Spring 2015

Sensitivity of mixed models to computational algorithms of time series data

Gunaime Nevine

SENSITIVITY OF MIXED MODELS TO COMPUTATIONAL ALGORITHMS OF TIME SERIES DATA

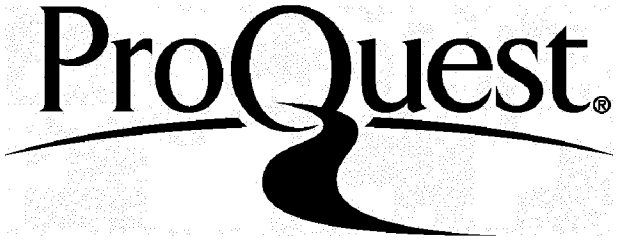
by

Nevine Gunaime, B.S., M.S.

A Thesis Presented in Partial Fulfillment
of the Requirements of the Degree
Doctor of Philosophy

COLLEGE OF ENGINEERING AND SCIENCE
LOUISIANA TECH UNIVERSITY

May 2015



LOUISIANA TECH UNIVERSITY

THE GRADUATE SCHOOL

January 16, 2015

Date

We hereby recommend that the dissertation prepared under our supervision by Nevine Gunaime

entitled SENSITIVITY OF MIXED MODELS TO COMPUTATIONAL ALGORITHMS OF TIME SERIES DATA.

be accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Computational Analysis and Modeling

Supervisor of Dissertation Research

Head of Department

CAM
Department

Recommendation concurred in:

Advisory Committee

Approved:

Director of Graduate Studies

Approved:

Dean of the Graduate School

Dean of the College

GS Form 13a
(6/07)

ABSTRACT

Statistical analysis is influenced by implementation of the algorithms used to execute the computations associated with various statistical techniques. Over many years, very important criteria for model comparison has been studied and examined, and two algorithms on a single dataset have been performed numerous times. The goal of this research is not comparing two or more models on one dataset, but comparing models with numerical algorithms that have been used to solve them on the same dataset.

In this research, different models have been broadly applied in modeling and their contrasting which are affected by the numerical algorithms in different SAS software procedures. Those model-algorithm combinations have been tested separately on three datasets: Box and Tiao Ozone data, simulated Tree Height-Age data, and Longleaf Pine Tree Diameter-Height (Taper) data.

Furthermore, results presented will be inclusive in describing the general conclusions by comparing the algorithms, then analyzing the behavior and performance of every algorithm based upon the verification and the results we have. In addition, algorithms' relative and absolute strengths and weaknesses will be identified. The decision will stand on well-known model selection criteria: Akaike Information Criterion (AIC), Schwarz's Bayesian Criterion (SBC), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants to the Prescott Memorial Library of Louisiana Tech University the right to reproduce, by appropriate methods, upon request, any or all portions of this Thesis. It is understood that "proper request" consists of the agreement, on the part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this Thesis. Further, any portions of the Thesis used in books, papers, and other works must be appropriately referenced to this Thesis.

Finally, the author of this Thesis reserves the right to publish freely, in the literature, at any time, any or all portions of this Thesis.

Author __ Nevine Gunaime _____

Date May 1, 2015

DEDICATION

To my family, my husband, and my lovely son.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I am grateful to my supervisors, Dr. Bogdan Strimbu and Dr. Mihaela Paun, for their extensive support in providing direction for my research, feedback on my work, and for their patience and encouragement.

In addition to my advisors, I would like to thank my thesis committee: Dr. Katie Evans, Dr. Weizhong Dai, Dr. Chokchai (Box) Leangsuksun, and Dr. Dexter Cahoy for all of their encouragement, insightful comments, and challenging questions.

To my mother, Thuraya Gareeb, who encouraged me to continue my graduate studies, and to my father, Mohamedali Gunaime, who called me every day to support me and make sure I finished.

To my labmate, Victor Strimbu, who helped me out with some of the technical issues. Special thanks to my friend, Krystal Stevens-Woods, for volunteering her time in helping to finalize this research.

I also want to thank the School of Forestry for providing me with a very pleasant lab to facilitate my research.

To my husband Ihsan Benten; I thank you for being a great supporter for me throughout all of my research journeys and my life. Finally and most importantly, to my lovely handsome boy who missed his patient mom for many long days.

CHAPTER 1

INTRODUCTION

Scientists possess the knowledge and understanding about the world, and they are capable of converting what they know into models to make complex information easier to realize. Stockburger (1996) defined a model as "a representation containing the essential structure of some object or event in the real world". Models are usually solved with algorithms, "which have a well-defined computational procedure that takes some value as input, and produces some values, as output" (Cormen *et al.*, 2009). Algorithms are used to find models, but many algorithms are available to execute this task. However, algorithms depend on implementation, which influence the performance of the algorithm (McGeoch, 1996). Therefore, algorithm comparison papers have become important (Coffin & Saltzman, 2000); however, they do exist in many scientific fields like engineering, computer science, and biology. Often, algorithm comparisons are based upon solution qualities or accuracy of their results.

From a general viewpoint of statistical results, any model depends on the numerical algorithm used to solve that model, which can be proven by testing more than one algorithm to solve the same model and compare the results. For some models, coefficients (parameters) of model variables intervene and play a role in affecting results, particularly if dealing with nonlinear models (Seppelt & Richter, 2005). Applying statistical analysis is a very strong implementation for assessing any algorithm's

performance. Statistics is the most important part of mathematical science, besides searching theoretic topics, which is used by researchers in many scientific branches to assemble, arrange, examine, and make implications of the data. Statistical models can help researchers in making convenient decisions on data in their work (McCullagh, 2002). When applying a statistical model on any type of dataset, it is best suited to execute through all types of algorithms associated with that model. Applying the same model on the same datasets with different algorithms yields consequences inclusive of different outputs.

One of the most broadly used tools for investigating multilevel data in statistics is regression analysis, which represents a relationship between variables (Montgomery, Peck, & Vining, 2006). This relationship describes the causative effect of one variable upon another. Regression equation is an approximation for the original relationship between variables. Furthermore, regression model also referred to as empirical model, can be a simple relation between one predictor variable and the response variable, or multiple which is between more than one predictor and the response variable. The best approach to replicate empirical models is the fix and random effects (Schielzeth & Forstmeier, 2009). When both effects appear together in a model, it is called the mixed effect model, and fixed effect is something the experimenter controls and is often repeatable. For example, a drug is tested in which experimenters provide a specified amount of a drug to one group and the other group gets the placebo. In this case, the drug or placebo is a fixed variable. Fixed effect is found to be more powerful because it produces smaller standard errors. In contrast, when an experimenter draws an individual (at random) from a population, random effects take place. As a means of providing a

comprehensive outlook, random effect can occur during a clinical trial by choosing random patients from one group for drug testing, which can be male or female; in this case, the gender is a random variable (Gelman, 2005). In essence, random effect is less powerful because it produces large standard errors.

Another important topic in statistics called time series must be defined to get a full understanding of this research; thus "time series is an ordered sequence of observations x_t ,each one being recorded at specific time t" (Brockwell & Davis, 1996). Time series analysis has been a significant topic in a variety of research fields such as business, agriculture, economics, engineering, geophysics, medicine, and social sciences (Wei, 2006).

1.1 Problem Statement

The applying scientists are focused on the details of their research and not on the particularity of the statistic methodologies. The quick improvement in computing technology has simplified the use of models and problem solving by using computer softwares. Consequentially, inducting researchers in applied sciences have a tendency to use the built-in default setting, which focuses on statistical softwares of their choice. However, we argue that the results are sensitive to settings embedded in the default options which can result in misleading or undefeatable conclusions. Furthermore, practitioners who focus on areas not related to statistics use the default setting of statistics software built for them. By using built-in procedures in any software, models can be fitted one equation at a time through different algorithms on the data, and every one of those algorithms has its own adjustment and derivation. On the other hand, we agree that the default option does not necessarily supply (indefinable) the results and researchers

should be aware of using this option. Unfortunately, this valuable information is often omitted in research reports, and from this point of view, the significance of studying the variation outcomes is a result out of applying different algorithms besides the default. Comparing algorithm papers are very wide spread, but unfortunately, most of them contain statistical summaries with few data analyses (Coffin & Saltzman, 2000). Data analysis is represented by EDA, which is an abbreviation for Exploratory Data Analysis, such as extracting important variables, detecting outliers, testing underlying assumptions, and determining optimal factor settings (Tukey, 1977).

In reality the exact true model of a given data set is unknown, but it could be predicted by practicing (running) well-known models or newly developed ones that are suitable for a research data numerous times by initiate different values for the model's parameters until the results meet the researcher expectation, such as the model converges and achieved the lowest mean squared error.

To be more precise, this issue is explored by using models broadly applied in forest growth and yield modeling, in addition to comparing their effects. For every model, four algorithms in the time series autoregressive process and three algorithms in the time series ARIMA model in Statistical Analysis System (SAS software) will be applied and assessed. As previously mentioned, model-algorithm combinations have been tested separately on three datasets: Ozone data, Tree Height-Age data, and Height-Diameter (Taper) of Longleaf Pine trees.

