csv"). Figure 1a, outlines the layout of this data transformation process. Data Transformation Services can be found within SQL Server Enterprise Manager. The location of the CSV files on the SQL server is D:\DB Dump\.

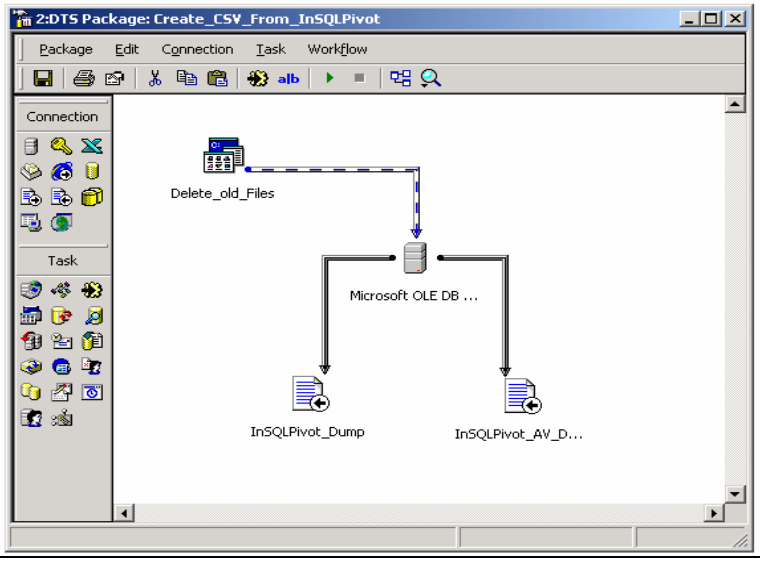


Figure 1a. Data Transformation Services window and illustration for data fusion system.

SQL Jobs. -- Three jobs were used to automate the process of inserting data into the "InSQLPivot" and "InSQLPivot_AV" tables and to generate the CSV files of these tables:

- InSQLPivot_Filler
- InSQLPivot_AV_Filler
- Create_CSVs_From_InSQLPivots.

The job "InSQLPivot_Filler" was used to execute the stored procedures "sp_Fill_InSQLPivot_From_Lab_DB" and "sp_Fill_From_eventsnapshot" every 10 minutes.

The job "InSQLPivot_AV_Filler" was also used to execute the stored procedures "sp_Fill_InSQLPivot_AV_From_Lab_DB" and "sp_Fill_From_ summarydata" every 10 minutes. The job "Create_CSVs_From_InSQLPivots" was used to execute the Data Transformation Service (DTS) Package "Create_CSVs_From_InSQLPivots" every 30 minutes to fuse the lab and process data. This DTS scheduling can be adjusted as required, as the scheduling merely dictates the frequency at which the "InSQLPivot" and "InSQLPivot_AV" tables are updated and the time(s) at which the CSV files are generated (Figure 2a) .

EventLogKey	DateTime	Comi	ProductI	TestNumber	ParallelEI	BunkerSpeed_BCL	BunkerSpeed_BSL	BunkerSpeed_TCL	BunkerSpeed_TSL	DividingRolls_BCL	DividingRolls_B
26109	12/12/2005 19:53	ok	7/16 RS	100618	58047.475	41	69	41	69	164	
26026	12/12/2005 14:03	ok	7/16 RS	100617	58653.225	39	61	39	61	164	
25973	12/12/2005 8:05	ok	7/16 RS	100616	54024.95	42	72	42	72	164	
25777	12/11/2005 19:45	ok	7/16 RS	100615	56764.7	46	73	44	73	164	
25686	12/11/2005 14:18	ok	7/16 RS	100614	55526.075	43	74	43	74	164	
25621	12/11/2005 7:30	ok	7/16 RS	100613	57665.1	43	69	43	70	164	
25322	12/10/2005 14:35	ok	7/16 RS	100610	79498.925	39	67	40	67	164	
25240	12/10/2005 8:27	ok	7/16 RS	100609	76640.95	44	71.5	44	72	164	
24977	12/9/2005 19:19	ok	7/16 RS	100608	65828.8	41	68	41	68	164	
24910	12/9/2005 13:53	ok	7/16 RS	100607	55315.675	44	61	44	61	164	
24830	12/9/2005 8:29	ok	7/16 RS	100606	61387.775	48	64	47	64	164	
24674	12/8/2005 22:16	ok	7/16 RS	100604	56287.475	37	60.5	37	61	164	
24546	12/7/2005 19:35	ok	7/16 RS	100603	63515.325	43	66	43	66	164	
24574	12/7/2005 14:36	ok	7/16 RS	100602	58926.3	44	71	44	71	164	
24545	12/7/2005 9:09	ok	19/32 RS	100601	161153.475	39	63	39	63	164	
24573	12/7/2005 2:26	ok	7/16 RS	100600	59603.325	29	48	29	48	164	
24544	12/6/2005 19:11	ok	7/16 RS	100599	55381.3	39	66	39	66	164	
24572	12/6/2005 14:18	ok	7/16 RS	100598	58865.25	46	66	46	66	164	
24543	12/6/2005 8:39	ok	7/16 RS	100597	69422.25	39	57	39	57	164	
24571	12/6/2005 0:31	ok	7/16 RS	100596	55201.25	41	68	41	68	164	
24542	12/5/2005 20:17	ok	7/16 RS	100595	57121.55	41	66	41	66	164	
24541	12/5/2005 7:17	ok	7/16 RS	100594	54550	29	45	29	45	164	
24570	12/5/2005 2:05	ok	7/16 RS	100593	56562	30.5	47	30	47	164	
24540	12/4/2005 20:33	ok	7/16 RS	100592	58170.75	33	47	33	47	164	
24569	12/4/2005 15:49	ok	7/16 RS	100591	54585.725	30	44	30	44	164	

Figure 2a. Example of automated fused database from OSB plant.

Adding a New Tag to be Stored in InSQLPivot and InSQLPivot_AV. -- New columns can be added to the "InSQLPivot" and "InSQLPivot_AV" tables to record additional snapshot and summary data process parameters. SQL Server Enterprise Manager will not facilitate the addition of extra columns to these tables, using the design view method, because they already contain such a large number of columns. Columns can however be added programmatically using SQL Query Analyzer. The following script will add a new column named "tagname" to the "InSQLPivot" and "InSQLPivot_AV" tables:

```
USE gann
BEGIN TRANSACTION
ALTER TABLE dbo.InSQLPivot ADD
        tagname float(53) NULL
ALTER TABLE dbo.InSQLPivot_AV ADD
        tagname float(53) NULL
GO
COMMIT
```

Note that the tag labeled "tagname" should already exist as an event snapshot within Industrial SQL before creating the additional column in the "InSQLPivot" and "InSQLPivot_AV" tables. The stored procedures used to automatically fill these tables will detect the addition of the new column and seek the data from the "v_eventsnapshot" and "v_summarydata" views.

APPENDIX B
Regression Tree Models in Guide Format

ILLUSTRATION 1B
MDF 0.750", n=200: Piecewise Simple Linear RT Model without node Pruning

Regression tree (75 Total Nodes, 38 Terminal Nodes):

```
Node 1: dCoreDRSP <=    7.50000
  Node 2: fShavOffT1 <=    0.13803
    Node 4: eDespMamp <=   78.55425
      Node 8: avgib-mean =  1.44500E+02
    Node 4: eDespMamp >   78.55425
      Node 9: avgib-mean =  1.39714E+02
  Node 2: fShavOffT1 >    0.13803
    Node 5: fShavOffT1 <=    0.52402
      Node 10: gFBlkDenst <=    5.00497
        Node 20: fCoreHTmpT <=  1.02565E+02
          Node 40: cSwgH20F <=   47.64060
            Node 80: avgib-mean =  1.37667E+02
          Node 40: cSwgH20F >   47.64060
            Node 81: avgib-mean =  1.43750E+02
        Node 20: fCoreHTmpT >  1.02565E+02
          Node 41: cSwOTemp <=  1.54425E+02
            Node 82: eDFld3Mamp <=  4.94517E+02
              Node 164: bFaceWFCur <=   94.12000
                Node 328: avgib-mean =  1.33400E+02
              Node 164: bFaceWFCur >   94.12000
                Node 329: avgib-mean =  1.47750E+02
            Node 82: eDFld3Mamp >  4.94517E+02
              Node 165: cSwFbrMst <=    8.01420
                Node 330: avgib-mean =  1.55600E+02
              Node 165: cSwFbrMst >    8.01420
                Node 331: fCoreMatMs <=    7.05117
                  Node 662: avgib-mean =  1.39000E+02
                Node 331: fCoreMatMs >    7.05117
                  Node 663: avgib-mean =  1.39500E+02
          Node 41: cSwOTemp >  1.54425E+02
            Node 83: dCoreGrndP <=   -1.56056
              Node 166: eDryrPrs <=    6.48971
                Node 332: avgib-mean =  1.49833E+02
              Node 166: eDryrPrs >    6.48971
                Node 333: eBoilrH20F <=   23.38240
                  Node 666: avgib-mean =  1.41500E+02
                Node 333: eBoilrH20F >   23.38240
                  Node 667: avgib-mean =  1.37000E+02
            Node 83: dCoreGrndP >   -1.56056
```

```
            Node 167: avgib-mean =  1.49857E+02
      Node 10: gFBlkDenst >    5.00497
        Node 21: cSwgScv2W <=    6.00008
          Node 42: fFaceMstM <=    7.85800
            Node 84: avgib-mean =  1.39500E+02
          Node 42: fFaceMstM >    7.85800
            Node 85: avgib-mean =  1.47500E+02
        Node 21: cSwgScv2W >    6.00008
          Node 43: aFaceBinLv <=   43.80000
            Node 86: cSwngGSF <=  2.03732E+03
              Node 172: avgib-mean =  1.48400E+02
            Node 86: cSwngGSF >  2.03732E+03
              Node 173: avgib-mean =  1.49600E+02
          Node 43: aFaceBinLv >   43.80000
            Node 87: avgib-mean =  1.49000E+02
    Node 5: fShavOffT1 >    0.52402
      Node 11: aCoreBinLv <=   47.79080
        Node 22: dCoreRTSFw <=  7.52119E+03
          Node 44: dCoreRsnS <=   12.90000
            Node 88: avgib-mean =  1.34750E+02
          Node 44: dCoreRsnS >   12.90000
            Node 89: avgib-mean =  1.19750E+02
        Node 22: dCoreRTSFw >  7.52119E+03
          Node 45: aShvAugSpd <=   31.00115
            Node 90: avgib-mean =  1.46250E+02
          Node 45: aShvAugSpd >   31.00115
            Node 91: avgib-mean =  1.42250E+02
      Node 11: aCoreBinLv >   47.79080
        Node 23: bDryFbrRt <=  1.91538E+04
          Node 46: hPrOpnTime <=   18.00415
            Node 92: avgib-mean =  1.28500E+02
          Node 46: hPrOpnTime >   18.00415
            Node 93: avgib-mean =  1.30000E+02
        Node 23: bDryFbrRt >  1.91538E+04
          Node 47: eDryESPTmp <=  7.44705E+02
            Node 94: avgib-mean =  1.41000E+02
          Node 47: eDryESPTmp >  7.44705E+02
            Node 95: avgib-mean =  1.26500E+02
Node 1: dCoreDRSP >    7.50000
  Node 3: dCoreScvWR <=    6.00034
    Node 6: hPrTempP <=  3.47248E+02
      Node 12: fShavOffT1 <=    0.37083
        Node 24: cSwgH20W <=   37.99810
          Node 48: cSwgScv2W <=    6.00009
            Node 96: avgib-mean =  1.43500E+02
          Node 48: cSwgScv2W >    6.00009
```

```
            Node 97: avgib-mean =  1.36800E+02
          Node 24: cSwgH20W >   37.99810
            Node 49: eDesp3KV <=   40.52945
              Node 98: avgib-mean =  1.41800E+02
            Node 49: eDesp3KV >   40.52945
              Node 99: avgib-mean =  1.32400E+02
        Node 12: fShavOffT1 >    0.37083
          Node 25: cSwMtrBSpd <=   33.32085
            Node 50: eDryrPrs <=    6.49700
              Node 100: avgib-mean =  1.47571E+02
            Node 50: eDryrPrs >    6.49700
              Node 101: avgib-mean =  1.34500E+02
          Node 25: cSwMtrBSpd >   33.32085
            Node 51: avgib-mean =  1.29667E+02
      Node 6: hPrTempP >  3.47248E+02
        Node 13: fHumidity <=   50.46480
          Node 26: avgib-mean =  1.38286E+02
        Node 13: fHumidity >   50.46480
          Node 27: avgib-mean =  1.12400E+02
    Node 3: dCoreScvWR >    6.00034
      Node 7: hPrTempP <=  3.46182E+02
        Node 14: avgib-mean =  1.26800E+02
      Node 7: hPrTempP >  3.46182E+02
        Node 15: aFaceBinLv <=   45.11000
          Node 30: avgib-mean =  1.35714E+02
        Node 15: aFaceBinLv >   45.11000
          Node 31: avgib-mean =  1.36286E+02
```

