Classification Tree Models of Forest Landowner Attitudes Towards Timber Harvests

"A Case Study of the Cumberland Plateau in Tennessee"

Timothy M. Young, PhD
Research Associate Professor
Forest Product Center

Don Hodges, PhD
Professor
Department of Forestry, Wildlife and Fisheries

Yingjin Wang
Graduate Research Assistant
Department of Statistics

Frank M. Guess, PhD
Professor
Department of Statistics

University of Tennessee
Contact: tmyoung1@utk.edu, 865.946.1119

Forest Products Industry

"Current State in the U.S.A."

- Industry/forests are undergoing unprecedented change
  - increasing wood fiber costs
  - land fragmentation/population growth
  - rising energy prices
  - falling final product prices with constrained credit markets
- Forest products industry contributed $240 billion to the U.S. economy and employed more than one million Americans in 2002 (U.S. Census Bureau 2004)
  - Tennessee industry contributes $21.7 billion to the economy and employed 180,000 Tennesseans in 2000 (Young, Rials and Hodges 2007)
- Identifying landowner attitudes towards timber harvests may be an important consideration for wood procurement

Outline

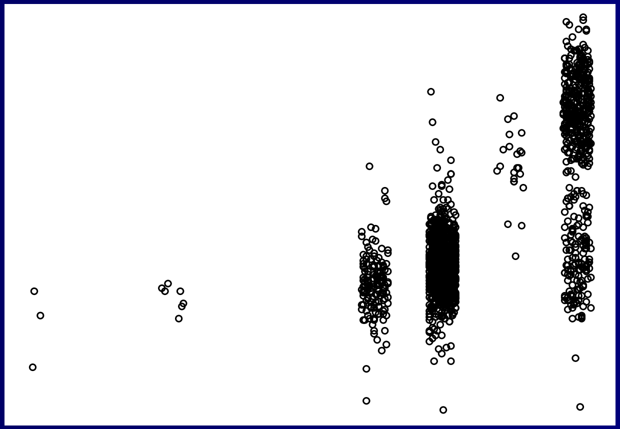
- Forest Landowner Survey
- Decision Trees
  - Regression Trees
  - Classification Trees
- Classification Tree Models of Forest Landowner Attitudes towards Timber Harvests
- Regression Tree Models of Mechanical Properties of Wood Composites

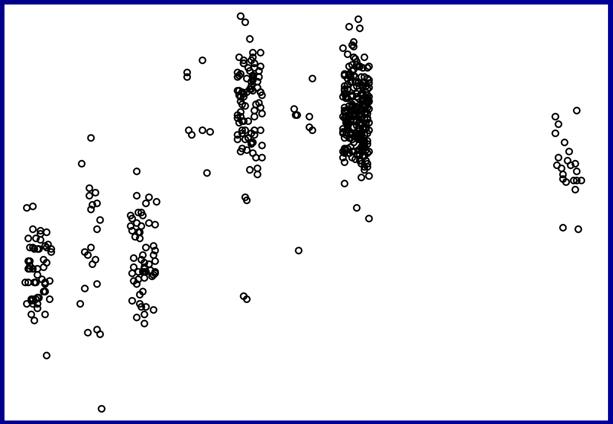
Forest Landowner Survey

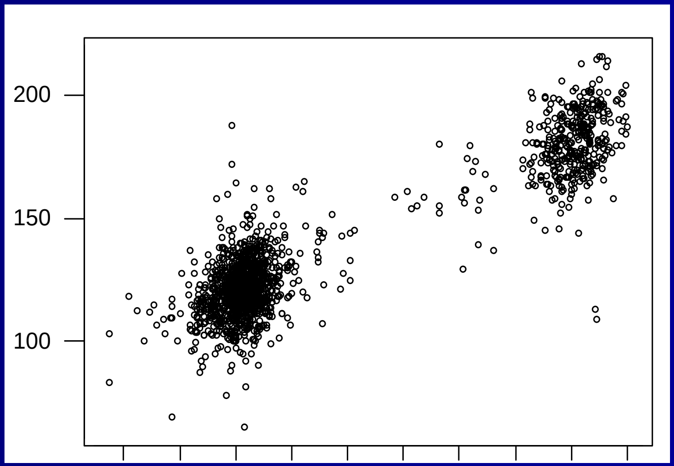
- In 2005, useable surveys from 495 private woodland owners (55% response rate) in the Northern Cumberland Plateau region of Tennessee were obtained.
- The purpose of the survey was to improve the understanding of forest landowner attitudes of commercial timber harvest and identify factors that influence the decisions to harvest.
- Survey Justification:
  - important to understand and advance the knowledge of forest landowner's attitude toward commercial timber harvest and forest land management (see Brubaker, Finley, and McDill, 2006; Wolf & Hufnagl-Eichiner, 2007);
  - may become critical if availability of wood fiber becomes economically scarce

Decision Trees

- Useful when Data are Heterogeneous







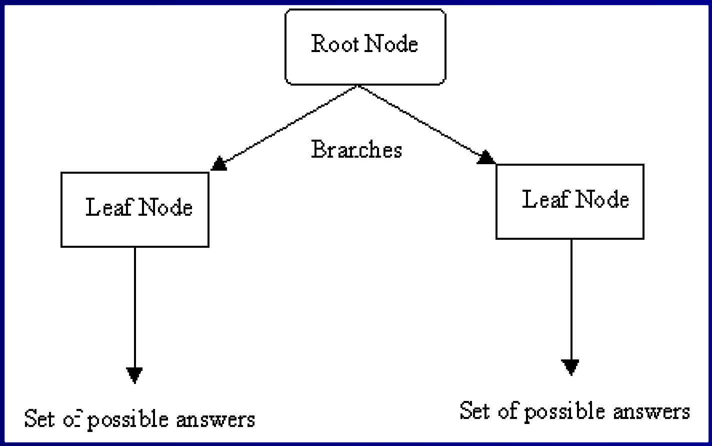
Decision Trees

- A decision tree is a model that is both predictive and descriptive.
- It is called a decision tree because the resulting model is presented in the form of a **tree structure**.
- The visual presentation makes the decision tree model very easy to understand and assimilate.
- As a result, the decision tree has become a very popular inductive or data mining technique.
- Important in the field of "Decision Theory"

Decision Trees

"Structure"

- Each box on the tree represents a node
- The top node is called the root node (Response or Y)
- Each node is connected to a test that splits its set of possible answers into subsets corresponding to different test results
- Each branch carries a particular test result's subset to another node.
- Nodes that are at the end of branches are called terminal nodes and play a special role when the tree is used for prediction.



Decision Trees

“Induction”

- Decision tree algorithms go through two phases:
  - tree growing or splitting phase;
  - followed by tree pruning phase.

- The **tree growing** phase is an iterative process which involves splitting the data into progressively smaller subsets. Each iteration considers the data in only one node. The first iteration considers the root node that contains all the data. Subsequent iterations work on derivative nodes that will contain subsets of the data.

Decision Trees

"Induction"