1.2 Objectives

The objective of this research is to assess the impact of using different algorithms under the same model. In other words, investigating how the data are sensitive to

algorithms based on respective models. Models' sensitivity terminology means describing how much the model output results are affected by changing the algorithms used by software procedure to execution that model.

This research, however, provides a clear understanding to the behavior and performance of each algorithm used to solve mixed models on the time series data. Furthermore, strengths and weaknesses will be identified for algorithms both relative and absolute. Above all, this study compares algorithms to determine which are best and the worst under each model.

In statistics, there is no miraculous procedure to provide the best model. Nevertheless, there are some criteria bench marks that should be measured to help compare two or more algorithms and choose the best. In this research, Akaike Information Criterion (AIC) and Coefficient of Determination (R^2) are the two guide methods for selecting the best and worst models which is called multivariate pairwise tests used to compare the algorithms performance; thusly, in such a case, two or more vectors are used for comparisons (Yildiz, Aslan, & Alpaydin, 2011). Therefore, the best and worst algorithm comparisons for similarities and differences will be investigated; comparing algorithms is done by classification errors (residuals) based upon different factors for each dataset.

Even though numerical algorithms for solving any model most of the time leads to the same conclusions, they still have major differences between them that makes one of them performs better than the rest. There are many reasons that make these algorithms vary and even some of them more preferable than others. One of the main important criteria that make one algorithm more preferable is time consumption, in which some

algorithms require periods of running time to be executed to solve the model better than other algorithms to produce results (Coffin & Saltzman 2000). In some research, especially in computer sciences, the running time is a very important factor in comparing algorithms, and researchers in that field utilize asymptotic notation for computing algorithm complexity by measuring the running time which should be as small as possible (Cormen, Leiserson, Rivest, & Stein, 2009). In statistics, some algorithms do not show very important details about the coefficients such as AIC or p value, which is very significant in explaining any model. One more important distinction in statistics visualization is a very important way for giving feasible meaning of a huge number of data which helps to get a clear understanding about the data is behavior. Some algorithms do not show graphics to demonstrate the tendency in the data or differences between variables, but rather show only numbers which force the users to utilize their memory for comparing, instead of visual explanation. Unfortunately, this depends on user memory, which makes the assessment imprecise and decision inaccurate since it differs from one user to another.

1.3 Hypotheses and Assumptions

Two assumptions and hypotheses were proposed to direct this research. The assumptions used to clarify the time series data are:

1- The future is unknown, but it is predictable.

2- Residuals of any Time Series data has at least one organized pattern.

The previous two assumptions were set up as starting points for testing the two hypotheses of this study, and these two hypotheses are:

1- The default option algorithm under any statistical software procedure is not always the perfect choice for a predefined model.

2- Model Selection task gives different results for selecting the best models among a set of candidate algorithms, depending on the selection criteria chosen.

To test the hypothesis, the research presented in this thesis considered three datasets:

1- Impact of computational algorithms on modeling Ozone.

2- Impact of autoregressive algorithms in modeling auto-correlated data.

3- Analytical assessment for autoregressive process impact on Height-Diameter (Taper) data of Longleaf Pine trees.

1.4 Thesis Structure

The objective of the research was accomplished by combining the results of three datasets. Setting models and numerical algorithms were used to solve them to form one diagnostic platform, which enables the identification of the best algorithms associated with each model, and are presented individually as three chapters within this thesis. One chapter is used as a background to define important facts, concepts, and models which are applied in this research. The rest of this thesis is organized into five chapters, followed by the details and the outlines for the contents of each chapter to address the research objective.

Chapter 2 provides a background of the main statistical concepts and models which are used in this research with their formulas, helping to obtain a greater understanding for the research steps. Section 2.1 provides a brief introduction about the chapter, and Section 2.2 addresses the regression models which are divided into Subsection 2.2.1 simple regression model, Subsection 2.2.2 multiple regression model

with Subsection 2.2.2.1 multicollinearity, and Subsection 2.2.3 mixed effect model. Section 2.3 is about the time series model, discussing auto regressive (AR) models in Section 2.3.1 and autoregressive integrated moving average (ARIMA) models explained in Section 2.3.2. The most significant part of this research is the selection criteria, which is a leading guide for selecting the best models to be discussed in Section 2.4 by having two types of model selection methods: Akaike Information Criteria (AIC) and Coefficient of Determination (R^2).

Chapter 3 serves three purposes based on Box and Tiao's data (1975). First, the intervention model analysis for the times series data is explained; and secondly, discussing the meaning of ARIMA procedure in SAS 9.3 (SAS Institute Inc., 2012). Research hypotheses were then tested by using the Ozone data by applying the ARIMA procedure and its three algorithms. Chapter 3 mimics the work completed on the Ozone data by Box and Tiao (1975), which was applied by using the ML algorithm of ARIMA process and extended to be used with CLS and ULS. Subsequently, the results of ARIMA procedure and applied in the autoregressive procedure in SAS was used. Finally, by using AIC and R^2, the best model was chosen among the final candidate models.

Chapter 4 will start with computational methods and numerical algorithms' significance. Next, the description for simulation data translates the theoretical results into real practice. Afterward, there will be an in depth explanation of the Height-Age data generated by using SAS, and clarification of the autoregressive procedure concept in SAS. Next, applying the simulated data to two models have been utilized in forest growth and yield modeling: Chapman Richard and Wyckoff models, beside Schumacher and Polynomial models. The results of those four models are implemented in the autoreg

procedure in SAS, then they are investigated and compared to its four algorithms: ML, ULS, ITYW, and YW (which is the default option). Exploring the models, as a result of each algorithm combined with the autoregressive procedure, will be completed through two stages, according to the selection of the best model based on AIC and R^2.

Chapter 5 begins with a brief introduction about model-algorithm comparisons, followed by a comprehensive discussion describing the Longleaf Pine tree, what they are and where they are spread. Then, an explanation of what the taper stands for, its significance, and usage with the other collected attributes will be analyzed. An assessment of the five taper model results was based upon ninety-one plots: Bruce, Coffre, Jimenez, Kozak, and Munro (Rojo, PeaStem, & Sanchez-Rodriguez, 2005). Next, four autoregressive numerical algorithms will be investigated based on residual results from the taper models, and the final most favorable models are chosen according to AIC and R^2.

Chapter 6 concludes the thesis by connecting and summarizing earlier chapters with significant results and decisions of the research. It will also discuss strengths and weaknesses throughout all major discoveries while showing the hypotheses of the thesis by relating them to algorithms. This research discusses the impact of applying different numerical algorithms on some important forest growth and yield models. Additionally, all tests and outcomes are examined theoretically and future researchers can use findings by testing other types of models with different datasets.

CHAPTER 2

BACKGROUND AND REVIEW

2.1 Introduction

This chapter is an extensive review of some basic statistical models used in this research, with explanation of their definitions, formulas, and graphs. These models will lead us to discuss and investigate the research findings and go over the main points of the data to support the methodology of the research.

2.2 Regression Models

The regression model is a statistical technique for evaluating and assessing a relationship between variables. This relationship is represented by an equation between a response (dependent) variable and one or more predictor (regressor or independent) variables (Draper & Smith, 1998). The regression model can be linear or nonlinear and is divided into two main types: simple regression model and multiple regression model.

2.2.1 Simple Regression Model

The simple regression model is an investigation of how the changes in one variable affect another. The equation of a straight line relating two variables is $y = \beta_0 + \beta_1 x + \varepsilon$, where y is the dependent variable, x is the independent variable, β_0 and β_1 are the intercept and the slope, respectively, and ε is a random error. The errors

are assumed to be normal, having a mean of zero, an unknown constant variance σ^2 (homoscedasticity), and finally they have to be independent (uncorrelated) (Montgomery, Peck, & Vining, 2006).

The residual plots are an efficient approach to investigate whether the fit of the regression model is strong or weak. These residuals have to be a random shape with no pattern to make sure the residuals are not correlated (Tsai, Cai, & Wu, 1998).

2.2.2 Multiple Regression Model

When a regression model includes more than one predictor variable, it is called a multiple regression model, and it is represented by $y_t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon$ (Montgomery, Peck, & Vining, 2006). Multiple regression is basically the same as simple regression which is used to determine the goodness of the fit of the data and reliability of the model. The numbers of the independent variables are controlled by the research indicators in its hypothesis. The same rules in the simple regression model should be applied on the residuals, which resulted from the multiple regression model.

2.2.2.1 *Multicollinearity*

When there is no linear relationship in existence among regressors, they are orthogonal (statistically independent), which should occur in any multiple regression model. However, if this does not occur, a serious problem called multicollinearity has taken place, which is a big issue that has influenced the efficiency of the regression model (Montgomery, Peck, & Vining, 2006).

One of the most important ways to discover whether there is a multicollinearity in the model or not is to build a simple regression model between each pair of the

independent variables and observe R^2 (R^2 will be discussed in Section 2.4.2). If it is close to 1, this means there is a strong multicollinearity.