Node label	No. Cases cases	fit	Mat. rank	Node D-mean	Node MSE	Node R^2	Split variable
1	245	200	2	1.390E+02	2.02E+02	0.0790	dCoreDRSP
2	148	127	2	1.412E+02	2.02E+02	0.1092	fShavOffT1
4	14	13	2	1.419E+02	2.27E+02	0.6002	eDespMamp
8T	6	6	2	1.445E+02	3.57E+01	0.9569	NONE
9T	8	7	2	1.397E+02	1.08E+02	0.8111	NONE
5	134	114	2	1.411E+02	1.72E+02	0.1274	fShavOffT1
10	91	79	2	1.445E+02	1.55E+02	0.1204	gFBlkDenst
20	66	56	2	1.434E+02	1.40E+02	0.2326	fCoreHTmpT
40	10	10	2	1.401E+02	6.97E+01	0.7387	cSwgH20F
80T	6	6	2	1.377E+02	5.63E+01	0.8824	NONE
81T	4	4	2	1.438E+02	3.82E-01	0.9941	NONE
41	56	46	2	1.441E+02	1.18E+02	0.3185	cSwOTemp
82	30	25	2	1.427E+02	1.07E+02	0.4870	eDFld3Mamp
164	13	9	2	1.398E+02	6.52E+01	0.8119	bFaceWFCur
328T	5	5	2	1.334E+02	1.00E+01	0.9822	NONE
329T	8	4	2	1.478E+02	3.29E+00	0.9769	NONE

165	17	16	2	1.444E+02	7.88E+01	0.5085	cSwFbrMst
330T	6	5	2	1.556E+02	4.72E+00	0.9750	NONE
331	11	11	2	1.393E+02	3.16E+01	0.6265	fCoreMatMs
662T	5	5	2	1.390E+02	4.83E+00	0.9397	NONE
663T	6	6	2	1.395E+02	7.30E+00	0.9440	NONE
83	26	21	2	1.458E+02	7.69E+01	0.4581	dCoreGrndP
166	18	14	2	1.438E+02	6.53E+01	0.4747	eDryrPrs
332T	8	6	2	1.498E+02	8.74E+00	0.9336	NONE
333	10	8	2	1.392E+02	3.79E+01	0.6085	eBoilrH20F
666T	4	4	2	1.415E+02	2.31E-05	1.0000	NONE
667T	6	4	2	1.370E+02	7.04E-01	0.9965	NONE
167T	8	7	2	1.499E+02	1.37E+01	0.9337	NONE
21	25	23	2	1.471E+02	1.05E+02	0.3587	cSwgScv2W
42	8	8	2	1.435E+02	3.90E+01	0.8213	fFaceMstM
84T	4	4	2	1.395E+02	4.79E+00	0.9868	NONE
85T	4	4	2	1.475E+02	2.69E-01	0.9988	NONE
43	17	15	2	1.490E+02	6.17E+01	0.5942	aFaceBinLv
86	10	10	2	1.490E+02	4.21E+01	0.6885	cSwngGSF
172T	5	5	2	1.484E+02	1.42E+01	0.8771	NONE
173T	5	5	2	1.496E+02	6.01E+00	0.9753	NONE
87T	7	5	2	1.490E+02	1.44E+01	0.9519	NONE
11	43	35	2	1.334E+02	1.21E+02	0.2887	aCoreBinLv
22	19	16	2	1.358E+02	7.62E+01	0.5592	dCoreRTSFw
44	9	8	2	1.272E+02	1.92E+01	0.8936	dCoreRsnS
88T	5	4	2	1.348E+02	6.88E-01	0.9973	NONE
89T	4	4	2	1.198E+02	2.68E-01	0.9958	NONE
45	10	8	2	1.442E+02	4.25E+00	0.8596	aShvAugSpd
90T	6	4	2	1.462E+02	2.02E-01	0.9849	NONE
91T	4	4	2	1.422E+02	6.51E-02	0.9989	NONE
23	24	19	2	1.315E+02	9.08E+01	0.4916	bDryFbrRt
46	9	8	2	1.292E+02	3.45E+01	0.7473	hPrOpnTime
92T	5	4	2	1.285E+02	1.64E+01	0.9585	NONE
93T	4	4	2	1.300E+02	4.52E-01	0.9652	NONE
47	15	11	2	1.331E+02	2.68E+01	0.8877	eDryESPTmp
94T	5	5	2	1.410E+02	4.54E+00	0.9896	NONE
95T	10	6	2	1.265E+02	6.99E+00	0.8931	NONE
3	97	73	2	1.352E+02	1.63E+02	0.1418	dCoreScvWR
6	71	54	2	1.358E+02	1.46E+02	0.1774	hPrTempP
12	59	42	2	1.382E+02	1.15E+02	0.2262	fShavOffT1
24	31	21	2	1.389E+02	8.50E+01	0.3800	cSwgH20W
48	13	11	2	1.405E+02	2.99E+01	0.5060	cSwgScv2W
96T	7	6	2	1.435E+02	2.97E+00	0.9282	NONE
97T	6	5	2	1.368E+02	1.41E+01	0.8355	NONE
49	18	10	2	1.371E+02	5.01E+01	0.7997	eDesp3KV
98T	5	5	2	1.418E+02	1.38E+01	0.8848	NONE
99T	13	5	2	1.324E+02	1.31E+01	0.9723	NONE

25	28	21	2	1.375E+02	7.41E+01	0.5738 cSwMtrBSpd
50	15	15	2	1.406E+02	3.81E+01	0.6669 eDryrPrs
100T	7	7	2	1.476E+02	9.83E+00	0.9210 NONE
101T	8	8	2	1.345E+02	1.29E+01	0.6594 NONE
51T	13	6	2	1.297E+02	2.93E+01	0.9102 NONE
13	12	12	2	1.275E+02	2.93E+01	0.8701 fHumidity
26T	7	7	2	1.383E+02	4.23E+00	0.8767 NONE
27T	5	5	2	1.124E+02	3.03E+00	0.9296 NONE
7	26	19	2	1.336E+02	9.24E+01	0.6214 hPrTempP
14T	7	5	2	1.268E+02	1.15E+00	0.9982 NONE
15	19	14	2	1.360E+02	2.34E+01	0.8503 aFaceBinLv
30T	10	7	2	1.357E+02	4.08E+00	0.9609 NONE
31T	9	7	2	1.363E+02	6.78E+00	0.9750 NONE

Node 1: Intermediate node
A case goes into Node 2 if dCoreDRSP <= 7.5000000E+00
dCoreDRSP mean = 6.61138335000000

Node 2: Intermediate node
A case goes into Node 4 if fShavOffT1 <= 1.3803350E-01
fShavOffT1 mean = 0.399134422598425

Node 4: Intermediate node
A case goes into Node 8 if eDespMamp <= 7.8554250E+01
eDespMamp mean = 85.3898846153846

Node 8: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	2.1699E+02	26.88			
fShavOffT3	-1.3148E+02	-9.42	3.6447E-01	5.5131E-01	8.1370E-01

Node 9: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-3.2634E+01	-0.87			
fCore2Face	3.6277E+00	4.63	4.2000E+01	4.7509E+01	5.6000E+01

Node 5: Intermediate node
A case goes into Node 10 if fShavOffT1 <= 5.2401650E-01
fShavOffT1 mean = 0.441650622807018

Node 10: Intermediate node
A case goes into Node 20 if gFBlkDenst <= 5.0049700E+00
gFBlkDenst mean = 4.81147670886076

Node 20: Intermediate node
A case goes into Node 40 if fCoreHTmpT <= 1.0256500E+02
fCoreHTmpT mean = 111.773216071429

Node 40: Intermediate node
A case goes into Node 80 if cSwgH20F <= 4.7640600E+01
cSwgH20F mean = 48.8630500000000

Node 80: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	8.5738E+01	8.61			
cSwngMPwr	7.0067E-02	5.48	4.0568E+02	7.4113E+02	1.2023E+03

Node 81: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	4.0779E+02	28.27			
dCoreMtrBS	-4.5028E+00	-18.31	5.6742E+01	5.8640E+01	6.0275E+01

Node 41: Intermediate node
A case goes into Node 82 if cSwOTemp <= 1.5442550E+02
cSwOTemp mean = 152.331543478261

Node 82: Intermediate node
A case goes into Node 164 if eDFld3Mamp <= 4.9451650E+02
eDFld3Mamp mean = 487.753720000000

Node 164: Intermediate node
A case goes into Node 328 if bFaceWFCur <= 9.4120000E+01
bFaceWFCur mean = 94.2264111111111

Node 328: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	3.5985E+02	20.36			
bFaceBnLvl	-3.0777E+00	-12.85	6.4378E+01	7.3576E+01	8.0969E+01

Node 329: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.6391E+02	82.88			
fFaceHDP	3.8352E+01	9.20	-6.4367E-01	-4.2137E-01	-1.6484E-01

Node 165: Intermediate node
A case goes into Node 330 if cSwFbrMst <= 8.0141950E+00
cSwFbrMst mean = 8.09143812500000

Node 330: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.1232E+02	27.30			
hPrPTime	3.1742E+00	10.82	9.8917E+00	1.3633E+01	1.7342E+01

Node 331: Intermediate node
A case goes into Node 662 if fCoreMatMs <= 7.0511700E+00
fCoreMatMs mean = 7.14003000000000

Node 662: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-4.7508E+01	-1.74			
hPrTCyclP	9.1776E-01	6.83	1.9172E+02	2.0322E+02	2.1216E+02

Node 663: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.3698E+03	-7.45			
fHumidPrs	1.0448E+01	8.21	1.4333E+02	1.4446E+02	1.4547E+02

Node 83: Intermediate node
A case goes into Node 166 if dCoreGrndP <= -1.5605650E+00
dCoreGrndP mean = -1.43533238095238

Node 166: Intermediate node
A case goes into Node 332 if eDryrPrs <= 6.4897100E+00
eDryrPrs mean = 6.48609214285714

Node 332: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.9732E+03	-6.97			
hPrStmP	1.4838E+01	7.50	1.4232E+02	1.4308E+02	1.4421E+02

Node 333: Intermediate node
A case goes into Node 666 if eBoilrH20F <= 2.3382400E+01
eBoilrH20F mean = 23.4775625000000

Node 666: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-3.6107E+02	-1760.90			
aChipPrct	7.0010E+00	2451.14	7.1000E+01	7.1785E+01	7.3000E+01

Node 667: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-6.9074E+01	-7.99			
eM2236Spd	7.7037E+00	23.86	2.5000E+01	2.6750E+01	2.8000E+01

Node 167: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	7.3410E+00	0.43			
fShaveOff1	5.6036E+01	8.39	2.2193E+00	2.5433E+00	2.7424E+00

Node 21: Intermediate node

A case goes into Node 42 if cSwgScv2W <= 6.0000800E+00

cSwgScv2W mean = 6.00075739130435

Node 42: Intermediate node

A case goes into Node 84 if fFaceMstM <= 7.8580000E+00

fFaceMstM mean = 7.96750125000000

Node 84: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.1005E+05	-12.20			
cSwgScv2W	1.8367E+04	12.22	5.9980E+00	5.9990E+00	6.0000E+00

Node 85: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-2.2756E+02	-25.01			
dCoreRsnS	2.9074E+01	41.23	1.2500E+01	1.2900E+01	1.3500E+01

Node 43: Intermediate node

A case goes into Node 86 if aFaceBinLv <= 4.3800000E+01

aFaceBinLv mean = 40.9454533333333

Node 86: Intermediate node

A case goes into Node 172 if cSwngGSF <= 2.0373200E+03

cSwngGSF mean = 1926.10200000000

Node 172: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	8.9653E+01	7.00			
aChipSloLv	8.9089E-01	4.63	5.3292E+01	6.5942E+01	7.5487E+01