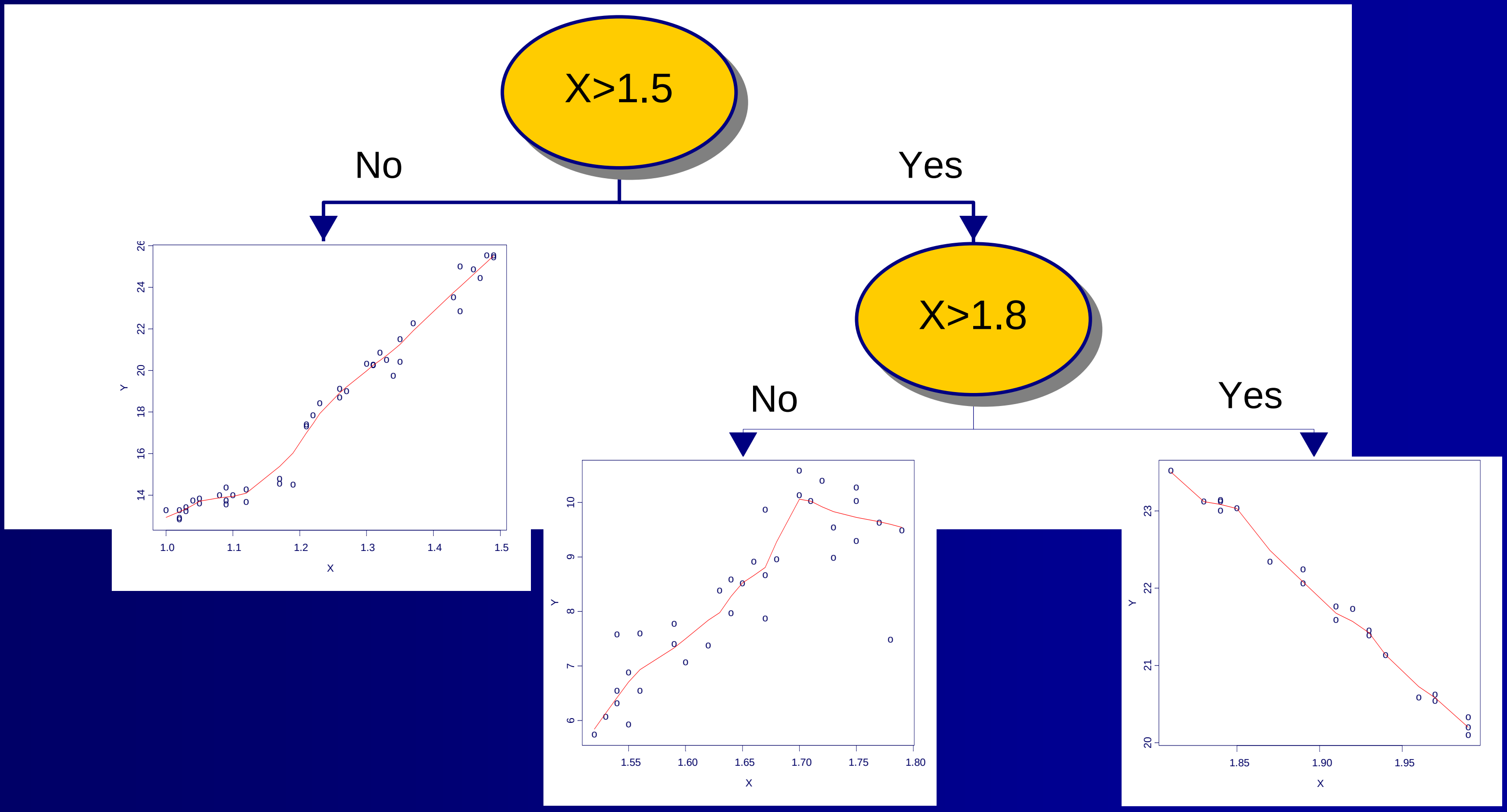
- Tree-building algorithms usually have several **stopping rules**. These rules are usually based on several factors including maximum tree depth, minimum number of elements in a node considered for splitting, or the minimum number of elements that must be in a new node.

- **Pruning** removes splits and the subtrees created by them.

- In some implementations, pruning is controlled by user, e.g., the computed difference between the resulting nodes falls below a threshold and is insignificant.

Decision Trees

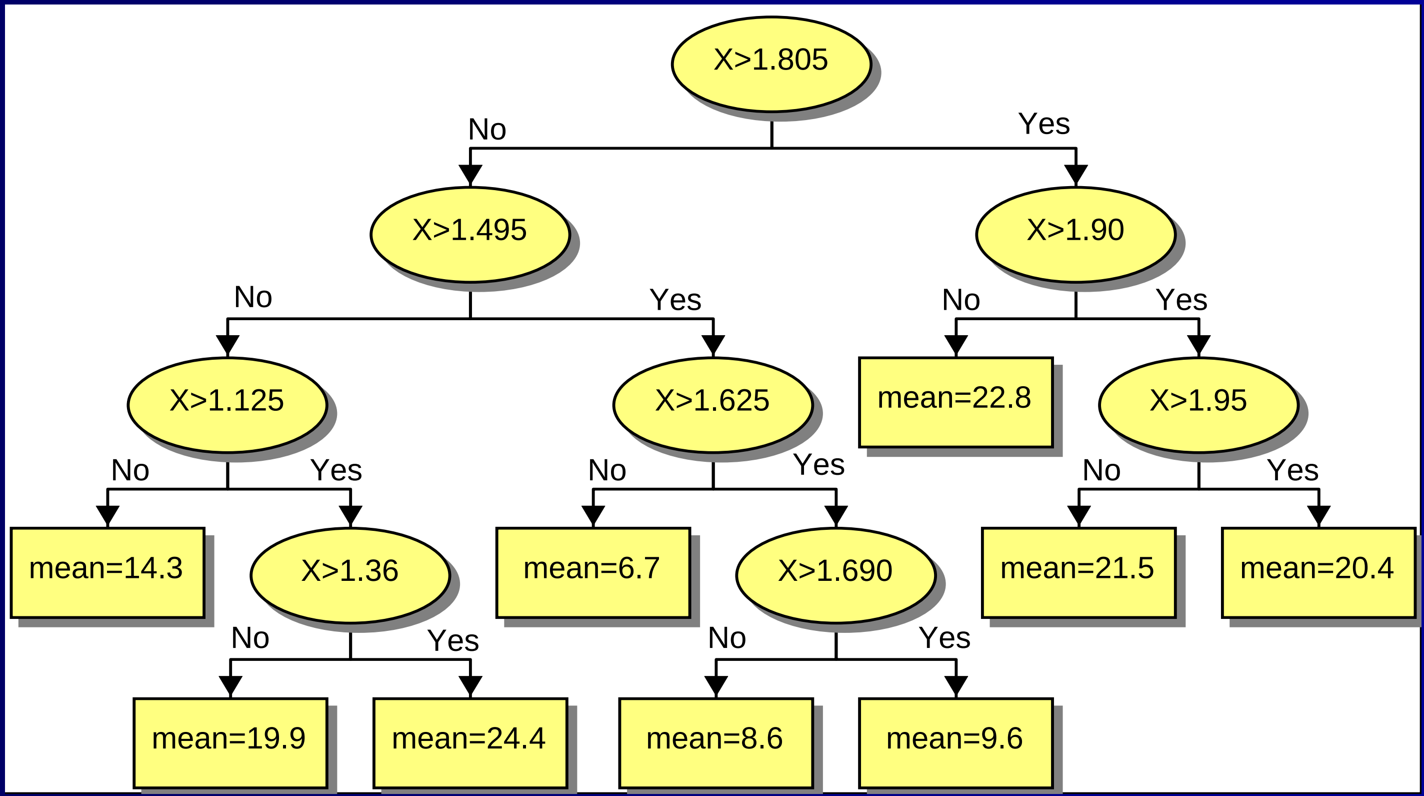
"Visually Interpretable"



Decision Trees

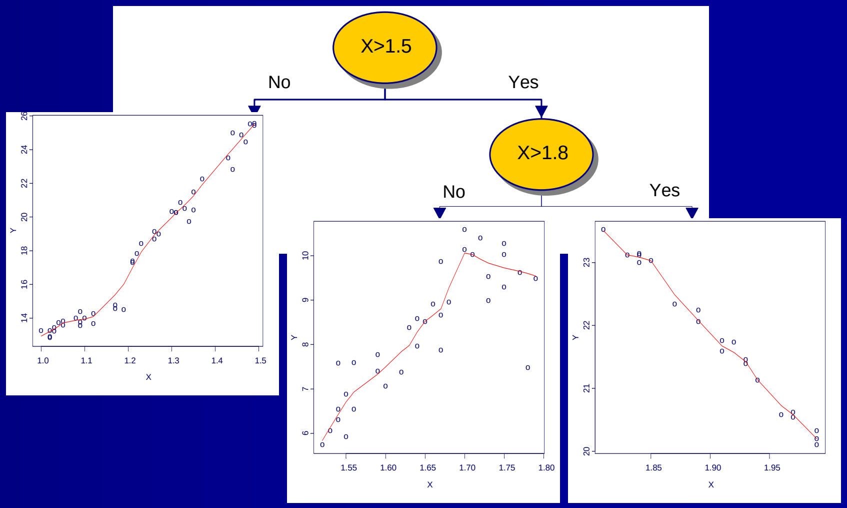
"Visually Interpretable"