There are many ways to correct the model after discovering the multicollinearity in it. First, remove one of the strong correlated variables, but this can result in a non-significant model. Secondly, increase the sample size of the model, but unfortunately, it is hard to recollect data in some research. Thirdly, combine one or more of the strongest correlated variables into a single variable without changing the idea of the research hypothesis. Finally, leave the model as is and describe the entire variable's significance or weakness, and investigate the problem and purpose of the solutions (Fennessey & D'Amico, 1980).

2.2.3 Mixed Effect Models

In the past few years, many researchers and practitioners have taken mixed model courses for the important role it plays in many fields, as it is an extension of the multiple regression model and its significance expanded too. As mentioned earlier (in Chapter 1), the mixed effect models (or called mixed models) are statistical models which consist of fixed and random effects, and they can be linear or nonlinear. The fixed effect variables are predefined by the experimenter and are the same for all groups in one study. Additionally, random effect variables are drawn from larger distribution and the experimenter cannot have control over it.

Mixed models investigate grouped or repeated measures data in which the response is measured at fixed time points. These repeated measures data are correlated and data differs from the time series data by multiple subjects which are included; the number of measurements per subject is generally not very large (Moser, 2004).

The general form of the mixed models is $Y = \beta X + \alpha Z + \varepsilon$, where βX is the fixed effect part and αZ is the random effect. The detailed equation explains two important classifications about the mixed models, which are within groups (observations or treatments) or between groups (levels) effects. Since there can be several treatments or single treatments evaluated at different points of time, the mixed model equation is in the form of: $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \alpha_{0i} + \alpha_{i1} z_{1ij} + \cdots + \alpha_{iq} z_{qij} + \varepsilon$, where β and x represent the fixed effect part and α and z represent the random part of the mixed effect model, i represents the number of the group or the level, and j represents the number of observations (treatment) in each group (Draper & Smith, 1998).

The most significant improvement of mixed-effects models, is permitting the experimenter to simultaneously take into account all factors that likely participate in the understanding of the data structure, while having the ability to fit models of large unbalanced data (Baayen, Davidson, & Bates, 2008). The preceding talk about the mixed models will be used in Chapter 5.

2.3 Time Series Models

"Time series analysis is an ordered sequence of observation" (Wei, 2006). The order in the sequence is usually from beginning to end depending on time. This recorded sequence is always in the shape of electrical voltage. **Figure 2-1** is an example of a general time series which is in a shape of electrical signals.

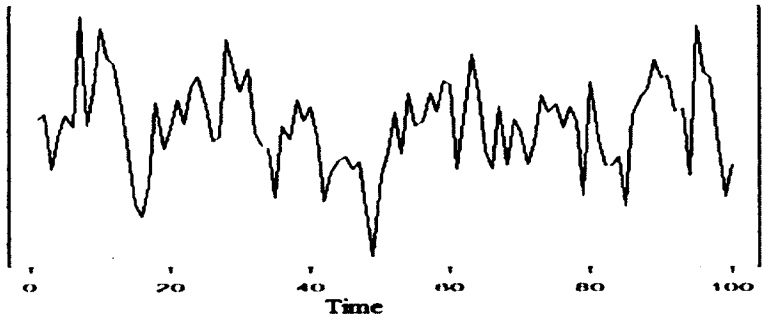


Figure 2-1: An example of a time series.

Time series analysis is used for modeling the data, recognizing the pattern in the correlated data, and predicting the future based on past and present data.

Time series is denoted as Z_t, where {t = 0, ±1, ±2,... } with mean $\mu_t = E(Z_t)$, variance $\sigma_t^2 = (Z_t - \mu_t)^2$, and the covariance function between Z_{t_1} and Z_{t_2} will be $\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2})$. The autocorrelation function of any time series measures how a series correlates with itself at different lags, and it is always one at lag zero. Consequently, the correlation between Z_{t_1} and Z_{t_2} time series will be $\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}$. If the autocorrelation between Z_{t_1} and Z_{t+k}, where k is a times lag, the equation will be $\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma(t, t+k)}{\sqrt{\sigma_t^2}\sqrt{\sigma_{t+k}^2}}$ (Wei, 2006). In the same manner, the partial auto correlation function is the partial correlation coefficients between the series and lags, but only after removing the best linear estimate of the series. Therefore, the partial autocorrelation between Z_t and Z_{t+k} time series will be:

$$P_k = \frac{Cov[(Z_t - \hat{Z}_t),(Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{Var(Z_t - \hat{Z}_t)}\sqrt{Var(Z_{t+k} - \hat{Z}_{t+k})}}$$, where $\hat{Z}_t$ and $\hat{Z}_{t+k}$ are the best linear estimate of Z_t and Z_{t+k}, respectively (Baayen, Davidson, & Bates, 2008).

Time series can be one of four cases:

1- The sequence varies in a fixed level, and in this case, it is called stationary in the mean.

2- The sequence in this case varies not in a fixed level, but in an upward (increasing) trend, and in this case, the variance is increasing, and it is nonstationary in the mean and the variance.

3- This case is the opposite of the second case; it is a downward (decreasing) trend with a decreasing variance, and it is also nonstationary in the mean and the variance.

4- The sequence has a frequent pattern as a consequent of seasonal variation, and it is also a nonstationary time series.

In cases 2, 3, and 4, transformation has to be complete to convert them to the stationary time series for removing the tendency, and differencing is one of the best ways to make the transformation. Nonstationary time series can happen in the mean, the variance, or both. Moreover, ARIMA is the famous homogenous nonstationary time series models (Wei, 2006).

2.3.1 Autoregressive Time Series Models (AR)

There is an important series that has to be mentioned before addressing AR, which is called a white noise (WN) process. White noise is a sequence of uncorrelated random variables from a fixed distribution. It is denoted as a_t with $E(a_t) = \mu_a = 0$,

constant variance $Var(a_t) = \sigma_a^2$, and covariance $\gamma_k = Cov(a_t, a_{t+k}) = 0$, for all $k \neq 0$ (Wei, 2006).

The autoregressive model is a representation to express a stationary time series process by regressing the dependent variable at time t on its own past values, and adding the white noise to it. The autoregressive model of order p is denoted as AR (P), and in the form of $\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-p} + a_t$ where $\dot{Z}_t = Z_t - \mu$ (Wei, 2006). Previous discussions about the AR model and white noise will be used in Chapters 3, 4, and 5.

2.3.2 Autoregressive Integrated Moving Average Time Series Models (ARIMA)

ARIMA is a homogenous nonstationary time series which can be transformed to the stationary time series by having a suitable degree of differencing. The general ARIMA model of (p,q) order and after d differencing degree, is in the form of $\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t$ (Brockwell & Davis, 1996). All previous discussions about the time series and the ARIMA model will be used in Chapter 3.

2.4 Model Selection Criteria

The best model of any research is the singular model for making implications from the data. The other unselected set of models which are clearly not significant will be maintained for more investigation. At the model selection stage there are data and a set of candidate models with statistic derivations which select the best model that will be based upon these derivations (Burnham & Anderson, 2004). Many researchers consider the robust inference test to be the best method for comparing models (Platt, 1964).There is no exclusive statistical procedure or method for selecting the best model, but there are

various methods that have been proposed, and each one of them has its own properties (Draper & Smith, 1998).

Even though the model selection criteria provides the same conclusion most of the time, they do not guide to the same solution when applied on the same data set. In any research, the performance measurement, of the model is considered by observing some statistical numbers, such as the minimum number of errors or coefficient of determination, to see how well data points fit a statistical model, and based upon those performance measurements, the best model will be selected. In this research, we will discuss two of the most significant model selection criteria, which are Akaike Information Criteria (AIC) and Coefficient of Determination (R^2).

2.4.1 Akaike Information Criteria (AIC)

Akaike information criterion was named according to a Japanese statistician who formulated it in the early 1970s (Akaike, 1974). As of late, AIC has had a significant impact in statistical model evaluations and has been known as one of the most important selection models for choosing the most favorable approximating model among all the candidates (Bozdogan, 2000). Furthermore, the best model has the smallest expected inconsistency among the candidates. In other words, the model with the lowest AIC value is selected as the best model to fit the data (Bozdogan, 2000; Wang, 2000).

AIC is a measure of the goodness of the fit and the complexity of the model, and is calculated as the following formula: $AIC = -2\log L(\hat{\theta}) + 2k$, where $L(\hat{\theta})$ is the maximized likelihood function, $-2\log L(\hat{\theta})$ is the lack of fit component, and k is the number of estimated parameters in the model (Bozdogan, 2000).

2.4.2 Coefficient of Determination (R^2)

Coefficient of determination (R^2) is an extremely important indicator of how the variability in y is explained by the regression model. In essence, R^2 demonstrates how well the data points fit a statistical model. It can be larger by adding more variables to the model and the magnitude of it depends on the range of the variability in the regressor variable (Montgomery, Peck, & Vining 2006). R^2 is represented by:

$$R_{tot}^2 = \frac{SS_E}{SS_T} \qquad \textbf{Equation 2-1}$$

where SS_T is the sum of squares for the response variable and SS_E is the final error sum of squares. The model with highest R^2 value is selected as the best model to fit the data.