Node 173: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.9254E+02	47.05			
cCO0046	-1.4218E+00	-10.89	2.0000E+01	3.0200E+01	3.8000E+01

Node 87: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	3.1174E+03	8.09			
bFaceFanCr	-3.1610E+01	-7.70	9.3089E+01	9.3906E+01	9.4207E+01

Node 11: Intermediate node

A case goes into Node 22 if aCoreBinLv <= 4.7790800E+01
aCoreBinLv mean = 50.8490028571429

Node 22: Intermediate node

A case goes into Node 44 if dCoreRTSFw <= 7.5211900E+03
dCoreRTSFw mean = 8043.66562500000

Node 44: Intermediate node

A case goes into Node 88 if dCoreRsnS <= 1.2900000E+01
dCoreRsnS mean = 12.4625000000000

Node 88: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.3013E+03	-24.51			
gMxActual	4.0775E+02	27.05	3.4884E+00	3.5218E+00	3.5583E+00

Node 89: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.7194E+02	71.60			
dCoreBnLvl	-7.8809E-01	-21.86	6.1576E+01	6.6222E+01	7.8628E+01

Node 45: Intermediate node

A case goes into Node 90 if aShvAugSpd <= 3.1001150E+01
aShvAugSpd mean = 31.4754750000000

Node 90: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.1053E+04	-11.28			
cSwgScv2W	1.8665E+03	11.42	5.9980E+00	5.9999E+00	6.0017E+00

Node 91: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	3.2012E+02	78.07			
fCoreHTmpT	-1.5542E+00	-43.40	1.0889E+02	1.1444E+02	1.1780E+02

Node 23: Intermediate node

A case goes into Node 46 if bDryFbrRt <= 1.9153850E+04

bDryFbrRt mean = 19568.2052631579

Node 46: Intermediate node

A case goes into Node 92 if hPrOpnTime <= 1.8004150E+01

hPrOpnTime mean = 23.2520750000000

Node 92: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.5395E+02	36.16			
hPrPPMTimS	-1.2214E+02	-6.79	1.1667E-01	2.0833E-01	4.0000E-01

Node 93: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.8989E+02	23.60			
eBoilrStmP	-2.1343E-01	-7.45	2.6145E+02	2.8059E+02	2.9339E+02

Node 47: Intermediate node

A case goes into Node 94 if eDryESPTmp <= 7.4470450E+02

eDryESPTmp mean = 741.973272727273

Node 94: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.2610E+04	-16.73			
hPrTempP	3.6777E+01	16.91	3.4618E+02	3.4672E+02	3.4734E+02

Node 95: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.4062E+02	52.66			
cSwgOutlet	-5.6689E-03	-5.78	1.0233E+03	2.4903E+03	4.1474E+03

Node 3: Intermediate node
A case goes into Node 6 if dCoreScvWR <= 6.0003400E+00
dCoreScvWR mean = 6.00020506849315

Node 6: Intermediate node
A case goes into Node 12 if hPrTempP <= 3.4724800E+02
hPrTempP mean = 344.592055555555

Node 12: Intermediate node
A case goes into Node 24 if fShavOffT1 <= 3.7083350E-01
fShavOffT1 mean = 0.371528562619048

Node 24: Intermediate node
A case goes into Node 48 if cSwgH20W <= 3.7998100E+01
cSwgH20W mean = 35.7147809523810

Node 48: Intermediate node
A case goes into Node 96 if cSwgScv2W <= 6.0000950E+00
cSwgScv2W mean = 6.00023909090909

Node 96: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	8.9397E+02	8.56			
gMxAverage	-2.1314E+02	-7.19	3.4915E+00	3.5210E+00	3.5581E+00

Node 97: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-5.4643E+01	-1.11			
eBoilrStmP	6.8746E-01	3.90	2.6172E+02	2.7848E+02	2.8987E+02

Node 49: Intermediate node
A case goes into Node 98 if eDesp3KV <= 4.0529450E+01
eDesp3KV mean = 40.3451400000000

Node 98: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.4352E+03	5.33			
fCoreHumid	-2.7372E+01	-4.80	4.6921E+01	4.7254E+01	4.7749E+01

Node 99: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max

Constant	-6.7361E+04	-10.24			
dCoreScvWR	1.1251E+04	10.26	5.9962E+00	5.9991E+00	6.0002E+00

Node 25: Intermediate node

A case goes into Node 50 if cSwMtrBSpd <= 3.3320850E+01

cSwMtrBSpd mean = 32.0288190476190

Node 50: Intermediate node

A case goes into Node 100 if eDryrPrs <= 6.4970000E+00

eDryrPrs mean = 6.44042533333333

Node 100: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.4121E+05	7.64			
dCoreScvWR	-2.3511E+04	-7.63	5.9991E+00	6.0000E+00	6.0003E+00

Node 101: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.2805E+02	56.15			
fShavOffT3	1.1603E+01	3.41	-2.4510E-01	5.5604E-01	1.0209E+00

Node 51: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.4515E+02	44.20			
hPrPPMTimS	-1.0717E+02	-6.37	0.0000E+00	1.4444E-01	3.5000E-01

Node 13: Intermediate node

A case goes into Node 26 if fHumidity <= 5.0464800E+01

fHumidity mean = 38.9410416666667

Node 26: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.5536E+02	52.36			
cSwPltPI	-8.4100E+01	-5.96	1.2235E-01	2.0304E-01	3.0422E-01

Node 27: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	9.6705E+01	37.01			
fShavOffT1	5.3930E+01	6.29	1.9067E-01	2.9102E-01	4.0473E-01

Node 7: Intermediate node

A case goes into Node 14 if hPrTempP <= 3.4618150E+02
hPrTempP mean = 344.544052631579

Node 14: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	3.9073E+02	60.81			
eAmmonaF	-9.7070E+02	-41.19	2.3511E-01	2.7189E-01	2.9571E-01

Node 15: Intermediate node

A case goes into Node 30 if aFaceBinLv <= 4.5110000E+01
aFaceBinLv mean = 45.4139357142857

Node 30: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.6606E+02	58.40			
dCoreRSSpd	-5.6235E-01	-11.08	2.4913E+01	5.3971E+01	6.9760E+01

Node 31: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.2407E+04	14.12			
hPrTempP	-3.5373E+01	-13.96	3.4630E+02	3.4690E+02	3.4740E+02

ILLUSTRATION 2B
OSB IB 7/16" RS, n=100: Mixed Stepwise RT Model for All Possible Subsets
No Prune

Regression Tree (23 Total Nodes, 12 terminal nodes):

Node 1: MTCLMoiLev <= 5.18742
  Node 2: MSTSLOFSpA <= 54.90909
    Node 4: IB-mean = 48.91745
  Node 2: MSTSLOFSpA > 54.90909
    Node 5: MF3PasCont <= 2.49750E+02
      Node 10: MSBCLOFDSP <= 6.17598
        Node 20: IB-mean = 38.23089
      Node 10: MSBCLOFDSP > 6.17598
        Node 21: IB-mean = 46.54200
    Node 5: MF3PasCont > 2.49750E+02
      Node 11: PMI736 <= 13.00000
        Node 22: PMI746 <= 2.52000E+03
          Node 44: Dry1Out <= 2.53500E+02
            Node 88: IB-mean = 40.12000
          Node 44: Dry1Out > 2.53500E+02
            Node 89: IB-mean = 43.34900
        Node 22: PMI746 > 2.52000E+03
          Node 45: IB-mean = 40.03717
      Node 11: PMI736 > 13.00000
        Node 23: IB-mean = 45.96050
Node 1: MTCLMoiLev > 5.18742
  Node 3: MTCLMoiLev <= 5.41640
    Node 6: IB-mean = 48.54244
  Node 3: MTCLMoiLev > 5.41640
    Node 7: PrsClosTim <= 56.50000
      Node 14: PZI740 <= 71.50000
        Node 28: WaxRatCL <= 3.16349E+02
          Node 56: IB-mean = 54.90971
        Node 28: WaxRatCL > 3.16349E+02
          Node 57: IB-mean = 52.45867
      Node 14: PZI740 > 71.50000
        Node 29: IB-mean = 47.20767
    Node 7: PrsClosTim > 56.50000
      Node 15: IB-mean = 61.05073

Node label	No. Cases cases	fit	Mat. rank	Node D-mean	Node MSE	Node R^2	Split variable
1	194	101	3	4.787E+01	6.98E+01	0.2648	MTCLMoiLev
2	128	59	3	4.385E+01	4.08E+01	0.2156	MSTSLOFSpA
4T	13	11	3	4.892E+01	1.44E+01	0.6687	NONE
5	115	48	3	4.269E+01	3.86E+01	0.2170	MF3PasCont
10	38	17	3	4.214E+01	3.86E+01	0.4308	MSBCLOFDSP

20T	9	9	3	3.823E+01	2.22E+01	0.3330	NONE
21T	29	8	3	4.654E+01	4.66E+01	0.4883	NONE
11	77	31	3	4.299E+01	2.23E+01	0.5053	PMI736
22	47	19	3	4.111E+01	2.60E+01	0.3483	PMI746
44	29	13	3	4.161E+01	3.60E+01	0.2180	Dry1Out
88T	21	7	2	4.012E+01	3.42E+01	0.0559	NONE
89T	8	6	2	4.335E+01	6.13E+01	0.0000	NONE
45T	18	6	2	4.004E+01	0.00E+00	1.0000	NONE
23T	30	12	3	4.596E+01	1.22E+01	0.7553	NONE
3	66	42	3	5.353E+01	5.70E+01	0.4569	MTCLMoiLev
6T	16	9	3	4.854E+01	2.23E+01	0.7368	NONE
7	50	33	3	5.489E+01	5.67E+01	0.4850	PrsClosTim
14	33	22	3	5.181E+01	4.44E+01	0.5803	PZI740
28	25	16	3	5.353E+01	5.18E+01	0.4860	WaxRatCL
56T	9	7	2	5.491E+01	1.56E+02	0.0411	NONE
57T	16	9	3	5.246E+01	5.68E+01	0.2761	NONE
29T	8	6	2	4.721E+01	7.47E+01	0.4292	NONE
15T	17	11	3	6.105E+01	4.24E+01	0.4905	NONE

Regression Tree Least Squares Functions:

Node 1: Intermediate node
A case goes into Node 2 if MTCLMoiLev <= 5.1874247E+00
MTCLMoiLev mean = 5.13977918654455

Node 2: Intermediate node
A case goes into Node 4 if MSTSLOFSpA <= 5.4909090E+01
MSTSLOFSpA mean = 60.6910631530508

Node 4: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	4.4506E+01	13.16			
BnkSpdTCL	-2.8398E+00	-3.89	1.1000E+01	2.8727E+01	4.6000E+01
BnkSpdTSL	2.0342E+00	4.00	1.4000E+01	4.2273E+01	6.8000E+01

Node 5: Intermediate node
A case goes into Node 10 if MF3PasCont <= 2.4975000E+02
MF3PasCont mean = 406.708333333333

Node 10: Intermediate node
A case goes into Node 20 if MSBCLOFDSP <= 6.1759762E+00
MSBCLOFDSP mean = 6.09058882200000

Node 20: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	4.3041E+01	1.31			
BnkSpdBSL	-1.1550E+00	-1.44	5.6000E+01	6.4889E+01	7.3000E+01
BnkSpdTCL	1.5940E+00	0.95	4.0000E+01	4.4000E+01	4.7000E+01

Node 21: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.6141E+02	2.38			
BnkSpdBCL	-2.0793E-01	-0.08	4.1000E+01	4.3688E+01	4.6000E+01
BnkSpdBSL	-1.5166E+00	-1.29	6.3000E+01	6.9750E+01	7.5000E+01

Node 11: Intermediate node

A case goes into Node 22 if PMI736 <= 1.3000000E+01

PMI736 mean = 6.48064506341935

Node 22: Intermediate node

A case goes into Node 44 if PMI746 <= 2.5200000E+03

PMI746 mean = 2226.36842105263

Node 44: Intermediate node

A case goes into Node 88 if Dry1Out <= 2.5350000E+02

Dry1Out mean = 250.500000000000

Node 88: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	4.1500E+01	12.34

Node 89: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	4.3349E+01	16.60

Node 45: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	4.0037E+01	18.47

Node 23: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	6.0536E+01	18.22			
DyBiLBoBSL	-7.5753E-01	-2.96	0.0000E+00	4.1708E+01	5.4500E+01
DyBiLBoCL	2.7232E-01	1.57	0.0000E+00	6.2500E+01	7.5000E+01