- Tree model with many partitions



Decision Trees

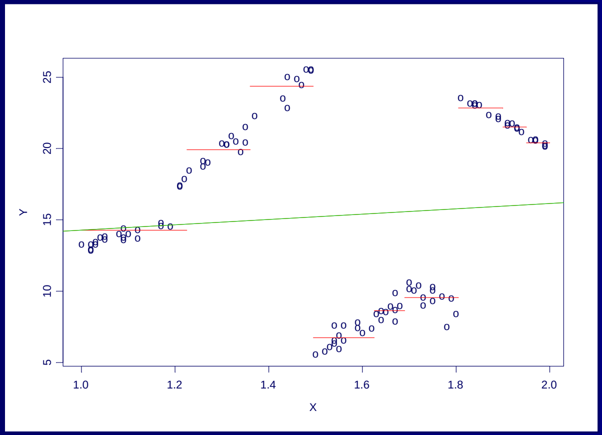
- Decision Tree
  - Categorical Data – “Classification Tree”
  - Numerical Data – “Regression Tree”
- Regression Tree
  - Piecewise estimate of a regression function
  - Constructed by recursively partitioning the data and sample space



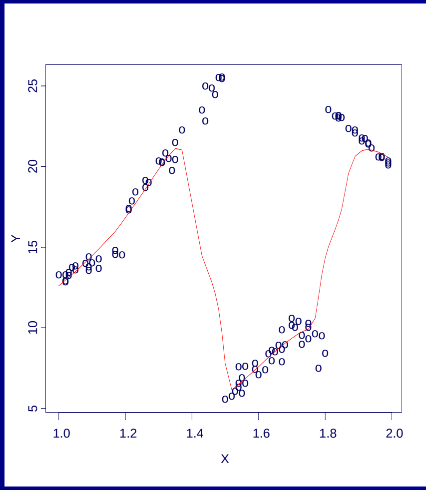
“Visually Interpretable Models”

Decision Trees

address the challenges of modeling data which are inherently heterogeneous



MLR fit vs. RT piecewise linear fit



LOESS

DT methods are in the spirit of Gleser (1996)'s **First Law of Applied Statistics**, "i.e., two individuals using the same statistical method on the same data should arrive at the same conclusion.

Decision Tree

Algorithms

- AID (*"Automatic Interaction Detection"*), Morgan and Sunquist (1963)
- CART© (*"Classification and Regression Trees"*), Brieman and Friedman (1984)
- MARS© (*"Multivariate Adaptive Regression Splines"*), Friedman (1991)
- M5, Quinlan (1992)
- FIRM (*"Formal Inference-based Recursive Modeling"*), Hawkins (1997)
- GUIDE (*"Generalized, Unbiased, Interaction Detection and Estimation"*), Loh (2002)
- CRUISE (*"Classification Rule with Unbiased Selection and Estimation"*), Kim and Loh (2001)

Decision Trees

"Classification Trees (CT) and Selection Bias"

- Most CT algorithms have a weakness of selection bias, i.e., when a variable appears in a split, it is hard to know if the variable is indeed the most important, or if the selection is due to bias.
  - Two sources of bias:
    - when variables differ greatly in their numbers of splits;
    - when variables differ in their proportions of missing values.

Methods

"Classification Trees from CRUISE"

FUTURE

- CART, QUEST, FIRM, CHAID, C4.5, FACT are biased toward selecting variables with more missing values.
- CRUISE:
  - trees often have prediction accuracy at least as high as those of CART and QUEST, two other highly accurate algorithms;
  - fast computation speed by employing multiway splits, precludes the use of greedy search methods;
  - free of selection bias;
  - sensitive to local interaction between variables, it has all the above properties with or without missing values in the learning sample.

Methods

"CRUISE"

- There are three split methods in CRUISE:
  - Univariate Splits
  - Linear Combination Splits
  - Univariate Splits with Node Models
- Univariate Splits:
  - Step 1: Selection of split variable
    - test the most significant pair, choose the most significant variable using a contingency table χ^2-test, select the variable with the smallest p-value;
  - Step 2: Selection of split point
    - If X is numerical perform a Box-Cox transformation and apply linear discriminant analysis (LDA) to the X values to find the split point (recall LDA is most effective when the data are normally distributed with the same covariance matrix)
    - If X is a categorical variable, it is first converted to a 0-1 vector and then processed as a numerical variable.

Methods

"CRUISE"

- Linear Combination Splits (has greater prediction accuracy with fewer terminal nodes, although this does not translate to improved interpretation):

  - Step 1: each categorical variable is transformed into a dummy vector and then projected onto the largest discriminant coordinate; this maps each categorical variable into a numerical variable.

  - Step 2: performs a principal component analysis (PCA) of the correlation matrix of the variables, principal components with small eigenvalues are dropped to reduce the influence of noise variables.

  - Step 3: LDA is applied to the remaining principal components to find the split.

Methods

"CRUISE"

- Univariate Splits with Node Models (fits a linear discriminant model to the best two-variable plot at each node):

  - If Univariate Splits result in a large tree or an extremely simple one (due to over-pruning), one solution is to employ linear combination splits but such splits may be difficult to interpret if they involve more than two variables.

  - Univariate splits with node models uses univariate splits and also fits a linear discriminant model to the best two-variable plot at each node. Such models can be used for class prediction which also simplifies the tree structure without sacrificing interpretability.

Results

"Tree Harvest vs. No Harvest"

Split type	Variable Selection Method	Split Method	# of Terminal Nodes	Important Variables	Re-substitution Misclassification Rate	Cross-validation Misclassification Rate
Univariate splits	Step 1	Exhaustive Search	4	q44_farmer; q41_years; q14_plan; q36_lndout;	0.3333	0.3960
		Linear Discriminant Analysis	5	q44_farmer; q41_years; q14_plan; q36_lndout; q38_live; q54_income;	0.3030	0.4000
	Step 2	Exhaustive Search	4	q39_arealife; q1_acres; q2_pctwd; q50_yrborn	0.3010	0.4000
		Linear Discriminant Analysis	2	q39_arealife; q1_acres	0.3535	0.3919
Linear combination splits			1		0.2909	0.3798

Results

"Tree Harvest vs. No Harvest"

Split type	Variable Selection method	Split Method	Pairwise Variable Selection Method	# of Terminal Nodes	Important variables	Re-substitution Misclassification Rate	Cross-validation Misclassification Rate
Univariate splits with node models	Step 1	Exhaustive search	MANOVA	2	q39_arealife; q45_impwdinc	0.3859	0.3939
		Exhaustive search	LDF	4	q44_farmer; q41_years; q14_plan; q36_lndout	0.2485	0.4040
		Linear discriminant analysis	MANOVA	2	q39_arealife q45_impwdinc	0.3859	0.3899
		Linear discriminant analysis	LDF	5	q44_farmer; q3_tracts q54_income; q30_prptxfair q45_impwdinc	0.3273	0.4141