CHAPTER 3

IMPACT OF COMPUTATIONAL ALGORITHMS ON MODELING OZONE

3.1 Introduction

Numerical algorithms are fundamental tools used to solve problems in computational science and are applied to known models or to newly developed ones. Although any given model can potentially employ more than one numerical algorithm, the selection of the algorithm stays with the researcher, which will use the best algorithm that will provide the solution. However, not all the numerical algorithms provide for the same problem, completely similar results. One algorithm's result can be more precise than another's.

When performing statistical analysis, the common practice is to use built-in procedures available in any statistical software package with default numerical algorithm embedded in that procedure. Furthermore, practitioners possess the ability to look for the other available algorithms besides the default, even though the results are sensitive to the algorithm used. Consequently, studying the impact of different numerical algorithms under the same procedure on solving the same problem is the scope of our paper, since they can provide significantly different results.

Our goal consists not in comparing two or more models of one dataset, but in comparing the numerical algorithms used to solve models of the dataset. After presenting the results, we will follow with the analysis of the behavior and performance of the models and the description of the conclusions based upon verification and results. The idea of this study is to make an assessment of different algorithms used to solve two time series procedures on a time series data. Observations are repeatedly recorded in a period of time, which is called the time series data; modeling this type of data is a significant topic in many scientific fields (Meek, Chickering, & Heckerman, 2002). The most complex implementation of time series models are applied in physical and environmental sciences (Shumway & Stoffer, 2011).

Delivering the ideology of this study requires a time series data, which is an intervention model for Ozone data chosen since it is considered one of the most important applications in economic and environmental science. This data set represents the Oxidant pollution in downtown Los Angeles from 1955 to 1972 and was studied by Box and Tiao (1975). Ozone is a chemical substance and it is referred to as O_3, which is also called oxidant, and a high concentration of Ozone is known as Oxidant pollution. Box and Tiao completed their work by using the stochastic model for the noise with the ARIMA process. The Maximum Likelihood (ML) algorithm is the numerical algorithm elected to solve the model. Two other algorithms are selected to solve ARIMA, namely Conditional Least square (CLS) and Unconditional Least Square (ULS) algorithms. The comparison is based on the results of the three algorithms selected. The second phase of the study mimics the ARIMA procedure by using the autoregressive process and comparing the outcomes of its four numerical algorithms: Yule Walker (YW), Iterative Yule Walker

(ITYW), Maximum Likelihood (ML), and Unconditional Least square (ULS) algorithms. Highlight discuss

We underline the sensitivity of the data to the chosen algorithms by applying a set of key input variables and parameter values. To test how the data reacts to different algorithms, major factors were chosen to be the guideline in building decisions. The analysis is based on the noise (residuals) of the data, which is a very important benchmark in testing the model's performance and assists in learning the data efficiently, while making the results easier to be interpreted and compared.

3.2 Methods

3.2.1 Data Description

As mentioned earlier, Ozone data will be solved as a time series regression model, with Ozone as a response variable and three regressors:

- x_1 which is an intervention variable with zero value for periods prior to 1960 and 1 for post 1960;
- *Winter* is 1 from November to May starting with the beginning of 1966 and zero otherwise; a
- *Summer* which is 1 from June to October starting from the beginning of 1966 and zero otherwise.

The year 1960 was chosen to be the intervention year to determine the values of x_1, because an important event occurred in the early 1960s which impacted the Ozone level. This event consisted of the opening of the Golden State Freeway, creating a new law, resulting in the reduction of the fraction of hydrocarbons in the gasoline sold. The law was produced to reduce the allowable proportion of hydrocarbons in the gasoline,

affecting the Ozone level. *Summer* and *Winter* are also considered as intervention variables because of the different temperatures and sunlight intensity that has different effects on oxidant pollution. Additionally, their effects have been an intervention to the Ozone level since 1966, which was the starting year of inventing new cars with engines that produced lower O_3 than before. **Table 3-1** is summarizing the Ozone variable and the three independent variables.

Table 3-1: Summary of Standard Statistical characteristics for Ozone data.

Ozone dataset N=228				
Variable	**Mean**	**SD**	**Min**	**Max**
Ozone	3.77268519	1.49163379	1.2	8.7
X1	0.7368421	0.441316	0	1
Summer	0.1754386	0.381179	0	1
Winter	0.245614	0.431398	0	1

Figure 3-1 shows the graph for the Ozone level from 1955 to 1973. The graph shows a downward trend with a peak Ozone level of 8.7, which was observed in September of 1966, and the lowest was 1.3 in January and December of 1975, and December of 1977.

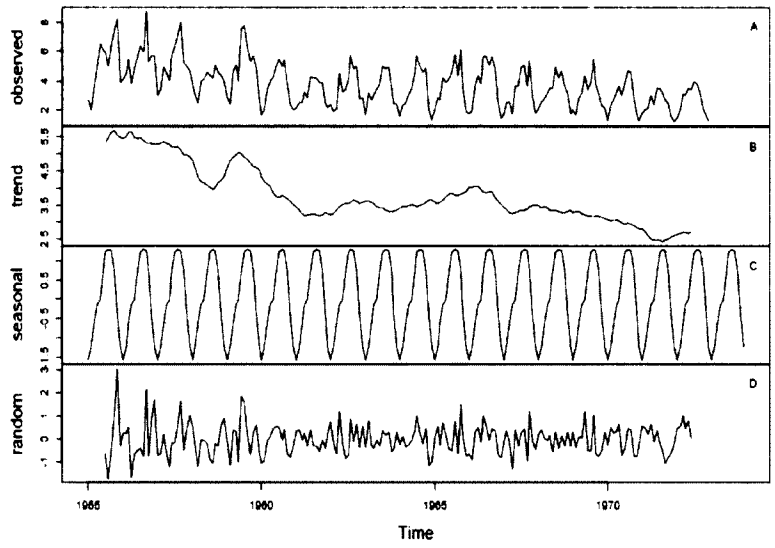


Figure 3-1: (A) Ozone level in downtown Los Angeles from Januar 1955 to December 1972, (B) the estimated trend component,(C) the estimated seasonal component, and (D) the estimated irregular component.

One should notice that the variance is not constant through time and since the Ozone data has seasonal components, it has to be adjusted by subtracting the predictable seasonal part from the series. As it was done in a previous study by Box and Tiao, the Ozone series was differenced at lag 12 because it is a monthly data. **Figure 3-2** shows the Ozone level series in the time interval 1955 to 1973 after it was differenced, and it is obvious that the seasonal component (downward trend) disappeared.

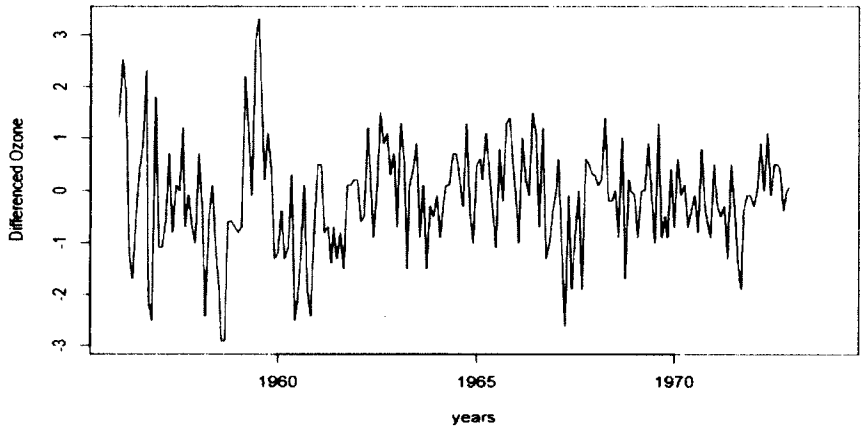


Figure 3-2: Differenced Ozone series from January 1955 to December 1972.

3.2.2 Models and Algorithms Description

To reach our goal of investigating and assessing the impact of different algorithms on Ozone data and to select the best algorithm, a set of runs were executed on the entire dataset. The appropriate numerical algorithms of investigation are used to solve two main time series processes which are the ARIMA and Autoreg procedures. Additionally, regression is the basic model the investigation was built on, and explaining the regression model within a time series platform means that the dependent variable is influenced by some time series independent variables. Furthermore, dependency should be expected between residuals of the time series values and the researcher's work to estimate the correlation between them (Shumway & Stoffer, 2011).

The first phase in this study mimicked the work of Box and Tiao by using the ARIMA procedure in SAS. Two separate runs will be executed, in addition to the one completed using the ML algorithm and these two runs will be based on two different algorithms: CLS and ULS. Akaike Information Criteria (AIC) is the selection criteria method for deciding on the best algorithm. Above all, Box and Tiao (1975) has more details about the ARIMA model, in which they explain how to fit this model on the Ozone data.

The second phase of the study consisted on applying a regression model on Ozone data (216 observations) by using the autoregressive procedure in SAS as a means of obtaining regression residuals, upon which the analyses is built. The autoregressive model explains the current value of the series based on the function of number p past values, which is the reason for representing the autoregressive model by AR (p).