Node 3: Intermediate node
A case goes into Node 6 if MTCLMoiLev <= 5.4163983E+00
MTCLMoiLev mean = 5.93038741752381

Node 6: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	5.7884E+01	8.21			
BnkSpdBSL	2.0537E+00	3.91	3.0000E+01	6.0111E+01	7.5000E+01
BnkSpdTCL	-3.4049E+00	-4.08	1.9000E+01	3.9000E+01	4.6000E+01

Node 7: Intermediate node
A case goes into Node 14 if PrsClosTim <= 5.6500000E+01
PrsClosTim mean = 55.0000000000000

Node 14: Intermediate node
A case goes into Node 28 if PZI740 <= 7.1500000E+01
PZI740 mean = 57.3181818181818

Node 28: Intermediate node
A case goes into Node 56 if WaxRatCL <= 3.1634933E+02
WaxRatCL mean = 727.071659205125

Node 56: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	5.6837E+01	9.02

Node 57: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	6.5805E+01	5.21			
BnkSpdBSL	6.4793E-01	0.92	2.4000E+01	5.5333E+01	6.6000E+01
BnkSpdTCL	-1.3139E+00	-1.25	1.8000E+01	3.7444E+01	4.7000E+01

Node 29: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	4.7208E+01	12.38

Node 15: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	7.4357E+01	9.97			
BnkSpdBCL	-1.8369E+00	-2.35	1.8000E+01	3.5364E+01	4.6000E+01
BnkSpdTSL	9.5415E-01	1.89	2.8000E+01	5.4136E+01	6.9000E+01

ILLUSTRATION 3B
OSB IB 7/16" RS, n=200: Mixed Stepwise RT Model for All Possible Subsets
No Prune

Regression tree (33 Total Nodes, 17 Terminal Nodes):

Node 1: PKI700AC1T <= 16.50000
  Node 2: MTCLMoiLev <= 4.84696
    Node 4: PPI700A <= 9.32500E+02
      Node 8: MD5OutTem <= 2.47250E+02
        Node 16: IB-mean = 41.18314
      Node 8: MD5OutTem > 2.47250E+02
        Node 17: IB-mean = 48.11538
    Node 4: PPI700A > 9.32500E+02
      Node 9: IB-mean = 40.13382
  Node 2: MTCLMoiLev > 4.84696
    Node 5: MDBBSLLev <= 62.50000
      Node 10: MBBSLSPFB <= 2.59241E+04
        Node 20: IB-mean = 63.61629
      Node 10: MBBSLSPFB > 2.59241E+04
        Node 21: PTI796C <= 2.23050E+02
          Node 42: IB-mean = 56.67327
        Node 21: PTI796C > 2.23050E+02
          Node 43: MBTRFTP <= 5.48991E+03
            Node 86: IB-mean = 52.34825
          Node 43: MBTRFTP > 5.48991E+03
            Node 87: IB-mean = 43.68167
    Node 5: MDBBSLLev > 62.50000
      Node 11: MBBSLRFTC <= 6.72257E+02
        Node 22: IB-mean = 47.57145
      Node 11: MBBSLRFTC > 6.72257E+02
        Node 23: IB-mean = 44.68070
Node 1: PKI700AC1T > 16.50000
  Node 3: PPC741B <= 97.00000
    Node 6: MFLSpdAPer <= 97.81804
      Node 12: PZI701St17 <= 3.69950E+03
        Node 24: IB-mean = 44.35308
      Node 12: PZI701St17 > 3.69950E+03
        Node 25: IB-mean = 44.52422
    Node 6: MFLSpdAPer > 97.81804
      Node 13: BnkSpdBSL <= 66.50000
        Node 26: IB-mean = 38.16686
      Node 13: BnkSpdBSL > 66.50000
        Node 27: PKI700SFOY <= 16.50000
          Node 54: IB-mean = 34.23363
        Node 27: PKI700SFOY > 16.50000
          Node 55: PKI700SFOY <= 18.50000
            Node 110: IB-mean = 44.96364
          Node 55: PKI700SFOY > 18.50000
            Node 111: IB-mean = 38.53143

Node 3: PPC741B > 97.00000
Node 7: MBWoFCV <= 31.50000
Node 14: IB-mean = 50.91569
Node 7: MBWoFCV > 31.50000
Node 15: IB-mean = 47.26390

Node label	No. Cases cases	fit	Mat. rank	Node D-mean	Node MSE	Node R^2	Split variable
1	293	200	3	4.571E+01	6.30E+01	0.2445	PKI700AC1T
2	175	111	3	4.688E+01	6.52E+01	0.2952	MTCLMoiLev
4	83	48	3	4.245E+01	3.84E+01	0.3597	PPI700A
8	34	20	3	4.569E+01	3.49E+01	0.3062	MD5OutTem
16T	9	7	2	4.118E+01	3.82E+01	0.1422	NONE
17T	25	13	3	4.812E+01	2.79E+01	0.3249	NONE
9T	49	28	3	4.013E+01	2.94E+01	0.5062	PKI700SFOY
5	92	63	3	5.026E+01	5.53E+01	0.4104	MDBBSLLev
10	54	32	3	5.467E+01	5.60E+01	0.4504	MBBSLSPFB
20T	11	7	2	6.362E+01	4.33E+01	0.0519	NONE
21	43	25	3	5.217E+01	4.48E+01	0.5099	PTI796C
42T	28	11	3	5.667E+01	3.03E+01	0.7076	NONE
43	15	14	3	4.863E+01	5.62E+01	0.2119	MBTRFTP
86T	9	8	3	5.235E+01	1.70E+01	0.7088	NONE
87T	6	6	2	4.368E+01	5.90E+01	0.0000	NONE
11	38	31	3	4.571E+01	2.69E+01	0.4654	MBBSLRFTC
22T	13	11	3	4.757E+01	1.67E+01	0.8518	NONE
23T	25	20	3	4.468E+01	1.54E+01	0.4163	PMI732
3	118	89	3	4.424E+01	5.61E+01	0.2064	PPC741B
6	68	55	3	4.150E+01	3.64E+01	0.2924	MFLSpdAPer
12	32	22	3	4.442E+01	3.68E+01	0.3933	PZI701St17
24T	13	13	3	4.435E+01	3.09E+01	0.6145	ResiRatCL
25T	19	9	3	4.452E+01	4.04E+01	0.3075	NONE
13	36	33	3	3.956E+01	2.51E+01	0.3754	BnkSpdBSL
26T	8	7	2	3.817E+01	0.00E+00	1.0000	NONE
27	28	26	3	3.993E+01	2.34E+01	0.5071	PKI700SFOY
54T	9	8	3	3.423E+01	3.82E+00	0.5899	NONE
55	19	18	3	4.246E+01	2.37E+01	0.4679	PKI700SFOY
110T	11	11	3	4.496E+01	1.96E+01	0.6321	NONE
111T	8	7	2	3.853E+01	0.00E+00	1.0000	NONE
7	50	34	3	4.866E+01	5.75E+01	0.2351	MBWoFCV
14T	17	13	3	5.092E+01	4.69E+01	0.2216	Dr1OutMois
15T	33	21	3	4.726E+01	3.39E+01	0.6233	PTI796D

Node 1: Intermediate node
A case goes into Node 2 if PKI700AC1T <= 1.6500000E+01
PKI700AC1T mean = 16.4800000000000

Node 2: Intermediate node
A case goes into Node 4 if MTCLMoiLev <= 4.8469572E+00
MTCLMoiLev mean = 4.99087761439640

Node 4: Intermediate node
A case goes into Node 8 if PPI700A <= 9.3250000E+02
PPI700A mean = 1800.72916666667

Node 8: Intermediate node
A case goes into Node 16 if MD5OutTem <= 2.4725000E+02
MD5OutTem mean = 251.375000000000

Node 16: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	4.2906E+01	14.27

Node 17: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	5.1340E+01	5.19			
BnkSpdBCL	-9.9614E-01	-2.17	1.9000E+01	4.2692E+01	4.9000E+01
DyBiLBoCL	5.6489E-01	1.79	3.9000E+01	6.9577E+01	8.0000E+01

Node 9: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	7.6142E+01	9.98			
DryWeBin2	-7.2477E-01	-2.67	1.1000E+01	2.4143E+01	2.8000E+01
DryWeBin4	-8.2400E-01	-3.91	1.2000E+01	2.2464E+01	2.7000E+01

Node 5: Intermediate node
A case goes into Node 10 if MDBBSLLev <= 6.2500000E+01
MDBBSLLev mean = 55.5555555555556

Node 10: Intermediate node
A case goes into Node 20 if MBBSLSPFB <= 2.5924136E+04
MBBSLSPFB mean = 33217.3470691563

Node 20: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	6.3662E+01	25.58

Node 21: Intermediate node
A case goes into Node 42 if PTI796C <= 2.2305000E+02
PTI796C mean = 221.751996704000

Node 42: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	7.3835E+01	15.32			
BnkSpdTCL	-2.8238E+00	-2.99	1.6000E+01	3.3500E+01	4.7000E+01
BnkSpdTSL	1.5658E+00	2.50	2.3000E+01	4.9455E+01	6.9000E+01

Node 43: Intermediate node
A case goes into Node 86 if MBTRFTP <= 5.4899058E+03
MBTRFTP mean = 5293.54849678571

Node 86: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-5.6765E+00	-0.09			
BnkSpdBCL	-8.4508E-01	-0.70	4.1000E+01	4.3250E+01	4.5000E+01
BnkSpdBSL	1.3921E+00	3.40	6.2000E+01	6.7938E+01	7.5000E+01

Node 87: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	4.3682E+01	17.07

Node 11: Intermediate node
A case goes into Node 22 if MBBSLRFTC <= 6.7225726E+02
MBBSLRFTC mean = 1783.75080588774

Node 22: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	4.4245E+01	11.19			
BnkSpdBCL	-1.8727E+01	-6.78	0.0000E+00	3.9273E+01	4.7000E+01
BnkSpdTCL	1.8833E+01	6.78	0.0000E+00	3.9227E+01	4.7000E+01

Node 23: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	7.1426E+01	4.10			
BnkSpdBCL	-9.5509E-01	-2.66	3.8000E+01	4.3250E+01	4.7000E+01
Dry2Out	5.9232E-02	2.10	1.2200E+02	2.4585E+02	2.8300E+02

Node 3: Intermediate node
A case goes into Node 6 if PPC741B <= 9.7000000E+01
PPC741B mean = 51.9325842696629

Node 6: Intermediate node
A case goes into Node 12 if MFLSpdAPer <= 9.7818043E+01
MFLSpdAPer mean = 88.3700883374545

Node 12: Intermediate node
A case goes into Node 24 if PZI701St17 <= 3.6995000E+03
PZI701St17 mean = 3699.36363636364

Node 24: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	6.5949E+01	9.68			
BnkSpdTSL	-5.5024E-01	-3.93	2.7000E+01	5.6308E+01	7.1000E+01
DyBiLBoTSL	3.0933E-01	2.76	0.0000E+00	3.0346E+01	7.0000E+01

Node 25: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	4.9459E+01	7.77			
BnkSpdBSL	1.5376E+00	1.19	1.4000E+01	4.8333E+01	6.5000E+01
BnkSpdTCL	-2.4016E+00	-1.29	1.1000E+01	3.3000E+01	4.6000E+01

Node 13: Intermediate node
A case goes into Node 26 if BnkSpdBSL <= 6.6500000E+01
BnkSpdBSL mean = 68.5454545454545

Node 26: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	3.8167E+01	26.81

Node 27: Intermediate node
A case goes into Node 54 if PKI700SFOY <= 1.6500000E+01
PKI700SFOY mean = 17.6153846153846

Node 54: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	9.0247E+01	4.32			
BnkSpdBCL	-2.6161E-01	-0.98	4.3000E+01	4.5750E+01	5.0000E+01
BnkSpdBSL	-6.1979E-01	-2.24	6.7000E+01	7.1062E+01	7.6000E+01

Node 55: Intermediate node
A case goes into Node 110 if PKI700SFOY <= 1.8500000E+01
PKI700SFOY mean = 18.5000000000000