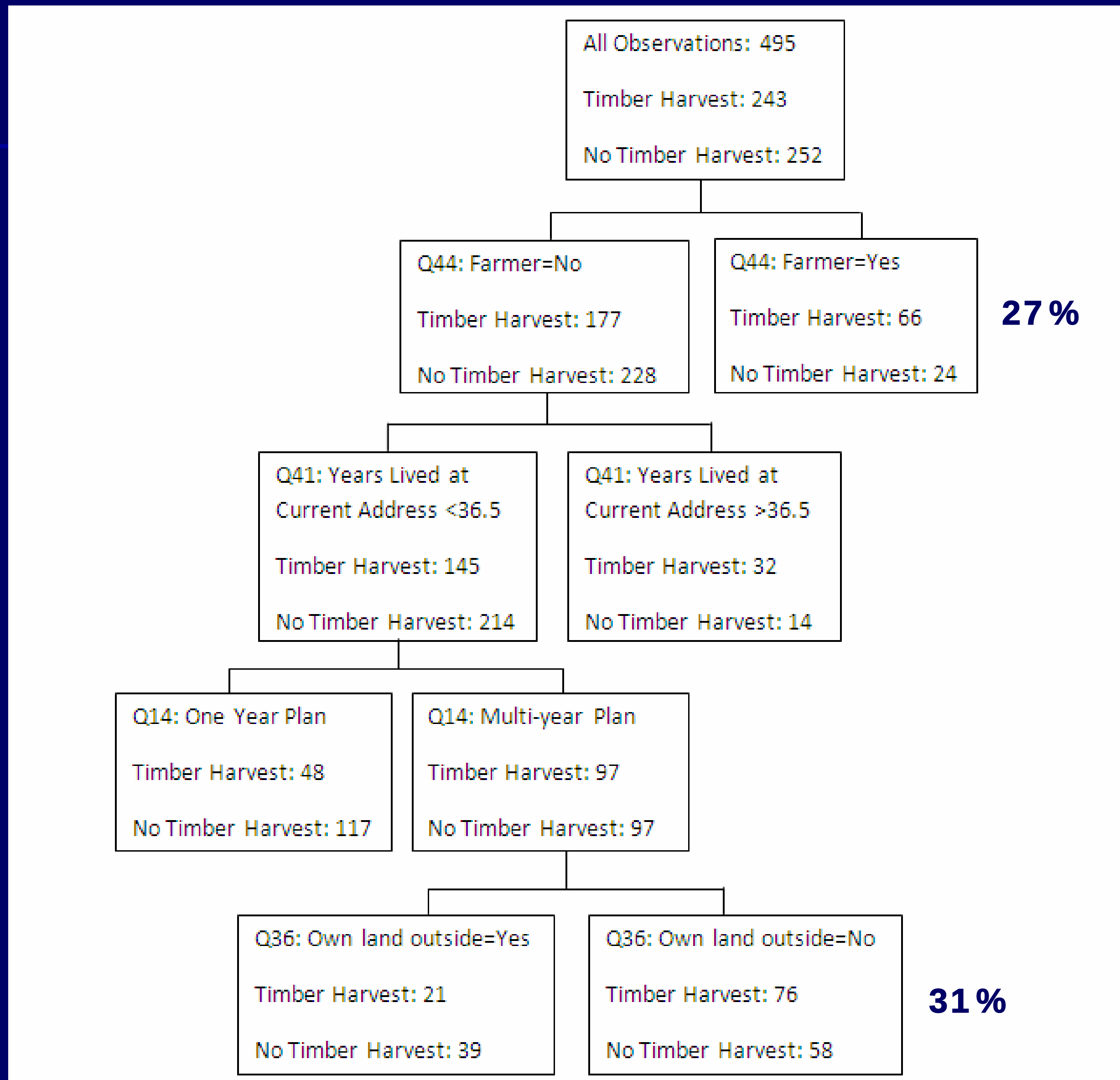
Results

"Tree Harvest vs. No Harvest"

Split Type	Variable Selection Method	Split Method	Pairwise Variable Selection Method	# of Terminal Nodes	Important Variables	Re-substitution Misclassification Rate	Cross-validation Misclassification Rate
Univariate splits with node models	Step 2	Exhaustive search	MANOVA	2	q39_arealife q45_impwdinc	0.3859	0.3939
		Exhaustive search	LDF	4	q39_arealife; q43_employ q54_income; q14_plan	0.3111	0.4182
		Linear discriminant analysis	MANOVA	2	q39_arealife q45_impwdinc	0.3859	0.3899
		Linear discriminant analysis	LDF	6	q44_farmer; q14_plan; q3_tracts q54_income; q45_impwdinc q48_educ	0.2909	0.4323

Results

"Classification Tree - Tree Harvest vs. No Harvest"



Results

"Classification Tree - Commercial vs. Non-Commercial Harvest"

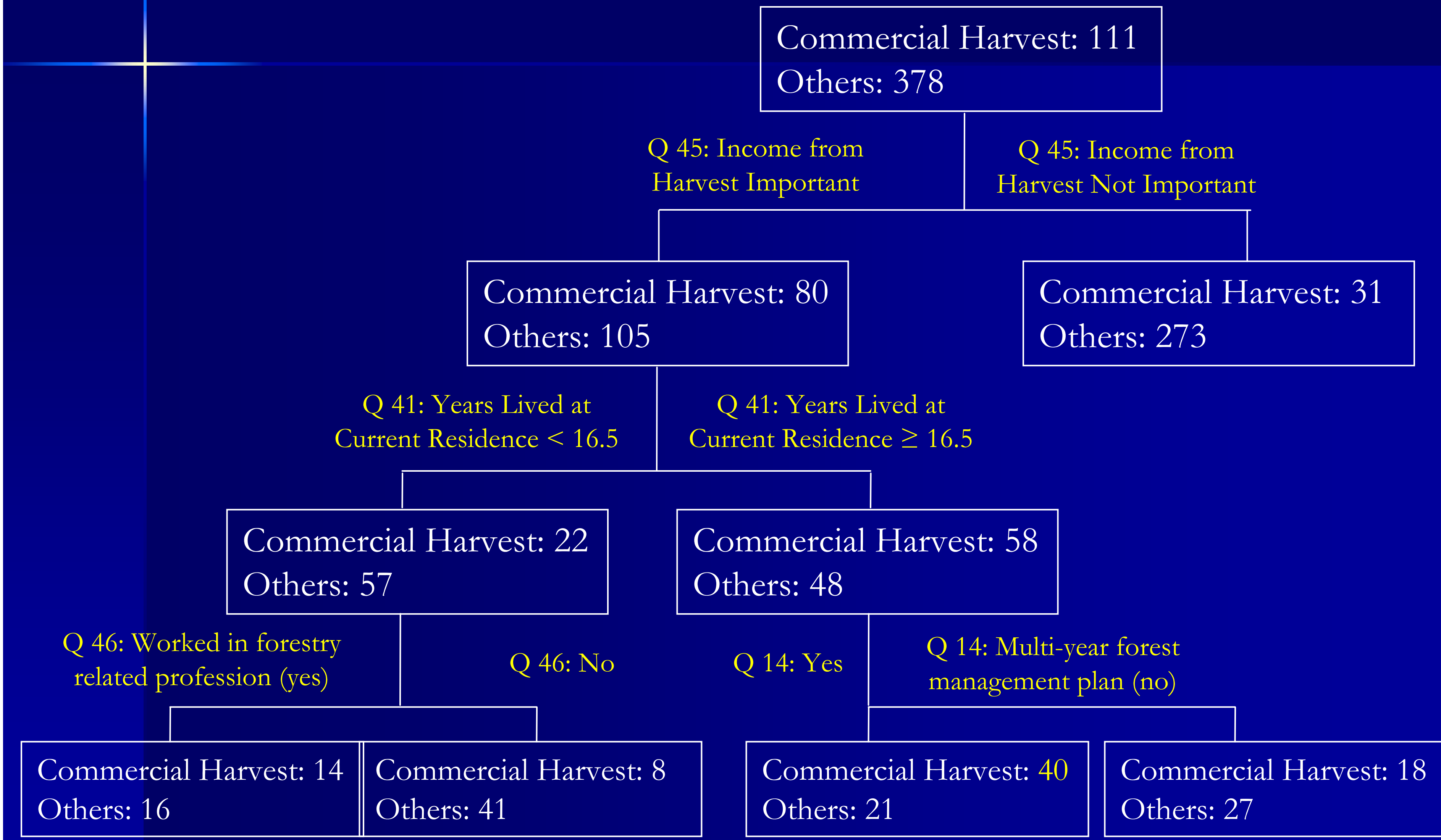
Split type	Variable Selection method	Split method		# of terminal nodes	Important variables	Re-substitution Misclassification Rate	Cross-validation misclassification Rate
Univariate splits	1D	Exhaustive search		2	q45_impwdinc	0.2911	0.3249
	2D	Exhaustive search		2	Q1_acres?	0.3038	0.3333
Linear combination splits				2		0.2152	0.3038
Univariate splits with node models	1D	Exhaustive search	MANOVA	2	q3_tracts; q45_impwdinc	0.3827	0.3911
	2D	Exhaustive search	MANOVA	2	Same tree as 1D method		

Results

"Classification Tree - Commercial vs. Non-Commercial Harvest"

Results

"Classification Tree - Commercial Harvest vs. All Others"