In the third phase, an autoregressive model is evaluated, given that it is the most familiar approach used in describing the time series data (Brockwell & Davis, 1996). This autoregressive process was applied on achieved errors from the second phase. Consequently, this obtained error is solved by four different numerical algorithms as was mentioned. Consecutively, four algorithms were applied to the differenced errors. The last step compared the results of each algorithm separately, based upon Akaike Information Criterion (AIC) and Coefficient of Determination (R^2), which are the two reference methods for selecting the best and the worst models. Choosing two selection criteria methods to compare the algorithm's performance is called the multivariate pairwise tests, and in such a case, two vectors are used for comparisons (Yildiz, Aslan, & Alpaydin, 2011).

3.2.2.1 ARIMA Algorithms Evaluations (Phase I)

Box and Tiao (1975) developed their model based on applying the ARIMA process on residuals of the following regression model:

$$ozone = \beta_0 + \beta_1 x1 + \beta_2 \text{Summer} + \beta_3 \text{Winter} + \varepsilon_t \qquad \textbf{Equation 3-1}$$

Where *Ozone* is the dependent variable, x_1, *Summer*, and *Winter* are independent variables; $\beta_0, \beta_1, \beta_2,$ and β_3 are the model parameters, and ε_t is the error term. Box and Tiao's process involves differencing Ozone and innovation x_1 seasonally while applying ARIMA on the error term. Applying ARIMA on the residuals of regression equation was completed without the autoregressive process. In addition, specifying the parameters at lag1 and lag 12 represent the moving-average part of the model due to these two lags having significant correlation. After going through the required steps, the final intervention model for the Ozone data is as follows:

$$\nabla_{12}\nabla ozone = \beta_1\nabla_{12}\nabla x1 + \beta_2 summer + \beta_3 w\,\text{int}\,er + \frac{(1-\theta_1 B)(1-\theta_2 B^{12})}{(1-B^{12})}a_t \qquad \textbf{Equation 3-2}$$

where $\nabla_{12}\nabla ozone$ and $\nabla_{12}\nabla x1$ are seasonally differenced *Ozone* and x_1, respectively, β_1, β_2, and β_3 are the model parameters, and a_t is the white noise.

Assessing the sensitivity of the Ozone data is achieved by comparing the three estimation methods, which are implemented for the ARIMA process of the residuals. The ARIMA procedure in SAS is carried out through three different numerical algorithms which are defined as the following: Conditional Least Square (CLS) is the default method, Maximum Likelihood (ML), and Unconditional Least Square (ULS).

The *Conditional Least Square Algorithm (CLS)* estimation is conditional on the assumption that past unobserved errors are equal to 0 and it produces estimates by minimizing the following equation (SAS, Inc., Cary, NC):

$$\sum_{t=1}^{n} a_t^2 = \sum_{t=1}^{n}(x_t - \sum_{i=1}^{\infty}\pi_i x_{t-i})^2 \qquad \textbf{Equation 3-3}$$

where π is computed from the estimates of ϕ and θ at each iteration, being AR and MA parameters for the series $x_1,\ldots,x_{t-1},x_t$ and $\sum_{t=1}^{n} a_t^2$ is $e'.e$ where e is the residuals matrix.

The *unconditional least squares algorithm (ULS)* produces estimates and also points to the exact least squares (ELS) method (SAS, Inc., 2012). The initial estimates are computed by using CLS algorithm, and ULS estimation obtained by minimizing the residual sum of square, as is represented in the following equation:

$$\sum_{t=1}^{n} a_t^2 = \sum_{t=1}^{n}(x_t - C_t V_t^{-1}(x_1,\cdots,x_{t-1})')^2 \qquad \textbf{Equation 3-4}$$

where C_t is the covariance matrix and V_t is the variance matrix for the series $x_1,...,x_{t-1},x_t$. $\sum_{t=1}^{n} a_t^2$ is $e'.e$ where e is the residuals matrix.

The *Maximum Likelihood (ML)* algorithm maximizes the likelihood function through nonlinear least square Marquardt's method. The initial estimates are computed using the CLS and ML algorithms and the last equation of MLE for a regression model is produced by minimizing the following sum of squares (SAS, Inc., 2012):

$$|H|^{1/n} e'e|H|^{1/n} \qquad \textbf{Equation 3-5}$$

where e is a vector of the residuals and H is the lower triangular matrix with positive elements on the diagonal, such that $HH' = \Omega$ where Ω is the determinant of the regression equation.

To assess the algorithms in this phase, Akaike Information Criteria (AIC) was used. AIC is a measure of the goodness of the fit of the model as well as the complexity of that model and is calculated as the following formula:

$$AIC = -2\ln(L) + 2k \qquad \textbf{Equation 3-6}$$

where L is the maximized value of the likelihood function for the estimated model and K is the number of estimated parameters.

3.2.2.2 Autoregressive Model (Phase II)

The autoregressive procedure in SAS is used to estimate linear regression models of a time series data, and the formula of the regression in this case is:

$$Y_t = X_t'\beta + V_t \qquad \textbf{Equation 3-7}$$

where Y_t is the response variable, X_t' is the regressor variable with β slope, and V_t is the error term. As was mentioned, the errors are correlated and represented by **Equation 3-8:**

$$V_t = -\varphi_1 V_{t-1} - \cdots - \varphi_m V_{t-m} + \varepsilon_t \qquad \textbf{Equation 3-8}$$

The regression model used in autoregressive procedure estimated the Ozone as a means of studying the relation between differenced Ozone as a dependent variable, with differenced intervention, *Summer*, and *Winter* as independent variables. The regression model is represented by the following equation:

$$\nabla_{12}\nabla ozone = \beta_1 \nabla_{12}\nabla x1 + \beta_2 summer + \beta_3 w\,\text{int}\,er + \varepsilon_t \qquad \textbf{Equation 3-9}$$

where $\nabla_{12}\nabla ozone$ and $\nabla_{12}\nabla x1$ are seasonally differenced at lag 12 Ozone and x1, respectively, $\beta_1, \beta_2,$ and β_3 are the model parameters, and ε_t is the residual. Ozone and intervention variables were differenced to mimic Box and Tiao's work in the ARIMA procedure.

The statistical analysis was performed using SAS 9.3 (SAS, Inc., 2012) version, and each of the model parameters estimators were evaluated, besides the mean square error (MSE) and coefficient of determination (R^2) (Kobayashi & Salam, 2000).

Mean Square Error: $$MSE = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n-p} \qquad \textbf{Equation 3-10}$$

Coefficient of determination: $$R^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \qquad \textbf{Equation 3-11}$$

where $Y_i, \hat{Y}_i, \bar{Y}$ are the actual, the predicted, and the average values for the dependent variable, respectively. Additionally, n is the total number of observations and p is the number of parameters in each model.

By the time the statistical analysis had been executed for the regression model, the residuals were prepared for the third phase. Since the Ozone data was already differenced

before fitting the autoregressive procedure is a time series data, the obtained errors are not serially correlated over time; the produced residual of the regression model was used in autoregressive procedure with four different algorithms in phase III.

3.2.2.3 Autoregressive Algorithms Evaluation (Phase III)

Evaluation of the algorithm's performance is conducted by applying autoregressive procedure in SAS. As it was mentioned earlier, the autoregressive procedure is used to estimate the linear regression model of a time series data. After considering the error of the differenced Ozone, which resulted from the previous stage, and since these residuals are a consequence of uncorrelated random variables, the outcome residuals are called the white noise process (a_t). This white noise has properties of zero mean, constant variance, and zero covariance.

As suggested by the Cleveland (1972) method for analyzing the autocorrelation, inverse auto correlation and partial auto correlation functions were applied to determine the autoregressive model order. As such, twelve was the best lag order for the given data. By stating twelve to be the order of the autoregressive error process for fitting AR (12), the error's autoregressive equation is in the form of **Equation 3-12** and the final equation was applied in autoreg procedure in SAS in the form of **Equation 3-13**:

$$(1-\phi_{12}B^{12})Z_t = a_t \qquad \textbf{Equation 3-12}$$

$$(1-\phi_{12}B^{12})\cdot residual = residual \qquad \textbf{Equation 3-13}$$

where a_t is the white noise.

At this stage of the research, the focus is on the residuals and how critical they are to the four different algorithms embedded in the autoregressive procedure, which helps in understanding the sensitivity of the Ozone to the intervention (x_1), *Summer*, and *Winter*

variables. Furthermore, comparing the data is achieved through applying four estimation methods, which are implemented for the autoregressive approach of residuals model, and algorithms are defined as the following: Yule Walker (YW), Iterative Yule Walker (ITYW), Maximum Likelihood (ML), and Unconditional Least Square (ULS).

The *Yule Walker Algorithm (YW)* algorithm is the default option in the autoregressive procedure in SAS. In a large sample from an AR (p) process, YW algorithm is represented by **Equation 3-14**:

$$C_t = \sum_{1}^{p} \phi_p C_{t-p} + \sigma_\varepsilon^2 \delta_{t,0} \qquad \textbf{Equation 3-14}$$

where C_t is the auto covariance function of the series, ϕ is the vector of autoregressive parameters which is the estimation of β, p is the autoregressive lag order, σ_ε^2 is the standard deviation of the residual, and $\delta_{t,0}$ is the Kronecker delta function.