Node 110: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.8288E+02	3.50			
BnkSpdBSL	-2.6262E+00	-3.49	6.7000E+01	6.9000E+01	7.2000E+01
DyBiLBoBSL	8.3554E-01	1.95	4.7000E+01	5.1818E+01	5.7000E+01

Node 111: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat
Constant	3.8531E+01	33.20

Node 7: Intermediate node
A case goes into Node 14 if MBWoFCV <= 3.1500000E+01
MBWoFCV mean = 34.8676470588235

Node 14: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	3.3700E+01	3.20			
DyBiLBoBSL	3.1929E-02	0.38	0.0000E+00	3.2462E+01	8.2000E+01
DyBiLBoCL	2.7262E-01	1.53	3.4000E+01	5.9346E+01	8.0000E+01

Node 15: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.4491E+02	7.71			
BnkSpdTCL	-1.3599E+00	-3.28	3.3000E+01	4.2476E+01	4.7000E+01
Dry3Out	-1.6252E-01	-3.56	1.2700E+02	2.4540E+02	2.7450E+02

ILLUSTRATION 4B
OSB Parallel EI 7/16" RS, n=100: Third-order RT Model No Prune

Regression tree (21 Total Nodes, 11 Terminal Nodes):

Node 1: PMI747 <= 2.37000E+03
  Node 2: Flk2StoSpd <= 3.50000
    Node 4: WeBi5Tot24 <= 1.39800E+03
      Node 8: MBTRFTP <= 4.91516E+03
        Node 16: ParlEI-mean = 5.70204E+04
      Node 8: MBTRFTP > 4.91516E+03
        Node 17: ParlEI-mean = 5.80785E+04
    Node 4: WeBi5Tot24 > 1.39800E+03
      Node 9: ParlEI-mean = 6.08591E+04
  Node 2: Flk2StoSpd > 3.50000
    Node 5: Flk3StoSpd <= 1.50000
      Node 10: ParlEI-mean = 5.92819E+04
    Node 5: Flk3StoSpd > 1.50000
      Node 11: ParlEI-mean = 6.30303E+04
Node 1: PMI747 > 2.37000E+03
  Node 3: MSTCLLev <= 54.50000
    Node 6: PPI740 <= 87.05000
      Node 12: PTI793 <= 1.64500E+02
        Node 24: ParlEI-mean = 5.76362E+04
      Node 12: PTI793 > 1.64500E+02
        Node 25: ParlEI-mean = 5.87825E+04
    Node 6: PPI740 > 87.05000
      Node 13: PTI796A <= 1.67550E+02
        Node 26: ParlEI-mean = 5.97514E+04
      Node 13: PTI796A > 1.67550E+02
        Node 27: ParlEI-mean = 6.01949E+04
  Node 3: MSTCLLev > 54.50000
    Node 7: MBBSLWFCV <= 48.50000
      Node 14: ParlEI-mean = 5.77287E+04
    Node 7: MBBSLWFCV > 48.50000
      Node 15: ParlEI-mean = 5.68532E+04

Node label	No. Cases cases	fit	Mat. rank	Node D-mean	Node MSE	Node Split R^2	variable
1	194	101	4	5.898E+04	1.17E+07	0.1608	PMI747
2	95	49	4	5.950E+04	1.14E+07	0.3688	Flk2StoSpd
4	68	31	4	5.853E+04	8.15E+06	0.4092	WeBi5Tot24
8	43	21	4	5.742E+04	2.91E+06	0.7504	MBTRFTP
16T	17	13	4	5.702E+04	2.10E+06	0.8496	NONE
17T	26	8	4	5.808E+04	8.22E+05	0.9511	NONE
9T	25	10	4	6.086E+04	3.38E+06	0.7850	NONE
5	27	18	4	6.116E+04	1.22E+07	0.5244	Flk3StoSpd
10T	10	9	4	5.928E+04	9.91E+05	0.9749	NONE

11T	17	9	4	6.303E+04	2.12E+06	0.8924	NONE
3	99	52	4	5.849E+04	8.99E+06	0.1695	MSTCLLev
6	70	36	4	5.902E+04	6.69E+06	0.3186	PPI740
12	53	19	4	5.818E+04	4.18E+06	0.6255	PTI793
24T	38	10	4	5.764E+04	1.66E+06	0.9076	NONE
25T	15	9	4	5.878E+04	8.48E+05	0.9202	NONE
13	17	17	4	5.996E+04	.06E+06	0.6636	PTI796A
26T	9	9	4	5.975E+04	5.48E+05	0.9684	NONE
27T	8	8	4	6.019E+04	5.55E+05	0.9280	NONE
7	29	16	4	5.729E+04	5.45E+06	0.6206	MBBSLWFCV
14T	15	8	4	5.773E+04	1.78E+06	0.9491	NONE
15T	14	8	4	5.685E+04	4.81E+05	0.9347	NONE

Node 1: Intermediate node
A case goes into Node 2 if PMI747 <= 2.3700000E+03
PMI747 mean = 2141.93069306931

Node 2: Intermediate node
A case goes into Node 4 if Flk2StoSpd <= 3.5000000E+00
Flk2StoSpd mean = 2.71428571428571

Node 4: Intermediate node
A case goes into Node 8 if WeBi5Tot24 <= 1.3980000E+03
WeBi5Tot24 mean = 1091.41935483871

Node 8: Intermediate node
A case goes into Node 16 if MBTRFTP <= 4.9151611E+03
MBTRFTP mean = 4567.32204978095

Node 16: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	4.2123E+04	14.14			
Dry1In	1.6849E+02	4.41	8.1000E+01	6.6131E+02	1.1490E+03
Dry1In^2	-2.9544E-01	-3.93			
Dry1In^3	1.3999E-04	3.60			

Lower and upper truncation bounds = 5.2122E+04 6.4638E+04

Node 17: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.9819E+06	-7.23			
Dry5Out	2.6159E+04	7.37	1.8500E+02	2.4794E+02	2.8600E+02
Dry5Out^2	-1.0983E+02	-7.29			
Dry5Out^3	1.5137E-01	7.19			

Lower and upper truncation bounds = 5.2842E+04 6.2945E+04

Node 9: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.9186E+06	-4.06			
MBBSLSP	1.4103E+04	4.13	3.4500E+02	4.1355E+02	5.0150E+02
MBBSLSP^2	-3.3151E+01	-4.06			
MBBSLSP^3	2.5725E-02	3.99			

Lower and upper truncation bounds = 5.5501E+04 6.7853E+04

Node 5: Intermediate node
A case goes into Node 10 if Flk3StoSpd <= 1.5000000E+00
Flk3StoSpd mean = 1.66666666666667

Node 10: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-1.8984E+07	-5.52			
MSBSLOFDSP	6.3933E+06	5.59	8.2360E+00	8.7784E+00	1.0009E+01
MSBSLOFDSP^2	-7.1407E+05	-5.65			
MSBSLOFDSP^3	2.6529E+04	5.71			

Lower and upper truncation bounds = 5.3487E+04 7.3270E+04

Node 11: Terminal node
Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-2.3549E+06	-4.42			
MTCLMoiLev	1.3712E+06	4.62	4.5484E+00	5.2418E+00	6.4819E+00
MTCLMoiLev^2	-2.5710E+05	-4.72			
MTCLMoiLev^3	1.5930E+04	4.81			

Lower and upper truncation bounds = 5.7786E+04 7.0383E+04

Node 3: Intermediate node
A case goes into Node 6 if MSTCLLev <= 5.4500000E+01
MSTCLLev mean = 52.9807692307692

Node 6: Intermediate node
A case goes into Node 12 if PPI740 <= 8.7050000E+01
PPI740 mean = 86.8166666666666

Node 12: Intermediate node
A case goes into Node 24 if PTI793 <= 1.6450000E+02
PTI793 mean = 163.315789473684

Node 24: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	1.0076E+07	7.99			
MFLCorWegt	-1.0251E+06	-7.92	1.8936E+01	1.9322E+01	2.0027E+01
MFLCorWegt^2	2.6211E+04	7.90			

Lower and upper truncation bounds = 5.3016E+04 6.5936E+04

Node 25: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	8.7677E+04	16.66			
MF2PasCont	-2.0020E+02	-5.51	1.9950E+02	4.6861E+02	9.3200E+02
MF2PasCont^2	3.7675E-01	5.26			
MF2PasCont^3	-2.0503E-04	-4.90			

Lower and upper truncation bounds = 5.4595E+04 6.3625E+04

Node 13: Intermediate node

A case goes into Node 26 if PTI796A <= 1.6755000E+02

PTI796A mean = 167.088232800000

Node 26: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	2.6157E+06	7.63			
ResiRatTSL	-1.7321E+04	-7.39	3.7227E+02	4.1646E+02	5.2142E+02
ResiRatTSL^2	3.8729E+01	7.30			
ResiRatTSL^3	-2.8553E-02	-7.19			

Lower and upper truncation bounds = 5.5206E+04 6.6791E+04

Node 27: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	6.0864E+04	126.43			
WaxRatBSL	-4.0478E+01	-5.46	6.6952E+00	4.0826E+02	1.0975E+03
WaxRatBSL^2	1.2619E-01	6.43			
WaxRatBSL^3	-8.1984E-05	-6.68			

Lower and upper truncation bounds = 5.5841E+04 6.4745E+04

Node 7: Intermediate node

A case goes into Node 14 if MBBSLWFCV <= 4.8500000E+01

MBBSLWFCV mean = 50.1875000000000

Node 14: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	5.7945E+04	43.43			
MBWoFCV	7.9232E+03	4.07	0.0000E+00	2.8375E+01	3.6000E+01
MBWoFCV^2	-5.4702E+02	-4.41			
MBWoFCV^3	9.2189E+00	4.71			

Lower and upper truncation bounds = 5.1375E+04 6.6557E+04

Node 15: Terminal node

Coefficients of least squares regression function:

Regressor	Coefficient	t-stat	Min	Mean	Max
Constant	-3.7609E+07	-6.56			
DryWeBin2	4.4140E+06	6.53	2.4000E+01	2.5375E+01	2.7000E+01
DryWeBin2^2	-1.7224E+05	-6.49			
DryWeBin2^3	2.2381E+03	6.45			

Lower and upper truncation bounds = 5.3685E+04 6.0493E+04

APPENDIX C
GUIDE Decision Tree Results for All Models Investigated

Table 1c. GUIDE decision tree results with no pruning for the IB of 0.500" MDF.

RT algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
No Pruning:								
Piecewise simple linear	100	20	13	7	9.9%	6.84	26.38	0.78
Piecewise simple linear	175	35	19	10	10.5%	6.91	26.58	0.78
Best third-order model	100	20	25	13	9.9%	11.87	12.54	0.96
Best third-order model	175	35	43	22	10.5%	2.92	23.74	0.96
Mixed stepwise	100	20	13	7	9.9%	13.81	16.85	0.78
Mixed stepwise	175	35	39	20	10.5%	12.82	13.63	0.26
Stepwise -all possible subsets	100	20	25	13	9.9%	8.10	31.93	0.69
Stepwise -all possible subsets	175	35	45	23	10.5%	9.12	16.34	0.62
Second-order model (shorter record length from Chapter IV)*	60	13	15	8	8.5%	2.16	15.74	0.96

*Models highlighted in blue are discussed in chapter (note quantile regression models are not listed).