Regression Trees

"Predicting Mechanical Strength Properties of MDF and OSB"

FUTURE

- Investigated parametric and non-parametric regression tree models for the internal bond (IB) of medium density fiberboard (MDF) and parallel EI of oriented strand board (OSB)

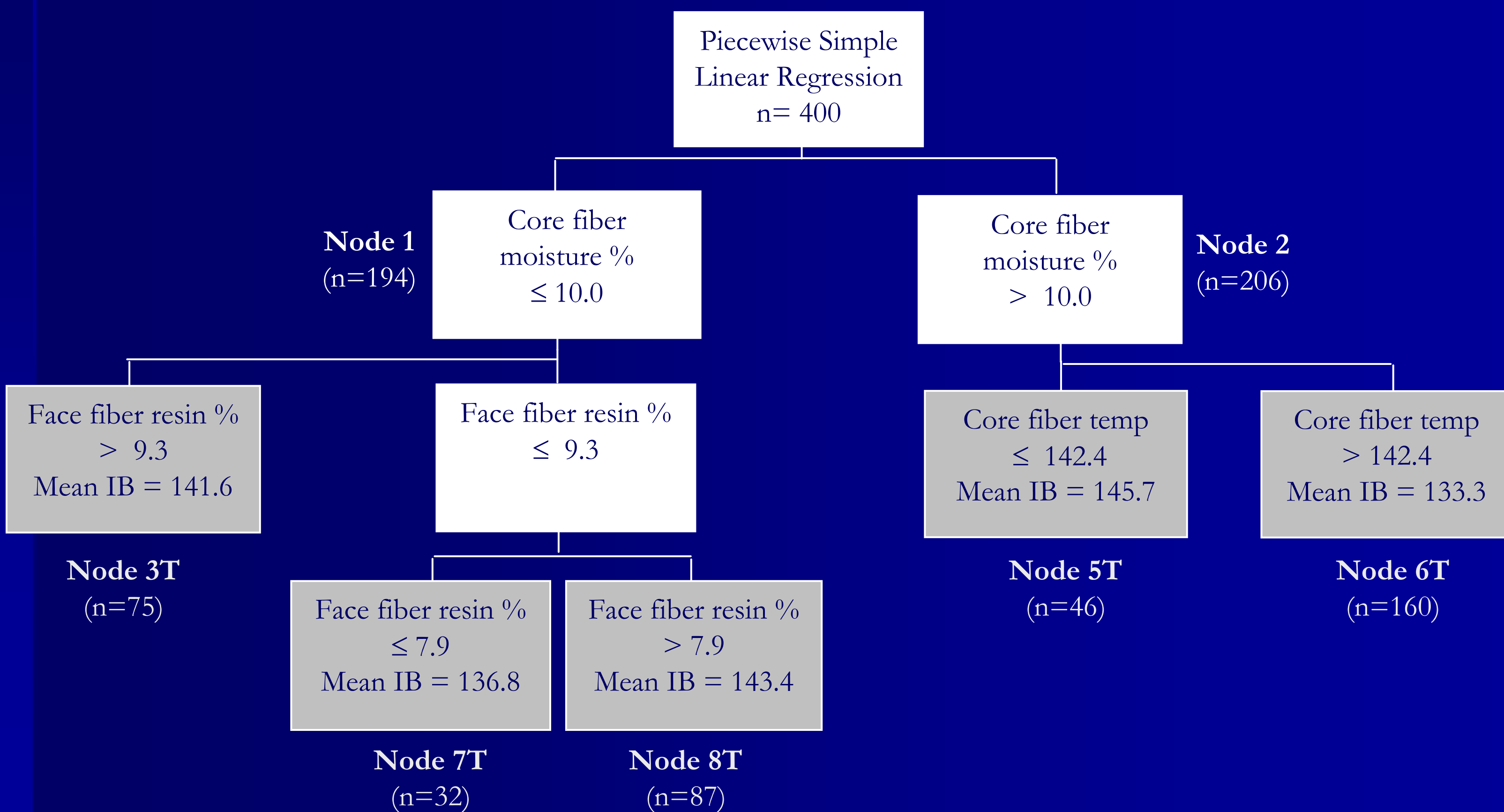
- Large time periods between destructive testing of strength properties and lack of knowledge of unknown causality lead to higher than necessary density targets and wood waste.

- Explored 1,335 regression tree models using GUIDE

Results

"MDF RT Model Information"

- **0.625" MDF –** Piecewise simple linear with v-fold cross-validation node pruning; note explanatory value of RT Model



Results

“MDF RT Models”

- **0.625” MDF** – Piecewise simple linear with v-fold cross-validation node pruning

Node 3T:	
SL Regression	**Coeff.**
Constant	-96.7
Press overall time set-point	1.20

Node 8T:	
SL Regression	**Coeff.**
Constant	120.8
Core refiner feeder screw spd.	0.036

Node 5T:	
SL Regression	**Coeff.**
Constant	295.6
Core fiber temperature	-1.36

Mean IB = 145.7 (highest)

Node 6T:	
SL Regression	**Coeff.**
Constant	118.5
Face refiner steam flow	-0.005

Mean IB = 133.3 (lowest)

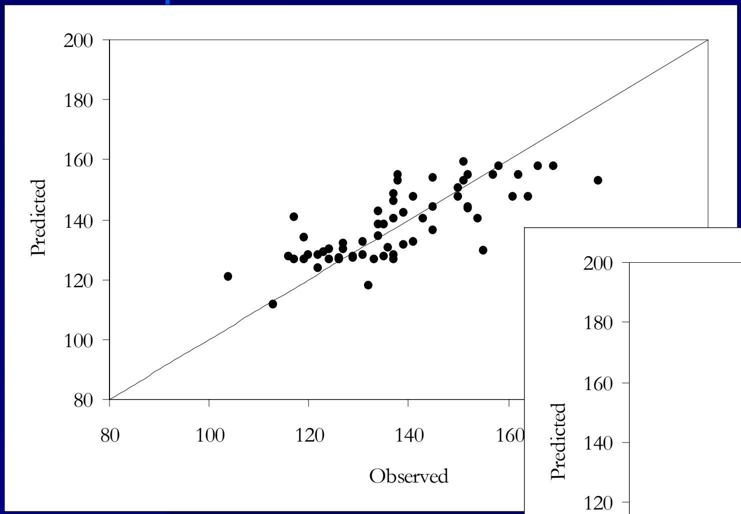
Node 7T:	
SL Regression	**Coeff.**
Constant	472.8
Swing digester pressure	-3.57

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

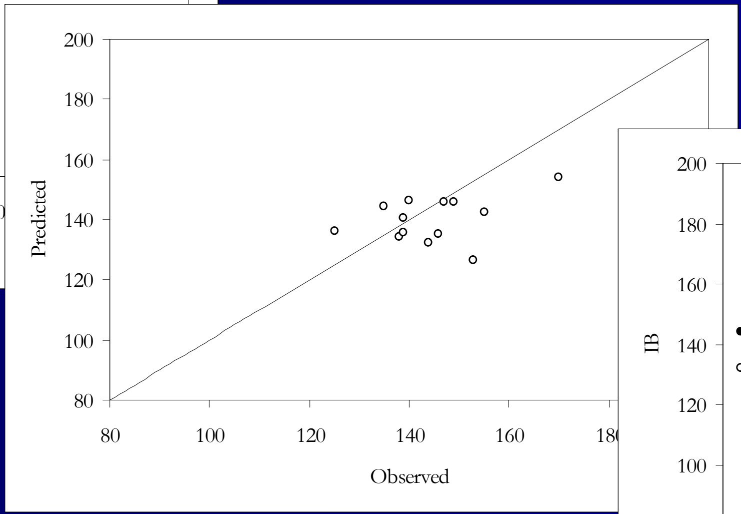
Results

"MDF RT Models"