The *Iterative Yule Walker algorithm* uses the resulted residuals which stems from the YW algorithm to create new estimators of ϕ and V (the error vector).

The *Maximum Likelihood algorithm* is efficient but the downside is that it needs a good starting point and cannot be computed for some data. Therefore, it is maximized by minimizing the objective function:

$$|L|^{1/N} e'e|L|^{1/N} \qquad \textbf{Equation 3-15}$$

where e is the residual vector and L is the likelihood function which represented as:

$$-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|V|) - \frac{S}{2\sigma^2} \qquad \textbf{Equation 3-16}$$

where σ^2 is the variance, V is the variance matrix of the error vector, $|V|$ is the determinant of V, N the number of the observations, and $S = e'.e$ (unconditional sum of squares of the model) where e is the transformed error.

The last algorithm that can be used to solve the autoregressive equation is the *Unconditional Least Squares Algorithm (ULS)*, which is computed by minimizing S with respect to the parameters β and ϕ_i (SAS, Inc., 2012). The best algorithm is decided based on the selection criteria. In addition, each model's intercept is estimated, as well as the residuals and the 12^{th} order of autoregressive process.

The strong implication test utilized by many researchers is the best method for contrasting models (Platt, 1964). Moreover, the supreme model of any research is the most favorable model for developing inferences from the data. In any research, the execution of any model or algorithm is measured by choosing extreme statistics, either the minimum number of errors (e.g., lowest value MSE or AIC), or the maximum number of coefficient of determination (R^2) (Dayton, 2003; Cameron & Windmeijer, 1995). This study uses both the lowest AIC and the highest coefficient of determination R^2, as it is mostly used in the time series and regression model analyses.

3.3 Results

Models are generally sensitive to the input variables and parameter values because their reactions differ when applying the numerical methods. Hence, to test the model's responses, three numerical algorithms with the ARIMA process and four with

the autoregressive process were applied to the Ozone model. Ozone regression models with ARIMA algorithm results are recorded in **Table 3-2** with n = 216. All variables were significant with $p < 0.0001$, except the Winter variable.

Table 3-2: Parameter estimates and fit statistics of the Ozone model with three different algorithms of the ARIMA process.

	Ozone lag (1)	Ozone lag (12)	X1 lag (12)	Summer	Winter	SE	AIC
CLS	-0.29983	0.59234	-1.2624	-0.2615	-0.082	0.856	520.52
ULS	-0.25264	1	-1.3742	-0.2027	-0.084	0.768	510.78
ML	-0.26684	0.76665	-1.3306	-0.2394	-0.080	0.797	501.77

From **Table 3-2** one can notice that all parameters have negative effect on the model, except the error when the ARIMA process is applied with MA (12), which has a positive effect. The highest AIC and SE values were discovered when CLS algorithm was applied, which explains why even though the CLS is the default option, it has the worst results among the three algorithms. Based upon the previous conclusion, a very important point has to be declared, which the hypothesis is claiming the default option algorithm under any statistical software procedure is not always the perfect choice for a predefined model. The best result was achieved according to AIC selection criteria by using the ML algorithm which explains why Box and Tiao depended on ML algorithm with the ARIMA process to perform their work, although it is not the default option, and it causes more work and time in order to be executed. **Equation 3-2** will be written as the following based on different algorithms from the results in **Table 3-2**.

CLS:

$$\nabla_{12}\nabla ozone = -1.2624\nabla_{12}\nabla x1 - 0.2615\,summer - 0.082\,w\,\text{int}\,er + \frac{(1+0.29983B)(1-0.59234B^{12})}{(1-B^{12})}a_t$$

ULS:

$$\nabla_{12}\nabla ozone = -1.3742\nabla_{12}\nabla x1 - 0.2027\,summer - 0.0842\,w\,\text{int}\,er + \frac{(1+0.25264B)(1-B^{12})}{(1-B^{12})}a_t \text{ and}$$

could be written as

$$\nabla_{12}\nabla ozone = -1.3742\nabla_{12}\nabla x1 - 0.2027\,summer - 0.0842\,w\,\text{int}\,er + (1+0.25264B)a_t$$

ML:

$$\nabla_{12}\nabla ozone = -1.3306\nabla_{12}\nabla x1 - 0.2394\,summer - 0.0802\,w\,\text{int}\,er + \frac{(1+0.26684B)(1-0.76665B^{12})}{(1-B^{12})}a_t$$

For all the previous models, MA (1) = -0.2924 and MA (12) = 0.4074 with white noise (a_t) = 0.945. Since 1973 had missing values throughout the entire year, the year was forecasted with 95% confident interval; the results were estimated by using the three algorithms as shown in **Table 3-3**.

Table 3-3: Forecasting 1973 by using three different algorithms.

1973	CLS		ULS		ML	
Months	**Forecast**	**SE**	**Forecast**	**SE**	**Forecast**	**SE**
January	1.4538	0.8562	1.2738	0.7677	1.4205	0.7966
February	1.8783	0.8938	1.5871	0.7919	1.8446	0.8244
March	2.5705	0.8938	2.2482	0.7919	2.4567	0.8244
April	2.8788	0.8938	2.7649	0.7919	2.8590	0.8244
May	3.1553	0.8938	2.9482	0.7919	3.1501	0.8244
June	2.8074	0.8938	2.7952	0.7919	2.7211	0.8244
July	3.3355	0.8938	3.3341	0.7919	3.3147	0.8244
August	3.4400	0.8938	3.3507	0.7919	3.4787	0.8244
September	2.8253	0.8938	3.2952	0.7919	2.0405	0.8244
October	2.0741	0.8938	2.7507	0.7919	2.3587	0.8244
November	1.6465	0.8938	2.3982	0.7919	1.8588	0.8244
December	1.2149	0.8938	1.5663	0.7919	1.2898	0.8244

From **Table 3-3** we notice that the squared errors for the forecasted months with the CLS algorithm were the highest among the three algorithms, which is additional verification that the default option is not always the perfect choice to execute the procedure.

Following the investigations on behalf of the ARIMA model analysis, the second phase was initialized by using the identical original Ozone data to start the autoregressive model assessment. As shown above, Equation 3.9 was applied and the results are reflected in **Table 3-4**.

Table 3-4: Parameter estimates and fit statistics of the Ozone model by the autoregressive process.

Variable	Estimate	SE	P	MSE	AIC	R^2
Intercept	-0.0176	0.0983	0.8582	1.04404	591.679393	0.0710
X1 diff	-1.1991	0.3109	0.0002			
Summer	-0.2281	0.1987	0.2524			
Winter	-0.0885	0.176	0.6155			

According to the outcomes, the differenced intervention variable is the only significant variable in the model and all variables had negative effects on the Ozone level. The coefficient of determination is very small, which is explained by the mentioned insignificant variables in the model. This insignificance could have been due to all three variables being dummy variables. **Figure 3-3** shows the forecasting series of the Ozone level in 1973 by using three algorithms: CLS, ULS, and ML.