Table 2c. GUIDE decision tree results with pruning by v-fold cross-validation for the IB of 0.500" MDF.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
Pruning by v-fold cross-validation								
Piecewise simple linear	100	20	3	2	9.9%	11.19	24.86	0.40
Piecewise simple linear	175	35	3	2	10.5%	11.51	12.40	0.41
Best third-order model	100	20	1	1	9.9%	11.57	14.47	0.33
Best third-order model	175	35	1	1	10.5%	12.12	12.63	0.33
Mixed stepwise	100	20	7	4	9.9%	11.69	23.10	0.64
Mixed stepwise	175	35	1	1	10.5%	7.89	16.26	0.72
Stepwise -all possible subsets	100	20	9	5	9.9%	11.69	23.10	0.61
Stepwise -all possible subsets	175	35	5	3	10.5%	10.94	19.12	0.46

Second-order model (shorter record length from Chapter IV)	60	13	3	2	8.5%	8.01	13.08	0.50

Table 3c. GUIDE decision tree results with no pruning for the IB of 0.625" MDF.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
No Pruning:								
Piecewise simple linear	100	20	15	8	9.7%	4.88	15.21	0.87
Piecewise simple linear	200	40	27	14	9.5%	5.64	20.94	0.82
Piecewise simple linear	300	60	43	22	9.9%	5.53	17.95	0.83
Piecewise simple linear	400	80	151	76	10.0%	3.00	26.60	0.95
Best third-order model	100	20	25	13	9.7%	1.76	19.89	0.98
Best third-order model	200	40	47	24	9.5%	3.90	24.28	0.94
Best third-order model	300	60	77	37	9.9%	2.52	20.32	0.97
Best third-order model	400	80	77	39	10.0%	5.06	20.80	0.87
Mixed stepwise	100	20	19	10	9.7%	13.00	10.94	0.07
Mixed stepwise	200	40	35	18	9.5%	12.57	13.93	0.10
Mixed stepwise	300	60	51	26	9.9%	12.17	15.49	0.20
Mixed stepwise	400	80	57	29	10.0%	12.49	14.59	0.19
Stepwise -all possible subsets	100	20	27	14	9.7%	9.22	11.20	0.48
Stepwise -all possible subsets	200	40	51	26	9.5%	10.92	18.72	0.33
Stepwise -all possible subsets	300	60	75	37	9.9%	10.25	19.88	0.43
Stepwise -all possible subsets	400	80	73	37	10.0%	7.81	21.96	0.68

Second-order model (shorter record length from Chapter IV)	62	13	17	9	11.1%	2.36	18.31	0.98

Table 4c. GUIDE results with pruning by v-fold cross-validation for the IB of 0.625" MDF.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
Pruning by v-fold cross-validation								
Piecewise simple linear	100	20	5	3	9.7%	8.26	13.44	0.62
Piecewise simple linear	200	40	3	2	9.5%	11.16	11.44	0.29
Piecewise simple linear	300	60	3	2	9.9%	11.50	13.59	0.28
Piecewise simple linear	400	80	9	5	10.0%	11.33	18.08	0.34
Best third-order model	100	20	1	1	9.7%	11.47	22.84	0.28
Best third-order model	200	40	3	2	9.5%	11.06	17.15	0.32
Best third-order model	300	60	3	2	9.9%	13.11	13.53	0.27
Best third-order model	400	80	7	4	10.0%	11.35	15.68	0.33
Mixed stepwise	100	20	1	1	9.7%	9.42	14.84	0.51
Mixed stepwise	200	40	1	1	9.5%	8.33	17.10	0.61
Mixed stepwise	300	60	3	2	9.9%	7.82	18.51	0.67
Mixed stepwise	400	80	3	2	10.0%	7.88	23.66	0.61
Stepwise - all possible subsets	100	20	19	10	9.7%	9.71	11.16	0.67
Stepwise - all possible subsets	200	40	2	1	9.5%	12.10	13.13	0.18
Stepwise - all possible subsets	300	60	2	1	9.9%	12.31	13.42	0.18
Stepwise - all possible subsets	400	80	7	4	10.0%	11.52	20.39	0.36

Second-order model (shorter record length from Chapter IV)	62	13	3	2	11.1%	9.55	11.29	0.60

Table 5c. GUIDE decision tree results with no pruning for the IB of 0.750"MDF.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
No Pruning:								
Piecewise simple linear	100	20	35	18	9.7%	3.25	23.35	0.94
Piecewise simple linear	200	40	75	38	10.6%	2.97	20.03	0.96
Best third-order model	100	20	19	10	9.7%	3.34	23.98	0.79
Best third-order model	200	40	23	12	10.6%	3.50	24.33	0.94
Mixed stepwise	100	20	13	7	9.7%	13.55	14.76	0.16
Mixed stepwise	200	40	31	16	10.6%	14.38	18.57	0.97
Stepwise - all possible subsets	100	20	27	14	9.7%	12.29	22.44	0.17
Stepwise - all possible subsets	200	40	47	24	10.6%	11.52	13.16	0.46

Second-order model (shorter record length from Chapter IV)	70	14	23	12	12.1%	2.01	30.10	0.98

Table 6c. GUIDE decision tree results with pruning by v-fold cross-validation for the IB of 0.750" MDF.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
Pruning by v-fold cross-validation								
Piecewise simple linear	100	20	1	1	9.7%	12.85	14.43	0.10
Piecewise simple linear	200	40	1	1	10.6%	14.15	15.75	0.07
Best third-order model	100	20	1	1	9.7%	12.36	11.58	0.15
Best third-order model	200	40	1	1	10.6%	14.00	17.88	0.10
Mixed stepwise	100	20	5	3	9.7%	12.86	13.69	0.10
Mixed stepwise	200	40	7	4	10.6%	13.44	18.80	0.17
Stepwise - all possible subsets	100	20	1	1	9.7%	12.54	16.52	0.15
Stepwise - all possible subsets	200	40	1	1	10.6%	14.14	14.72	0.13

Second-order model (shorter record length from Chapter IV)	70	14	1	1	12.1%	14.69	15.87	0.20

Table 7c. GUIDE decision tree results with no pruning for the IB of OSB.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
No Pruning:								
Piecewise simple linear	100	20	33	17	20.2%	2.18	15.01	0.95
Piecewise simple linear	200	40	69	35	19.9%	2.08	13.04	0.95
Piecewise simple linear	300	60	100	51	19.6%	2.07	14.61	0.95
Best third-order model	100	20	23	12	20.2%	4.27	7.14	0.96
Best third-order model	200	40	49	25	19.9%	4.27	10.30	0.94
Best third-order model	300	60	71	36	19.6%	4.67	15.05	0.95
Mixed stepwise	100	20	11	6	20.2%	7.70	9.28	0.36
Mixed stepwise	200	40	21	11	19.9%	7.65	8.67	0.28
Mixed stepwise	300	60	41	21	19.6%	7.79	11.30	0.25
Stepwise - all possible subsets	100	20	23	12	20.2%	5.79	9.24	0.64
Stepwise - all possible subsets	200	40	33	17	19.9%	7.03	7.44	0.73
Stepwise - all possible subsets	300	60	39	20	19.6%	5.07	13.35	0.69

Second-order model (shorter record length from Chapter IV)	59	12	15	8	19.6%	2.30	11.01	0.94

Table 8c. GUIDE decision tree results with pruning by v-fold cross-validation for the IB of OSB.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
Pruning by v-fold cross-validation								
Piecewise simple linear	100	20	2	1	20.2%	8.48	8.47	0.22
Piecewise simple linear	200	40	3	2	19.9%	7.53	8.41	0.31
Piecewise simple linear	300	60	2	1	19.6%	8.28	9.15	0.13
Best third-order model	100	20	2	1	20.2%	8.14	7.41	0.28
Best third-order model	200	40	2	1	19.9%	7.86	7.37	0.25
Best third-order Model	300	60	2	1	19.6%	8.43	9.46	0.16
Mixed stepwise	100	20	3	2	20.2%	5.42	8.35	0.68
Mixed stepwise	200	40	2	1	19.9%	7.03	7.44	0.40
Mixed stepwise	300	60	3	2	19.6%	7.28	9.36	0.35
Stepwise - all possible subsets	100	20	3	2	20.2%	6.68	9.01	0.52
Stepwise - all possible subsets	200	40	2	1	19.9%	7.86	7.37	0.25
Stepwise - all possible subsets	300	60	2	1	19.6%	8.15	9.80	0.18

Second-order model (shorter record length from Chapter IV)	59	12	1	1	19.6%	7.88	9.17	0.26

Table 9c. GUIDE decision tree results with no pruning for the Parallel EI of OSB.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
No Pruning:								
Piecewise simple linear	100	20	37	19	20.2%	3040	5583	0.96
Piecewise simple linear	200	40	65	33	19.9%	1476	5852	0.84
Piecewise simple linear	300	60	97	49	19.6%	1360	6679	0.91
Best third-order model	100	20	21	11	20.2%	920	4361	0.94
Best third-order model	200	40	45	23	19.9%	946	5337	0.94
Best third-order model	300	60	71	36	19.6%	1103	6128	0.94
Mixed stepwise	100	20	9	5	20.2%	3565	3888	0.05
Mixed stepwise	200	40	25	13	19.9%	3566	4207	0.08
Mixed stepwise	300	60	41	21	19.6%	4314	4097	0.14
Stepwise - all possible subsets	100	20	25	13	20.2%	2059	4670	0.68
Stepwise - all possible subsets	200	40	53	27	19.9%	2798	4853	0.43
Stepwise - all possible subsets	300	60	75	38	19.6%	4376	8066	0.11

Second-order model (shorter record length from Chapter IV)	58	12	15	8	8.5%	692	2479	0.96

Table 10c. GUIDE decision tree results with pruning by v-fold cross-validation for the Parallel EI of OSB.

RT Algorithm	n		Total Nodes	Terminal nodes	CV	RMSE	RMSEP	R^2 Tree Model
	Training	Validation						
Pruning by v-fold cross-validation								
Piecewise simple linear	100	20	2	1	6.24%	3528	3994	0.07
Piecewise simple linear	200	40	2	1	8.25%	3615	4185	0.05
Piecewise simple linear	300	60	2	1	7.74%	4414	3778	0.09
Best third-order model	100	20	2	1	6.24%	3356	3825	0.16
Best third-order model	200	40	2	1	8.25%	3615	4185	0.07
Best third-order model	300	60	2	1	7.74%	4382	3837	0.11
Mixed stepwise	100	20	5	3	6.24%	3040	5583	0.31
Mixed stepwise	200	40	21	11	8.25%	3560	4165	0.07
Mixed stepwise	300	60	29	15	7.74%	4388	4001	0.11
Stepwise - all possible subsets	100	20	2	1	6.24%	3475	4150	0.10
Stepwise - all possible subsets	200	40	2	1	8.25%	3547	4231	0.09
Stepwise - all possible subsets	300	60	5	3	7.74%	2990	5941	0.29

Second-order model (shorter record length from Chapter IV)	58	12	1	1	8.5%	3255	5988	0.10

APPENDIX D
Variable Description

Medium Density Fiberboard

	Industrial Tag Name	GUIDE name
1	Validate	Validate
2	idnum 625	idnum
3	shiftcode	shiftcode
4	tech	tech
5	avgib	avgib
6	aChipExitAugerSpeed_SI	aChipAugSp
7	aChipsRawWeight_WT	aChipWght
8	aChipStorageSiloLevel_LI	aChipSloLv
9	aCoreMeteringBinLevel_LI	aCoreBinLv
10	aFaceMeteringBinLevel_LI	aFaceBinLv
11	aRMSChipsPercent	aChipPrct
12	aShavingsExitAugerSpeed_SI	aShvAugSpd
13	aShavingsRawWeight_WI	aShvRawWgt
14	bDry_Fibre_Rate	bDryFbrRt
15	bFaceBlowValvePostion_ZI	bFaceBlwPs
16	bFaceFiberMoisture_Act	bFacFibrMs
17	bFaceGrindingPressure_PV	bFacGrdPrs
18	bFaceHotGasTemp_TI	bFaceTemp
19	bFaceMeteringBin_Spd	bFaceBnSpd
20	bFacePlatePosition_ZI	bFacePltPs
21	bFaceResinFlow_PV	bFaceResn
22	bFaceResinSolids	bFaceResnS
23	bFaceResinToWood_Act	bFaceResnW
24	bFaceResinToWood_SP	bFacResnWS
25	bFaceScavToWood_Act	bFacScvW
26	bFaceScavToWood_SP	bFacScvWS
27	bFaceWaterFlow_PV	bFacH20Flw
28	bFaceWaxDryFiberWeight_WI	bFacWxWgt
29	bFaceWaxFlow_PV	bFacWxFlw
30	bMainMotorPower_JI	bMotorPwr
31	bSwingDigesterPressure_PV	bSwgDigPrs
32	cCDry_Fibre_Rate	cDryFbrRt
33	cCI0023PT	cCI0023PT
34	cCO0046PCV	cCO0046
35	cPressStartPositionControl_SP	cPrsPstnCt
36	cResinDryFiberWeight_WI	cRsnFbrWgt
37	cSwing_Seperator_Outlet_PT_two	cSwgOutlet
38	cSwingBlowValvePosition_ZI	cSwBlwVlv
39	cSwingDigesterLevel_PV	cSwDgstrL