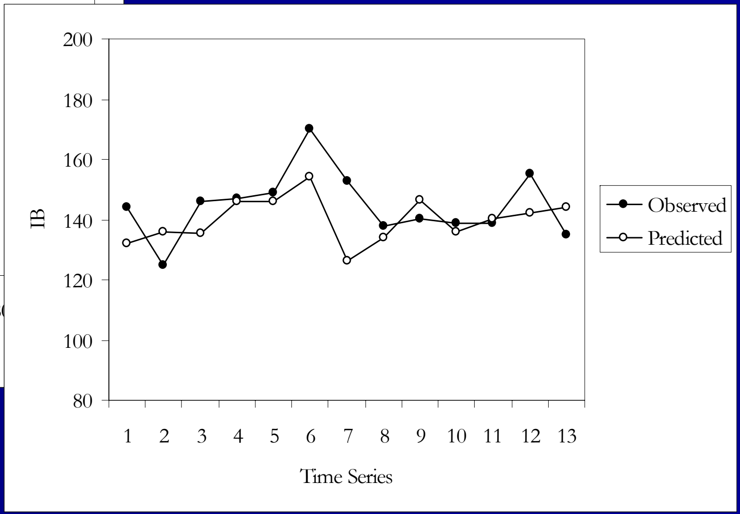
- **0.625" MDF**



Training data set



Validation data set

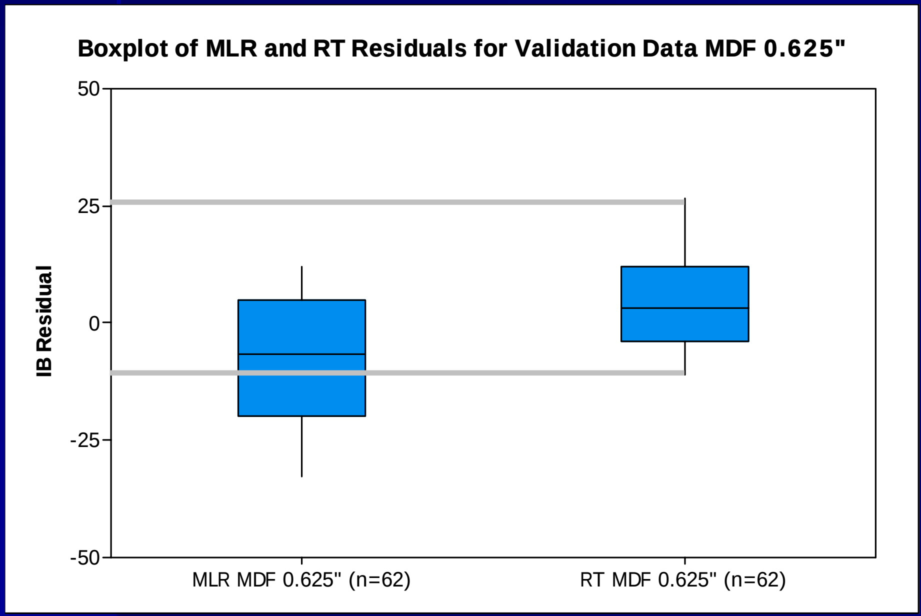


Validation data set as a time series

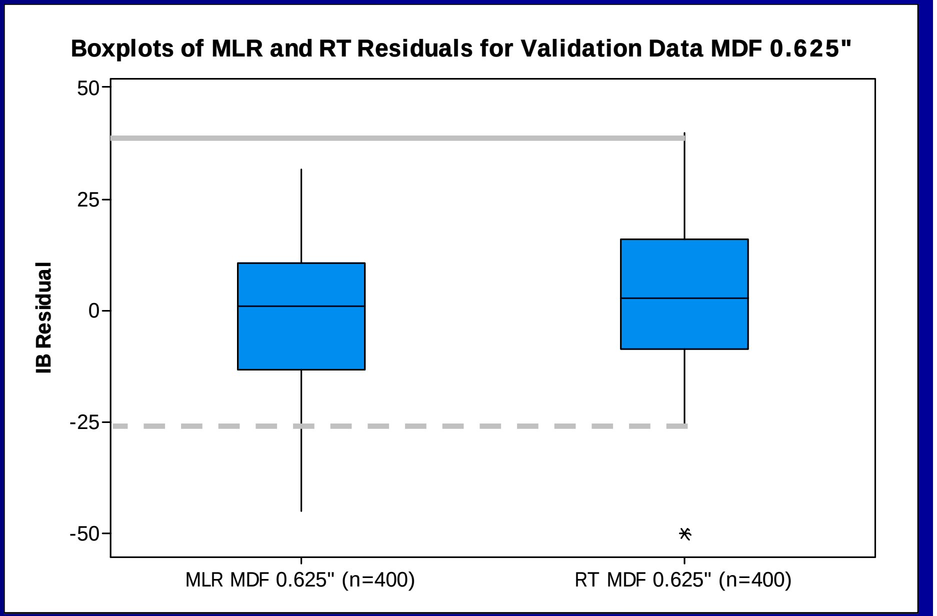
Results

"MDF RT Models"

- **0.625" MDF** Box Plots of Residuals (difficult product to model – Young and Guess 2002; Young and Huber 2004)



Short record length

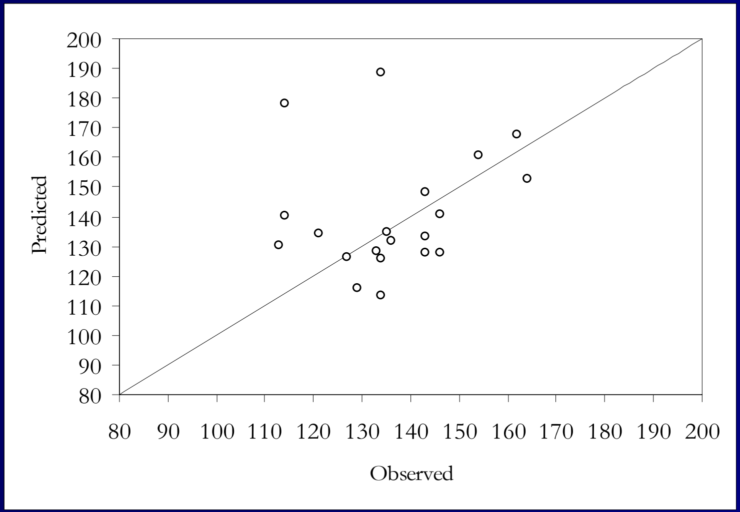


Full record length

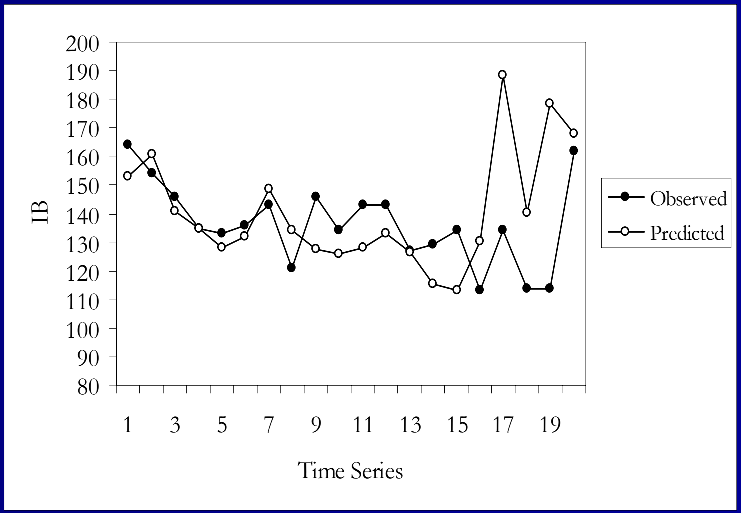
Results

"MDF RT Models"