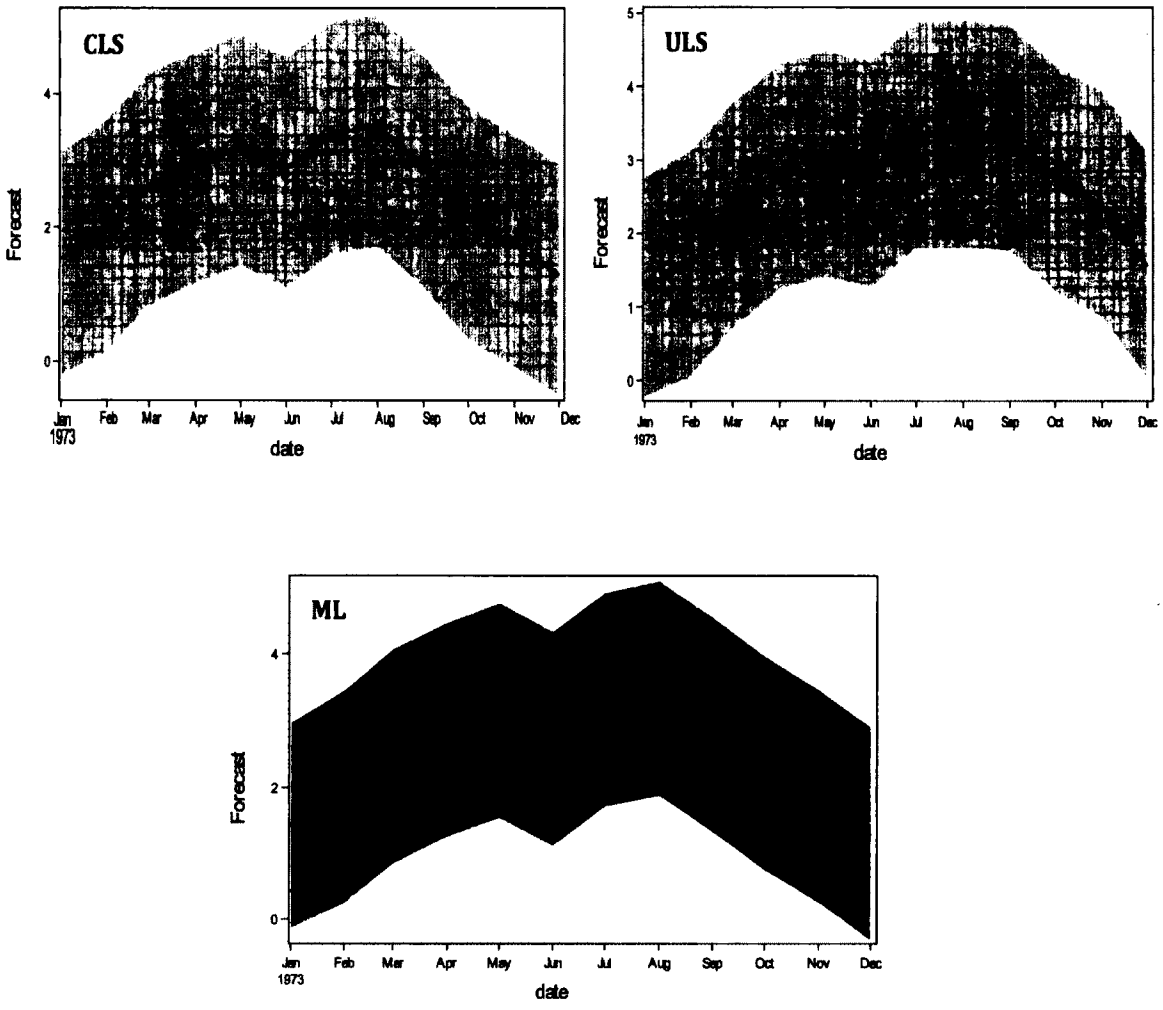


Figure 3-3: 1973 forecasting by three different algorithms.

By checking the time effects on the residuals of the Ozone model, **Figure 3-4** illustrates a decreasing funnel shape which means there is a correlation between them. For this reason, differencing was done to remove the correlation and prepare them for use in phase three of the study as stationary white noise residuals.

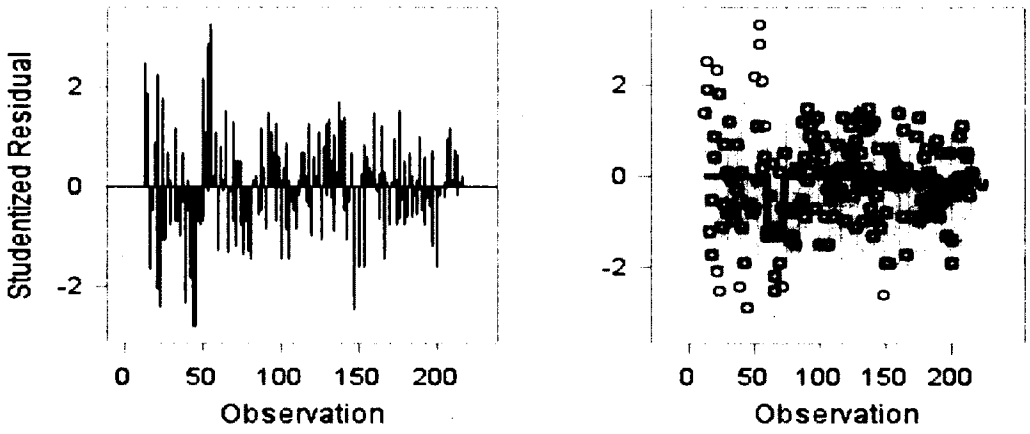


Figure 3-4: The residuals of the autoregressive model between the differenced Ozone with differenced intervention variable, Summer, and Winter.

After completing the analysis of phase two, the resulted residuals were differenced to remove the correlation among them and prepare them for the third phase of this study. As was mentioned in Subsection 3.2.2.3, the autoregressive model was applied on the linear differenced residuals model in the form of $residual_{1} = (1 - \phi_{12} B^{12}) \cdot residual$ to investigate the sensitivity of the Ozone data towards the four numerical algorithms. The autoregressive procedure was applied four times with four different algorithms on the differenced residual with lag 12 and the results are shown in **Table 3-5**.

Table 3-5: Autoregressive error model results for the four algorithms.

Algorithm	Residual	AR12 ε_{t-12}	MSE	AIC	R^2	Durbin Watson at lag 12		
						DW	Pr < DW	Pr > DW
ML	0.7085	-0.4951	0.76425	459.751036	0.3194	1.8784	0.4647	0.5353
ITYW	0.7164	-0.4231	0.77253	459.760418	0.3191	1.981	0.7413	0.2587
ULS	0.7044	-0.5334	0.76311	459.97695	0.3202	1.829	0.3287	0.6713
YW	0.7167	-0.4198	0.77311	462.232177	0.3067	1.9859	0.7526	0.2474

Assessing the errors' sensitivity to the algorithms and comparing performance and behavior with respect to every algorithm is the goal of this research. After that the results of the model parameter estimations were sorted ascending with respect to AIC and descending with respect to R^2, as shown in **Table 3-5**.

It is clear that the Ozone residuals had the least sensitive results with YW algorithm, and the highest impact was achieved when ML was applied. Additionally, one can notice that even though Yule Walker (YW) is the default option for the autoregressive model, it was not a good choice for the Ozone model when applied. This is an additional proof that the ML algorithm was the best choice for Box and Tiao for solving their model of the Ozone data. By using the Durbin Watson methodology (Watson, 1951) for testing serial correlation, research detected $DW > DW_U$ for all situations in **Table 3-5**, meaning we do not reject H_0 as there is no serial correlation (all the serial correlation $\rho_s = 0$). Preliminary results were expected since we already differenced residuals to remove the correlations among them. The most effective way of assessing any model's performance is through their residual which is accomplished through plotting the prediction errors for the Ozone model and their variables (Weiskittle, Hann, Kershaw, & Vanclay, 2011). In the regression analysis with time series data,

residual plots have been widely used to discover a model shortage (Anscombe, 1961, Chatterjee, & Hadi, 1980). Furthermore, graphical analysis of residuals is a very useful way to investigate the goodness of the fit for any regression model (Montgomery, Peck, & Vining, 2006). The differenced residuals are illustrated in **Figure 3-5,** clearly showing no pattern in the residuals' plot, as they are distributed randomly.

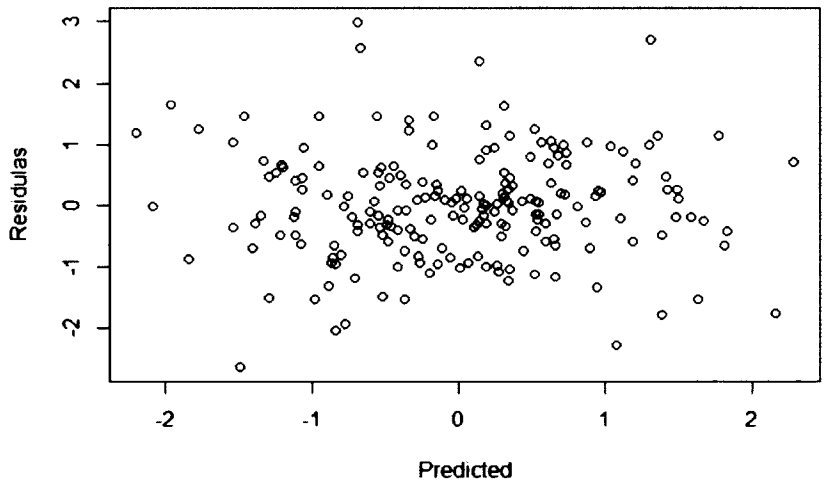


Figure 3-5: Times effect on differenced residuals of the Ozone model.

As was stated previously the residuals were modeled after differencing, **Figure 3-6** and **Figure 3-7** show the state of the residuals when the ML algorithm is applied. It is obvious from **Figure 3-6** that the residuals are independent and identical, normally distributed, and the white noise probabilities plot point out the resulting residuals (which are white noise). Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and Inverse Autocorrelation Function (IACF) are three important functions for any time series model. ACF is a bar chart used to test the correlation between time series coefficients and the time lag. Furthermore, if a trend is found indicating a correlation, dependency exists. Moreover, PACF is a bar chart of the amount of correlation between

variables and time lags, and IACF is the reverse of the calculated ACF of any time series. By investigating the partial autocorrelation function (PACF) plot in **Figure 3-7**, all partial autocorrelations are within the 5% significance error limit curves and there is no pattern observed, which means the AR (12) model was the suitable one to fit the data.