40	cSwingDigesterPressure_PV	cSwDgsPres
41	cSwingFiberMoisture_Act	cSwFbrMst
42	cSwingFiberWetWeight_WI	cSwFibWgt
43	cSwingGrindingPressure_PV	cSwgGrndP
44	cSwingHotGasTemp_TI	cSwgTemp
45	cSwingPlatePosition_ZI	cSwPltPI
46	cSwingResinFlow_PV	cSwgRsnF
47	cSwingResinToWood_Act	cSwgRsn2Wd
48	cSwingScavFlow_PV	cSwgScvF
49	cSwingScavToWood_Act	cSwgScv2W
50	cSwingWaterFlow_PV	cSwgH20F
51	cSwingWaterToWood_Act	cSwgH20W
52	cSwingWaxDryFiberWeight_WI	cSwgWxFbrW
53	cSwingWaxFlow_PV	cSwgWxF
54	cSwingWaxPercentSolids	cSwgWxSd
55	cSwingWaxToWood_Act	cSwgWx2W
56	cSwinResinToWood_SP	cSwgRsn2W
57	dCoreDigesterPressure_PV	dCoreDgstP
58	dCoreFiberWetWeight_WI	dCoreFbrW
59	dCoreGrindingPressure_PV	dCoreGrndP
60	dCoreGrindingSteamFlow_FI	dCoreGrndS
61	dCoreMeteringBin_Spd	dCoreMtrBS
62	dCoreMeteringScrew_Dmd	dCoreMtrSr
63	dCoreOutletTemp_PV	dCoreTemp
64	dCorePlatePosition_ZI	dCorePltP
65	dCoreResin_SP	dCoreRsnS
66	dCoreResinDryFiberWeight_WI	dCoreRsnW
67	dCoreResinFlow_PV	dCoreRsnF
68	dCoreResinToWood_Act	dCoreRsn2W
69	dCoreScavWaadRate_Act	dCoreScvWR
70	dCoreTotalWeight_Tons	dCorelWgh
71	dCoreWater_PV	dCoreH20
72	dCoreWaterToWood	dCoreWtr2W
73	dCoreWax_PV	dCoreWx
74	dCoreWaxToWood_Act	dCoreWx2W
75	dDry_Fibre_Rate	dDryFbrR
76	eAmmoniaFlow	eAmmonaF
77	eBoilerInjectionWaterFlow_FI	eBoilrH20F
78	eBoilerSteamFlow_FT	eBoilrStmF
79	eBoilerSteamPressure_PV	eBoilrStmP
80	eDesp_Field1_Kilovolts	eDesp1KV
81	eDesp_Field1_Milliamps	eDespMamp
82	eDesp_Field2_Kilovolts	eDesp2KV
83	eDesp_Field3_Kilovolts	eDesp3KV
84	eDryerMassFlow_FI	eDryrFlw

85	eDryerSupplyPressure_PI	eDryrPrs
86	eE_Vapour_Temp	eVaprTemp
87	eEI0311AI	eEI0311AI
88	eEI0338FT	eEI0338FT
89	eEI1071FT	eEI1071FT
90	eM2236_Spd	eM2236Spd
91	eM2237_Spd	eM2237Spd
92	eM2241_Spd	eM2241Spd
93	eM2243_Spd	eM2243Spd
94	fCoreBinBottomSpeed_SI	fCoreBtmSpd
95	fCoreFiberMatMoisture_MI	fCoreMatMst
96	fCoreHumidifierTemp_PV	fCoreHTmpP
97	fCoreHumidifierTemp_TI	fCoreHTmpT
98	fCoreHumidity_PV	fCoreHumid
99	fFaceBinBottomSpeed_SI	fFaceBtmSd
100	fFaceFiberMatMoisture_MI	fFaceMstM
101	fFaceHumidifierDraft_PV	fFaceHDP
102	fFaceHumidifierTemp_PV	fFaceHTP
103	fFormerThayerWeight_PV	fFormerWgt
104	fHumidifierSupplyPressue_PV	fHumidPrs
105	fMatCoreToFaceRatio_SP	fCore2Face
106	fShaveOffElevatorLevelone_LI	fShaveOff1
107	fShaveOffElevatorLeveltwo_LI	fShaveOff2
108	fShaveOffMatThicknesstwo_LI	fShavOffT2
109	fShaveOffMatThicknessthree_LI	fShavOffT3
110	fShaveOffMatThicknessfour_LI	fShavOffT4
111	gFormingWireSpeed_SI	gFrmSpd
112	gOOPAverageBoardWeight_PV	gAvgBrdWgt
113	gPrecompBottomBeltSpeed_SI	gPreBBSpd
114	hPressCloseoneTime_PV	hPrClsTime
115	hPressClosetwoTime_PV	hPrCls2Tim
116	hPressClosethreeTime_PV	hPrCls3Tim
117	hPressClosingTime_PV	hPrClsTimP
118	hPressFinalPositionHoldTime_SP	hPrFPHTime
119	hPressFinalPositionTrip_SP	hPrFPTrip
120	hPressOverallTime_PV	hPrAllTime
121	hPressOverallTime_SP	hPrAlTimeS
122	hPressPostionTime_PV	hPrPTime
123	hPressPPHoldTime_PV	hPrPPHTime
124	hPressPPHoldTime_SP	hPrPPHTimS
125	hPressPPMoveTime_SP	hPrPPMTimS
126	hPressPrePostitionX_SP	hPrPPXSP
127	hPressSteamPressure_PI	hPrStmP
128	hPressTemp_PV	hPrTempP
129	hPressTemp_SP	hPrTempS

130	hPressTotalCycleTime_PV	hPrTCyclP
131	gMeasurexTarget	gMsrexTgt
132	dCoreBlowLinePressure_PI	dCoreBLPrs
133	bFaceFibreBinLevel	bFaceBnLvl
134	dCoreFibreBinLevel	dCoreBnLvl
135	bFaceChipChuteLevel_PV	bFaceChpLv
136	bFaceDigesterLevel_PV	bFaceDigLv
137	bFaceDigesterPressure_PV	bFaceDigPr
138	bFaceEastFanCurrent_PV	bFaceFanCr
139	dCoreResinPressure_PI	dCoreResnP
140	bFaceGrindingSteamFlow_FI	bFaceGrdSF
141	bFaceOutletTemp_PV	bFaceTempP
142	bFacePlugFeederScrewSpeed	bFacePFSS
143	bFaceRefinerFeederScrewSpeed	bFaceRFSS
144	bFaceResinDryFiberWeight_WI	bFaceResnW
145	bFaceScavFlow_PV	bFaceScvF
146	bFaceWaterToWood_Act	bFaceH202W
147	bFaceWaxToWood_Act	bFaceWx2W
148	bFaceWestFanCurrent_PV	bFaceWFCur
149	cResinWaterTankTemp_PV	cResnH20Tp
150	cSwingChipChuteLevel_PV	cSwngChpL
151	cSwingGrindingSteamFlow_FI	cSwngGSF
152	cSwingMainMotorPower_JI	cSwngMPwr
153	cSwingMeteringBin_Spd	cSwMtrBSpd
154	cSwingOutletTemp_PV	cSwOTemp
155	cSwingPlugFeederScrewSpeed	cSwPgFSSpd
156	cSwingRefinerFeederScrewSpeed	cSwRFSSpd
157	cSwingRefinerTotalSteamFlow	cSwRTSFlw
158	dCoreBlowValvePosition_ZI	dCoreBVP
159	dCoreDustRatio_SP	dCoreDRSP
160	dCoreEastFanCurrent_PV	dCoreEFCur
161	dCoreFiberMoisture_Act	dCoreMoist
162	dCorePlugFeederScrewSpeed	dCoreFSSpd
163	dCoreRefinerFeederScrewSpeed	dCoreRSSpd
164	dCoreRefinerTotalSteamFlow	dCoreRTSFw
165	dCoreScavFlow_PV	dCoreScvFP
166	dCoreWaterToWood_Act	dCoreH202W
167	dMainMotorPower_JI	dMnMtrPwr
168	eDesp_Field2_Milliamps	eDFld2Mamp
169	eDesp_Field3_Milliamps	eDFld3Mamp
170	eDryerSupplyPressure_PV	eDryrSPrs
171	eDryESPOutletTemp_TI	eDryESPTmp
172	eE_Nox_Lbs_Hr	eENoxLbsHr
173	ePressSteamFlow_FI	ePrStmFlow
174	fFaceHumidifierTemp_TI	fFaceHTemp

175	fFaceHumidity_PV	fFaceHumid
176	fOutsideHeadTemperature	fOutHdTmp
177	fRelativeHumidity	fHumidity
178	fShaveOffElevatorLevelthree_LI	fShaveOff3
179	fShaveOffElevatorLevelfour_LI	fShaveOff4
180	fShaveOffMatThicknessone_LI	fShavOffT1
181	gBulkDensity	gBlkDensty
182	gFiberBulkDensity_PV	gFBlkDenst
183	gMeasurexActual	gMxActual
184	gMeasurexAverage	gMxAverage
185	hPositionTime	hPosTime
186	hPressFullOpenTime_PV	hPrOpnTime
187	hPressPrePositionTime_PV	hPPstnTime
188	hPressStartPositionControl_SP	hPrSCtrlSP

Oriented Strand Board

	Industrial Tag Name	GUIDE name
1	TestNumber	Number
2	Weight	Weight
3	Thickness	Thick
4	Density	Density
5	ParallelEI	ParlEI
6	PerpendicularEI	PerpEI
7	ParallelMM	ParlMM
8	PerpendicularMM	PerpMM
9	WetMM	WetMM
10	IB	IB
11	ThicknessSwell	Thicksweel
12	WaterAbsorption	WtAbsorpt
13	ParallelLE	ParlLE
14	PerpendicularLE	PerpLE
15	BunkerSpeed_BCL	BnkSpdBCL
16	BunkerSpeed_BSL	BnkSpdBSL
17	BunkerSpeed_TCL	BnkSpdTCL
18	BunkerSpeed_TSL	BnkSpdTSL
19	DryBinLiveBottom_BSL	DyBiLBoBSL
20	DryBinLiveBottom_CL	DyBiLBoCL
21	DryBinLiveBottom_TSL	DyBiLBoTSL
22	Dryer1_Inlet	Dry1In
23	Dryer1_Outlet	Dry1Out
24	Dryer2_Inlet	Dry2In
25	Dryer2_Outlet	Dry2Out
26	Dryer3_Inlet	Dry3In
27	Dryer3_Outlet	Dry3Out
28	Dryer4_Inlet	Dry4In
29	Dryer4_Outlet	Dry4Out
30	Dryer5_Inlet	Dry5In
31	Dryer5_Outlet	Dry5Out
32	WetBin_5_Total_Runtime_24_Hour	WeBi5Tot24
33	DryerWetBin1	DryWeBin1
34	DryerWetBin2	DryWeBin2
35	DryerWetBin3	DryWeBin3
36	DryerWetBin4	DryWeBin4
37	DryerWetBin5	DryWeBin5
38	Flaker1_StrokeSpeed	Flk1StoSpd
39	Flaker2_StrokeSpeed	Flk2StoSpd
40	Flaker3_StrokeSpeed	Flk3StoSpd
41	FormingLineSpeed	FomLinSpd
42	MainBlenderBSLResinFlowPIDPV	MBBSLRFPV
43	MainBlenderBSLWaxFlowPIDPV	MBBSLWFPV