- **0.750" MDF** – Non-parametric quantile regression model without node pruning



Validation data set

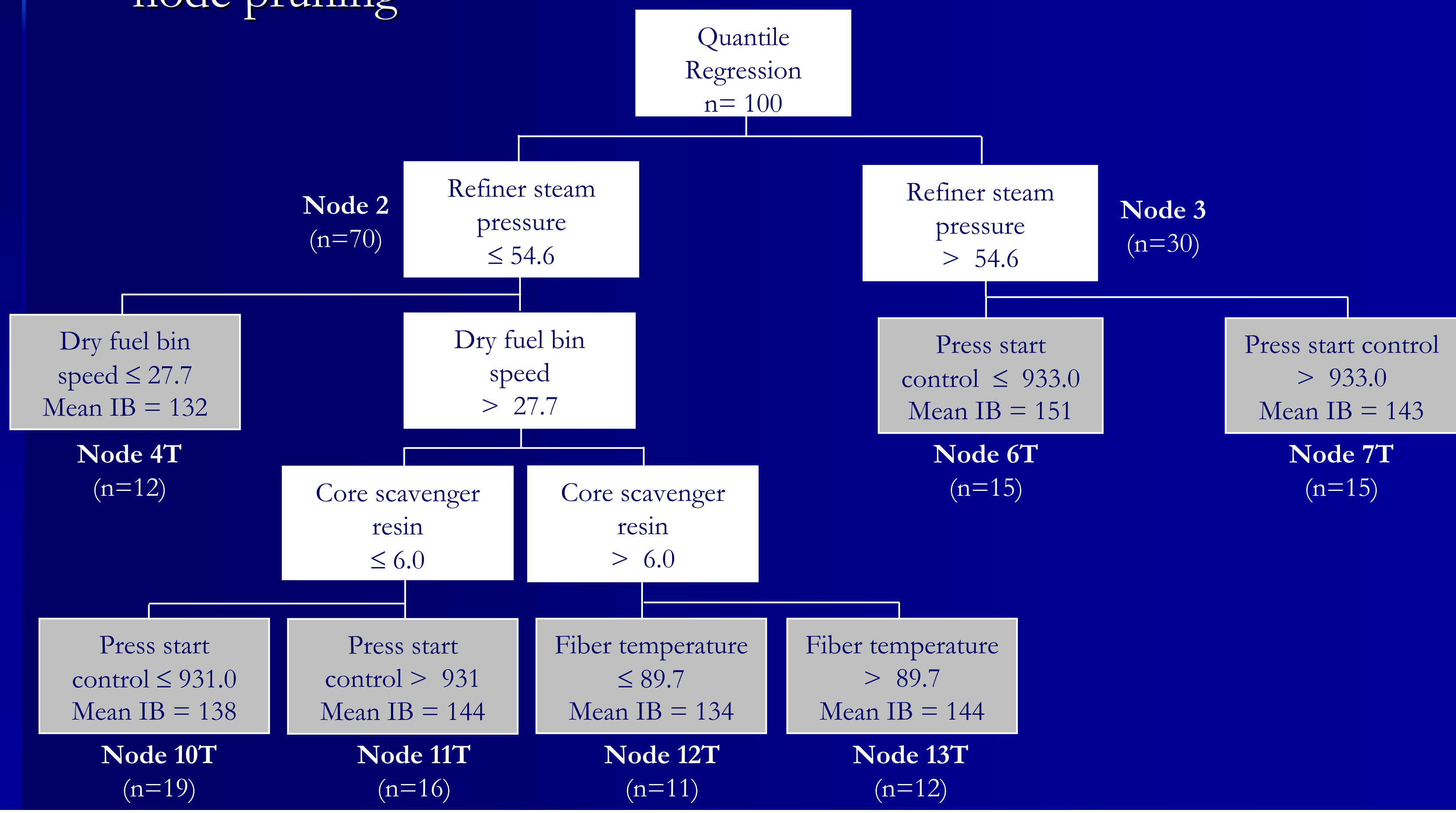


Validation data set as a time series

Results

“MDF RT Models”

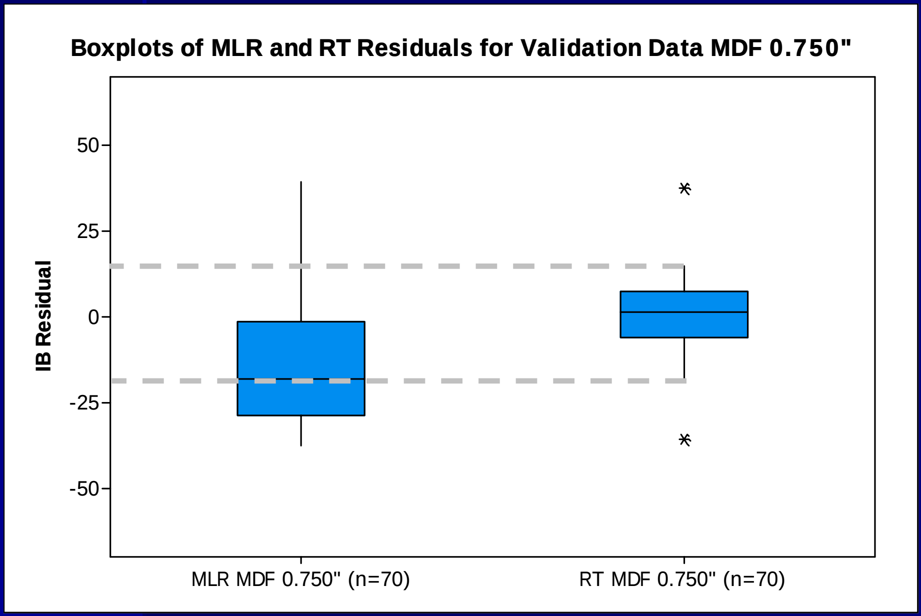
- **0.750” MDF** – Non-parametric quantile regression model without node pruning



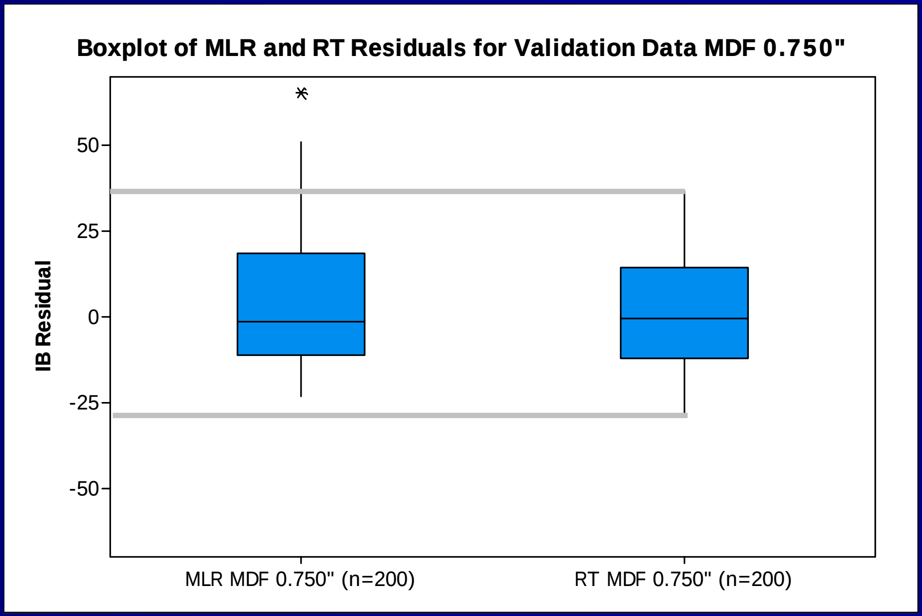
Results

“MDF RT Models”