44	MainBlenderBSLResinFlowPIDCV	MBBSLRFCV
45	MainBlenderBSLResinFlowPIDDeviation	MBBSLRFD
46	MainBlenderBSLResinFlowPIDRatioFB	MBBSLRFB
47	MainBlenderBSLResinFlowPIDRatioSP	NBBSLRSP
48	MainBlenderBSLResinFlowPIDSP	MBBSLSP
49	MainBlenderBSLResinFlowTotalCurrent	MBBSLRFTC
50	MainBlenderBSLResinFlowTotalPrev	MBBSLRFTP
51	MainBlenderBSLWoodFlowPIDCV	MBBSLWFCV
52	MainBlenderBSLWoodFlowPIDManCV	MBBSLMCV
53	MainBlenderBSLWoodFlowPIDPV	MBBSLPV
54	MainBlenderBSLWoodFlowPIDSPFB	MBBSLSPFB
55	MainBlenderBSLWoodTotalCurrent	MBBSLWTC
56	MainBlenderBSLWoodTotalPrev	MBBSLWTP
57	MainBlenderCoreResinFlowPIDCV	MBCRFCV
58	MainBlenderCoreResinFlowPIDDeviation	MBCRFD
59	MainBlenderCoreResinFlowPIDPV	MBCRFPV
60	MainBlenderCoreResinFlowPIDRatioFB	MBCRFRFB
61	MainBlenderCoreResinFlowPIDRatioSP	MBCRFRSP
62	MainBlenderCoreResinFlowPIDSP	MBCRFSP
63	MainBlenderCoreResinFlowTotalCurrent	MBCRFTC
64	MainBlenderCoreResinFlowTotalPrev	MBCRFTP
65	MainBlenderCoreWoodFlowPIDCV	MBCWoFCV
66	MainBlenderCoreWoodFlowPIDManCV	MBCWoFMCV
67	MainBlenderCoreWoodFlowPIDManCVFB	MBCWoFMVFB
68	MainBlenderCoreWoodFlowPIDPV	MBCWoFPV
69	MainBlenderCoreWoodFlowPIDSP	MBCWoFSP
70	MainBlenderCoreWoodFlowPIDSPFB	MBCWoFSPFB
71	MainBlenderCoreWoodTotalCurrent	MBCWoTC
72	MainBlenderCoreWoodTotalPrev	MBCWoTP
73	MainBlenderTSLResinFlowPIDCV	MBTRFCV
74	MainBlenderTSLResinFlowPIDDeviation	MBTRFDe
75	MainBlenderTSLResinFlowPIDPV	MBTRFPV
76	MainBlenderTSLResinFlowPIDRatioFB	MBTRFRFB
77	MainBlenderTSLResinFlowPIDRatioSP	MBTRFRSP
78	MainBlenderTSLResinFlowPIDSP	MBTRFSP
79	MainBlenderTSLResinFlowTotalCurrent	MBTRFTC
80	MainBlenderTSLResinFlowTotalPrev	MBTRFTP
81	MainBlenderTSLWoodFlowPIDCV	MBWoFCV
82	MainBlenderTSLWoodFlowPIDManCV	MBWoFMCV
83	MainBlenderTSLWoodFlowPIDManCVFB	MBWoFMCVFB
84	MainBlenderTSLWoodFlowPIDPV	MBWoFPV
85	MainBlenderTSLWoodFlowPIDSPFB	MBWoFSPFB
86	MainBlenderTSLWoodTotalCurrent	MBWoTC
87	MainBlenderTSLWoodTotalPrev	MBWoTP
88	MainBSLMoistureLevel	MBSLMoiLev

89	MainDryBinBSLLevel	MDBBSLLev
90	MainDryBinCoreLevel	MDBCLev
91	MainDryBinTSLLevel	MDBTSLLev
92	MainDryer1OutletTemp	MD1OutTem
93	MainDryer2OutletTemp	MD2OutTem
94	MainDryer3OutletTemp	MD3OutTem
95	MainDryer4OutletTemp	MD4OutTem
96	MainDryer5OutletTemp	MD5OutTem
97	MainFlaker_1_HMI_Watchdog	MF1HMIWatd
98	MainFlaker_1_Watchdog	MF1Watdog
99	MainFlaker_2_HMI_Watchdog	MF2HMIWatd
100	MainFlaker_3_HMI_Watchdog	MF3HMIWatd
101	MainFlaker_Support_HMI_Watchdog	MFSHMIWat
102	MainFlaker1FlakeTime	MF1Flakt
103	MainFlaker1PassCounter	MF1PasCont
104	MainFlaker2FlakeTime	MF2Flakt
105	MainFlaker2PassCounter	MF2PasCont
106	MainFlaker3FlakeTime	MF3Flakt
107	MainFlaker3PassCounter	MF3PasCont
108	MainFLSpeedActualFpm	MFLSpdAFpm
109	MainFLSpeedActualPercent	MFLSpdAPer
110	MainFLSpeedSPAcceptedFpm	MFLSpdAcpF
111	MainFLSpeedSPAcceptedPercent	MFLSpdAcpP
112	MainFLSpeedSPFpm	MFLSpdFpm
113	MainFormingLineCoreWeight	MFLCorWegt
114	MainFormingLineTearWeight	MFLTerWegt
115	MainFormingLineTotalWeight	MFLTotWegt
116	MainHotOil2OilTemp	MHOil2OilT
117	MainHotOilTotalFlow	MHOilTotFl
118	MainPressTemperature	MPTemp
119	MainSpreaderBCLLevel	MSpBCLLev
120	MainSpreaderBCLOFSpeedActual	MSpBCLOFSA
121	MainSpreaderBSLLevel	MSBSLLev
122	MainSpreaderBSLOFSpeedActual	MSBSLOFSpA
123	MainSpreaderCLWeightActual	MSCLWegAct
124	MainSpreaderSLWeightActual	MSSLWegAct
125	MainSpreaderTCLLevel	MSTCLLev
126	MainSpreaderTCLOFSpeedActual	MSTCLOFSpA
127	MainSpreaderTotalWeightActual	MSToWegAct
128	MainSpreaderTSLLevel	MSTSLLev
129	MainSpreaderTSLOFSpeedActual	MSTSLOFSpA
130	MainTCLMoistureLevel	MTCLMoiLev
131	MainTSLMoistureLevel	MTSLMoiLev
132	Moisture_BSL	MosturBSL
133	Moisture_CL	MosturCL

134	Moisture_TSL	MosturTSL
135	PressClosingTime	PrsClosTim
136	PressDecompression_3	PrsDecomp3
137	Pressi_AVG_POS	PrsAVGPOS
138	Pressi_HK700-Close3Rate	PrHK700C3R
139	Pressi_HK700-Decompress3Rate	PHK700D3Ti
140	Pressi_HK700-Decompress3Time	PHK700D3Tr
141	Pressi_HK700-FinalPositionTrip	PHK700FPTr
142	Pressi_HK700-OverallTime	PHK700OvTi
143	Pressi_HK700-PPMoveTime	PHK700PPMT
144	Pressi_HK700-PrePosn1	PHK700PP1
145	Pressi_HK700-PrePosnX	PHK700PPX
146	Pressi_HK701-Decompress1Trip	PHK71DP1Tr
147	Pressi_HK701-Decompress2Trip	PHK71DP2Tr
148	Pressi_HK701-Decompress3Rate	PHK71DP3Ti
149	Pressi_HK701-Decompress3Time	PHK71DP3Tr
150	Pressi_HK701-FinalPositionTrip	PHK71FPTri
151	Pressi_HK701-OverallTime	PHK701OvTi
152	Pressi_HK701-PrePosn1	PHK701PPo1
153	Pressi_HK701-PrePosnX	PHK701PPX
154	Pressi_KI700A-Close1Time	PKI700AC1T
155	Pressi_KI700C-Close2Time	PKI700CC2T
156	Pressi_KI700I-ClosingTime	PKI700ICTi
157	Pressi_KI700J-PositionTime	PKI700JPTi
158	Pressi_KI700M-OverallTime	PKI700MOTi
159	Pressi_KI700P-Decompress3Time	PKI700PD3T
160	Pressi_KI700Q-OpeningTime	PKI700QOTi
161	Pressi_KI700R-TotalCycleTime	PKI700RTCT
162	Pressi_KI700S-FullOpenTime	PKI700SFOY
163	Pressi_KI700T-ButtonTime	PKI700TBTi
164	Pressi_LI770	PLI770
165	Pressi_LI780	PLI780
166	Pressi_LI795	PLI795
167	Pressi_MI731	PMI731
168	Pressi_MI732	PMI732
169	Pressi_MI733	PMI733
170	Pressi_MI734	PMI734
171	Pressi_MI735	PMI735
172	Pressi_MI736	PMI736
173	Pressi_MI737	PMI737
174	Pressi_MI740	PMI740
175	Pressi_MI741	PMI741
176	Pressi_MI742	PMI742
177	Pressi_MI743	PMI743
178	Pressi_MI744	PMI744

179	Pressi_MI745	PMI745
180	Pressi_MI746	PMI746
181	Pressi_MI747	PMI747
182	Pressi_MI748	PMI748
183	Pressi_MI770	PMI770
184	Pressi_MI780	PMI780
185	Pressi_MI795	PMI795
186	Pressi_PC741A	PPC741A
187	Pressi_PC741B	PPC741B
188	Pressi_PC741C	PPC741C
189	Pressi_PI700A	PPI700A
190	Pressi_PI700B	PPI700B
191	Pressi_PI740	PPI740
192	Pressi_PI770	PPI770
193	Pressi_PIC700-PV	PPIC700PV
194	Pressi_QI700	PQI700
195	Pressi_TEMPOA_POS	PTEMPAPOS
196	Pressi_TEMPOB_POS	PTEMPBPOS
197	Pressi_TI700	PTI700
198	Pressi_TI770	PTI770
199	Pressi_TI793	PTI793
200	Pressi_TI796A	PTI796A
201	Pressi_TI796B	PTI796B
202	Pressi_TI796C	PTI796C
203	Pressi_TI796D	PTI796D
204	Pressi_UY741B-Output	PUY741BOut
205	Pressi_YI741-ASectSP	PYI741ASSP
206	Pressi_YZ741C	PYZ741C
207	Pressi_YZ743C	PYZ743C
208	Pressi_YZ744A	PYZ744A
209	Pressi_YZ745B	PYZ745B
210	Pressi_YZ747B	PYZ747B
211	Pressi_YZ747C	PYZ747C
212	Pressi_ZI701	PZI701
213	Pressi_ZI701-Step1	PZI701St1
214	Pressi_ZI701-Step12	PZI701St12
215	Pressi_ZI701-Step13	PZI701St13
216	Pressi_ZI701-Step15	PZI701St15
217	Pressi_ZI701-Step17	PZI701St17
218	Pressi_ZI701-Step2	PZI701Ste2
219	Pressi_ZI701-Step3	PZI701Ste3
220	Pressi_ZI701-Step4	PZI701Ste4
221	Pressi_ZI701-Step5	PZI701Ste5
222	Pressi_ZI701-Step6	PZI701Ste6
223	Pressi_ZI701-Step7	PZI701Ste7

224	Pressi_ZI701-Step9	PZI701Ste9
225	Pressi_ZI701B	PZI701B
226	Pressi_ZI740	PZI740
227	Pressi_ZIC700-Kp	PZIC700Kp
228	Pressi_ZIC700-PV	PZIC700PV
229	Pressi_ZIC700SP	PZIC700SP
230	Pressi_ZIC700-Vc	PZIC700Vc
231	Pressi_ZIC740SP	PZIC740SP
232	PressMainPressure	PreMaiPres
233	PressTemperature	PreTempt
234	PressTotalCycleTime	PreToCyTi
235	ResinRate_BSL	ResiRatBSL
236	ResinRate_CL	ResiRatCL
237	ResinRate_TSL	ResiRatTSL
238	WaxRate_BSL	WaxRatBSL
239	WaxRate_CL	WaxRatCL
240	WaxRate_TSL	WaxRatTSL
241	WeightScale_BSL	WegScaBSL
242	WeightScale_CORE	WegScaCORE
243	WeightScale_TSL	WegScaTSL
244	MainSpreaderBCLOFDensitySP	MSBCLOFDSP
245	MainSpreaderBSLOFDensitySP	MSBSLOFDSP
246	Dryer1OutletMoisture	Dr1OutMois
247	Dryer2OutletMoisture	Dr2OutMois
248	Dryer3OutletMoisture	Dr3OutMois
249	Dryer4OultetMoisture	Dr4OutMois
250	Dryer5OutletMoisture	Dr5OutMois

Vita

Timothy Mark Young was born in Appleton, Wisconsin on September 6, 1956. Tim graduated from Xavier High School in Appleton in May, 1974. He attended the University of Wisconsin Fox Valley Extension campus for two years before transferring to the University of Wisconsin at Madison. Tim received a B.S. in Forestry in December, 1979 and worked for one year for the U.S. Forest Service in Yaak, Montana in 1980. He returned to the University of Wisconsin and received a M.S. in Forest Economics in May, 1983.

Tim worked as a Research Associate for the University of Tennessee from 1983 to 1993 and completed a M.S. in Statistics from the University of Tennessee in December 1993. He worked for Georgia-Pacific Corporation from 1994 to 1998 in South Carolina as Director of Continuous Improvement and returned to the University of Tennessee in December of 1998. Tim was a Research Assistant Professor in the Forest Products Center from December 1998 to December 2003 and was promoted to Research Associate Professor in January 2004. In 2007, he obtained a Doctor of Philosophy degree from the University of Tennessee in Natural Resources with a concentration in Statistics.