- **0.750” MDF** Box Plots of Residuals



Short record length

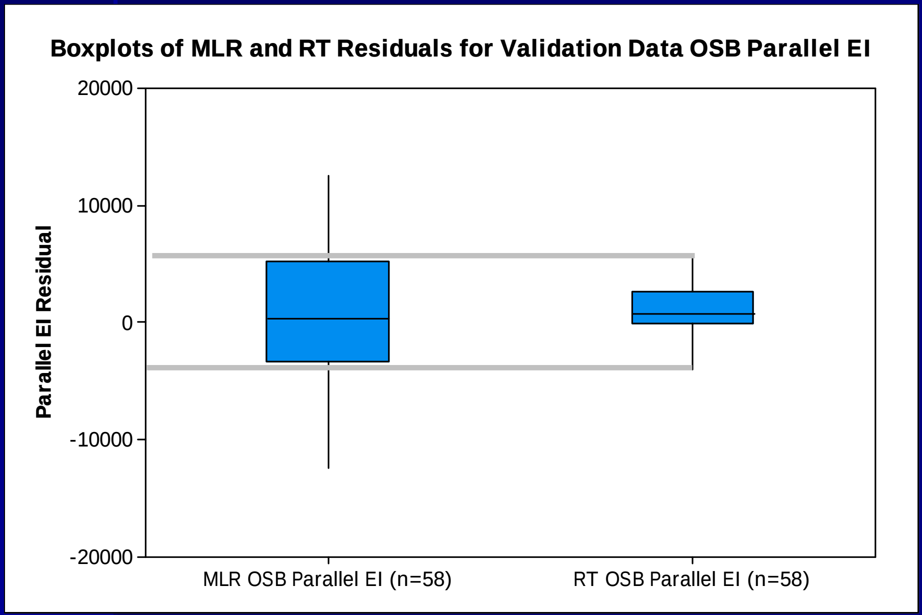


Full record length

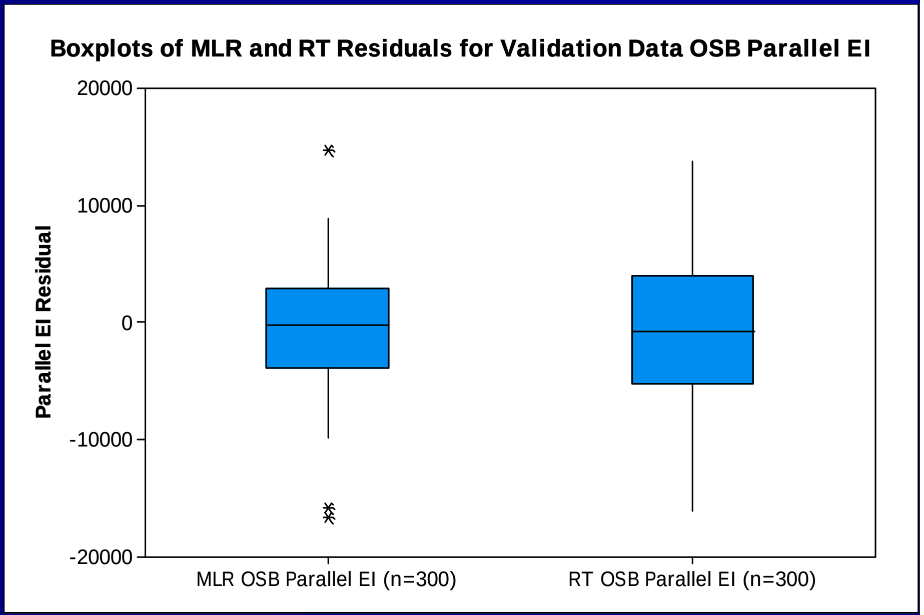
Results

"OSB RT Models"

- **OSB Parallel EI Box Plots of Residuals**



Short record length



Full record length

Results

"GUIDE's Ranking of Significant Regressors"

Useful for Practitioner (Top Ten Regressors):

OSB Internal Bond (n=300)		OSB 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

Conclusions

- Survey of Forest Landowners – Classification Tree
  - Commercial vs. Non-Commercial Harvest:
    - Forest landowners that conduct commercial tree harvests, it's all about the money ($$$) - no surprise
  - Timber Harvest vs. No Harvest:
    - 27% of timber harvests were by farmers; however, 31% of non-farmers that lived at the property < 36.5 years with a multi-year management plan, without any other land would harvest timber
  - Commercial Harvest vs. All Others:
    - 36% of commercial tree harvests were by owners with a multi-year management plan, lived at the current residence for at least 16.5 years and thought income from the harvest was important.
- Classification Trees are attractive by accounting for hidden interactions within heterogeneous data and are easy to interpret by practitioners

Conclusions

- Wood Composites Data:
  - RT models outperformed MLR models (RMSEP, residual plots, predictions in validation data)
  - RT models may improve model quality when Y is non-normal, e.g., OSB for short record lengths
  - RT models identified hidden interactions
  - Easy to interpret for practitioner in mill
  - GUIDE ranking of regressors very useful for the practitioner

Decision Trees

"Aligned with Philosophy of Accurately Quantifying Variation"

Box (1979): "All models are wrong but some are useful."

Quantile Regression
n= 100

Node 2 (n=70)
Refiner steam pressure ≤ 54.6

Node 3 (n=30)
Refiner steam pressure > 54.6

Dry fuel bin speed ≤ 27.7
Mean IB = 132
Node 4T (n=12)

Dry fuel bin speed > 27.7

Press start control ≤ 933.0
Mean IB = 151
Node 6T (n=15)

Press start control > 933.0
Mean IB = 143
Node 7T (n=15)

Core scavenger resin ≤ 6.0

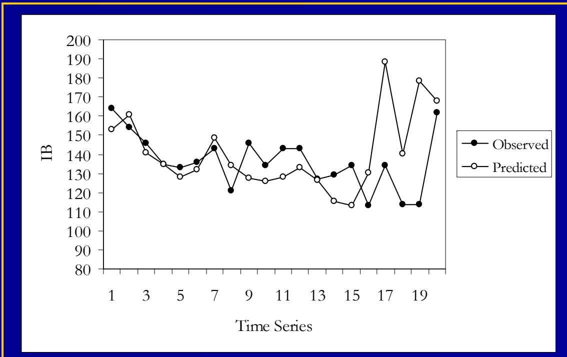
Core scavenger resin > 6.0

Press start control ≤ 931.0
Mean IB = 138
Node 10T (n=19)

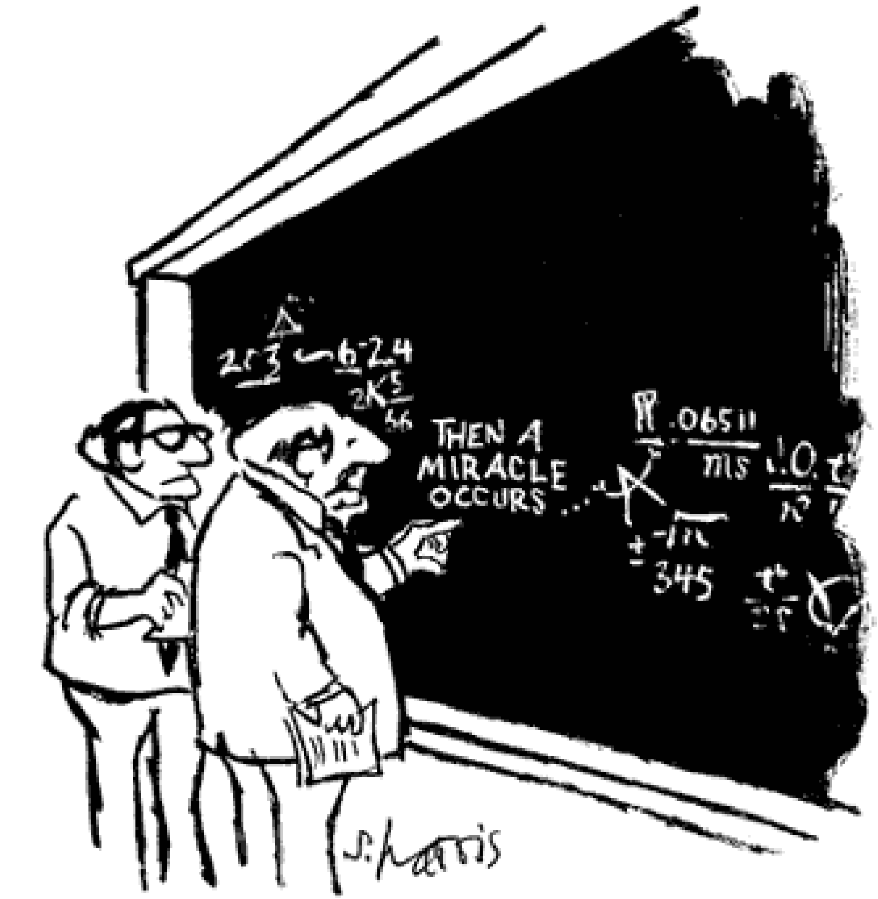
Press start control > 931
Mean IB = 144
Node 11T (n=16)

Fiber temperature ≤ 89.7
Mean IB = 134
Node 12T (n=11)

Fiber temperature > 89.7
Mean IB = 144
Node 13T (n=12)



Questions



"I think you should be more explicit here in step two."