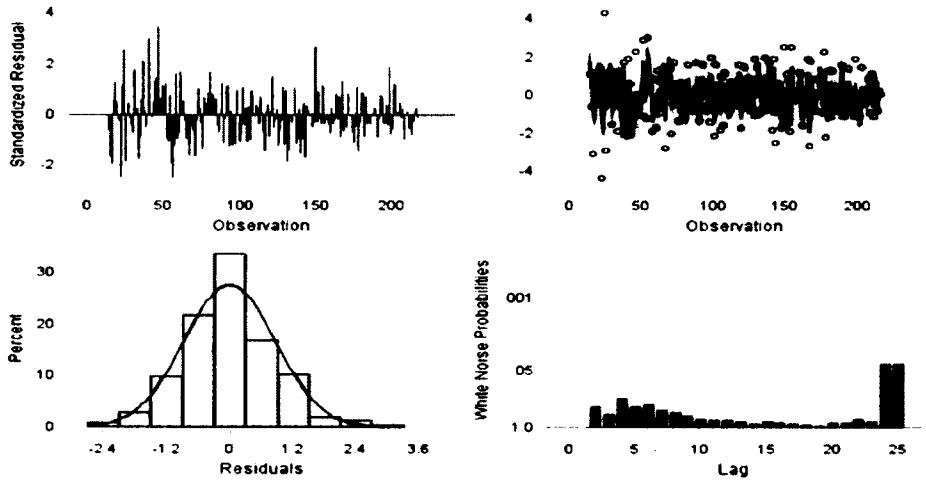


Figure 3-6: Complete information about plots the studentized residuals for the models with the nlag = 12, standardized, and histogram of the residuals and plots the white noise probabilities.

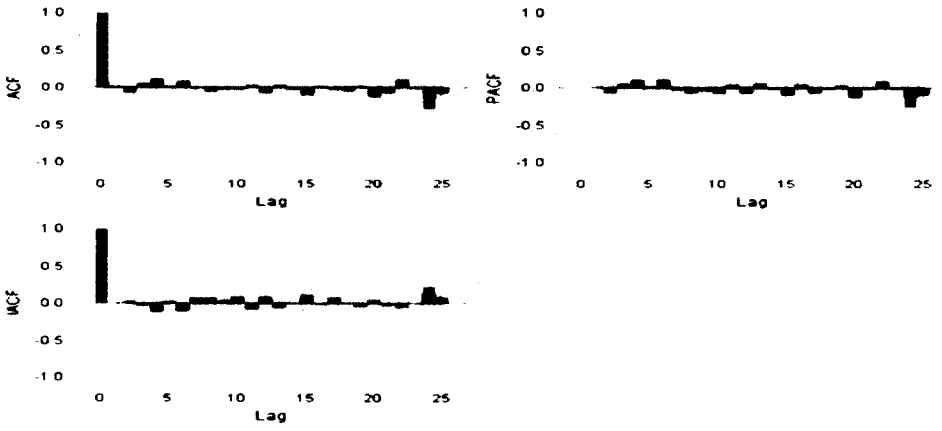


Figure 3-7: Autocorrelation, inverse autocorrelation, and partial autocorrelation function for the residuals.

3.4 Discussion

Ozone is a chemical substance called oxidant which is represented by O3, and a high concentration of it is known as the Ozone pollution. The first phases of this study was about mimicking Box and Tiao's work by using two different algorithms and comparing the results with their finding. This study was conducted by applying the ARIMA model with CLS and ULS algorithms. Following, algorithms were compared with the ML algorithm used in the study from 1975. In this research, differenced Ozone was predicted by differenced innovation, Summer, and Winter variables, and they referenced variables that had negative consequences on the Ozone level. Applying ARIMA on the residuals with lag 12 had an affirmative impact on the Ozone level.

Furthermore, The ML algorithm, with respect to the selection criteria, produced the best results when applied with the ARIMA model. In the second phase, the autoregressive process was used to fit the Ozone data, as done previously by using ARIMA. The resulted residuals from the second phase were differenced and refitted in the autoregressive process by using differenced residuals, which is the third phase. The third phase of our study had been repeated four times as a means of allowing the four autoregressive numerical algorithms (ITYW, ML, ULS, and YW) to be executed. After comparing the four algorithmic outcomes, the ML algorithm yielded the best result when used to solve the autoregressive model of the Ozone. Even though the Yule Walker (YW) algorithm is the default option when using the autoregressive procedure, it has the least significant result when used to solve the autoregressive model. Our accomplished study proves that Box and Tiao's choice of utilizing the ML algorithm is the best selection for working on the Ozone data.

3.5 Conclusion

The mixed models approach helps any researcher to consider all factors contribution to affecting the data. Ozone pollution data in downtown Los Angeles was very important in studying the oxidant pollution during the time of 1955-1972, and was investigated for the first time by Box and Tiao in 1975. The best model assessment is based upon the selection criterion measurement as the mean square error (MSE), AIC, or R^2. Furthermore, judgment of the best model or algorithm should be based upon more than one test or investigation; for instance, in this study AIC and R^2 were combined to select the best model.

More significantly, several numerical algorithms when associated with the Ozone model could behave better and have a higher impact than other algorithms associated with the same model when applied on the same dataset. In this study, we proved that the future is unknown but can be predicted as was done in forecasting 1973, which was unknown since we have the Ozone level up to December 1972. All analyses and results were obtained based upon the Ozone data, and how models and algorithms will perform when used with other data is uncertain to be the same. The basic idea of this study can be applied on any time series data. Moreover, for future work interested researchers can try any regression model to be fitted on the Ozone data by using the autoreg procedure in SAS as long as the model is linear. However, if the model is not linear, the NLIN procedure can be used instead but with stated initial parameters. Analyses which are based on the resulted residuals by differencing and mimicking them had been completed in phase three.

CHAPTER 4

IMPACT OF AUTOREGRESSIVE ALGORITHMS IN MODELING AUTO-CORRELATED DATA

4.1 Introduction

Computational methods and numerical algorithms have been a prolific topic in all scientific fields, which helps solving complicated models. Often one model can be solved by using more than one algorithm. To determine which algorithm solves a particular model best depends upon criteria selection. Therefore, important in modeling is not only the function selected to represent the processes or phenomenon of interest but the algorithms or methods used to solve them (Seppelt and Richter 2005).

In biometric investigations a central role is played by time series data, which record information on the same subject at various intervals, preferably evenly spaced. The importance of time series analyses has increased in significance throughout recent years, as a means of studying phenomena occurring at equal periods of time (Wei 2006). The focus of time series analysis consists in identification of relationships among various temporal connected variables (Bartlett 1978, Box *et al.* 1994, Gujarati 1995). Among biometric investigations growth and yield models are routinely investigated with time series techniques. Some of the most popular growth and yield models are focused on tree development, and describe changes of dendrometric attributes, such as height, volume or

diameter, with age. A vast palette of parametric models is available to relate total height of a tree with its age (Pretzsch 2009), which are either empirically or process based (Korzukhin *et al.* 1996).

Development of models describing growth or yield of a tree or stand recommends the inclusion in the modeling process besides mathematical representation of the relationships of interests the algorithms used to identify the respective relations. However, definite conclusions regarding a model and the algorithm employed to obtain the model are impossible when data supplied by experiments or observations are used, for at least two reasons. First, considering the simplifications associated with the modeling process and that the true model is not known in advance renders the inability to state the congruence between the found model and actual model. Secondly, measured data do not cover the entire spectrum of combinations required by a complex factorial experiment (Tabachnick and Fidell 2001). An alternative to recorded data is simulated data. Data simulation is the purpose of generating values from a well-known distribution that mimic real world phenomenon, and enables researchers to test hypotheses and forecast with various inputs and conditions. Simulation assists in various domains, such as decision making or risk assessment (Wang *et al.* 2008). Software that have the capability to solve complicate models are also able to generate data, such as SAS (SAS Institute 2010), R (Gentleman and Ihaka 2014) or SPSS (IBM Corporation 2014). Simulated data are popular in forestry research (Chuvieco *et al.* 2008, Strimbu 2014), and represents a fruitful avenue of investigating the impact of algorithms on model development. Therefore, the objective of this article is to assess the solutions supplied by different algorithms used to develop models with auto-correlated variables. The reason that

variables with an additional structure are used in this research is to ensure that the findings are not influenced by the violation of modeling assumptions, such as heteroskedasticity or co-linearity (Neter *et al.* 1996), as strict fulfillment of the analytical assumptions is required (Gujarati 1995).

4.2 Methods

The general approach of this research is based on the assumption that if the true model is known, then by using the appropriate modeling one should be able to reach a formula similar, if not identical, to the true model. Simulated data are based on known models, but to ensure that data resembles reality noise should be added. Generated data guarantees that all possible combinations of factors that can influence modeling assumptions are present in the analysis, particularly the type of noise used to simulate real variability. Conceptually the framework used to identify the impact of algorithms on analysis is based on the following serial steps **Figure 4-1**:

1. Generate data according to a deterministic functional relationship
2. Generate noise that mimic real variation
3. Add noise to the generated data
4. Using various algorithms, select models to fit the generated data, assuming that actual model used to generate data is unknown. The actual model will be included as a possibility. The selected models will fulfill all modeling requirements.
5. Compare the selected models with the models used to generate data.

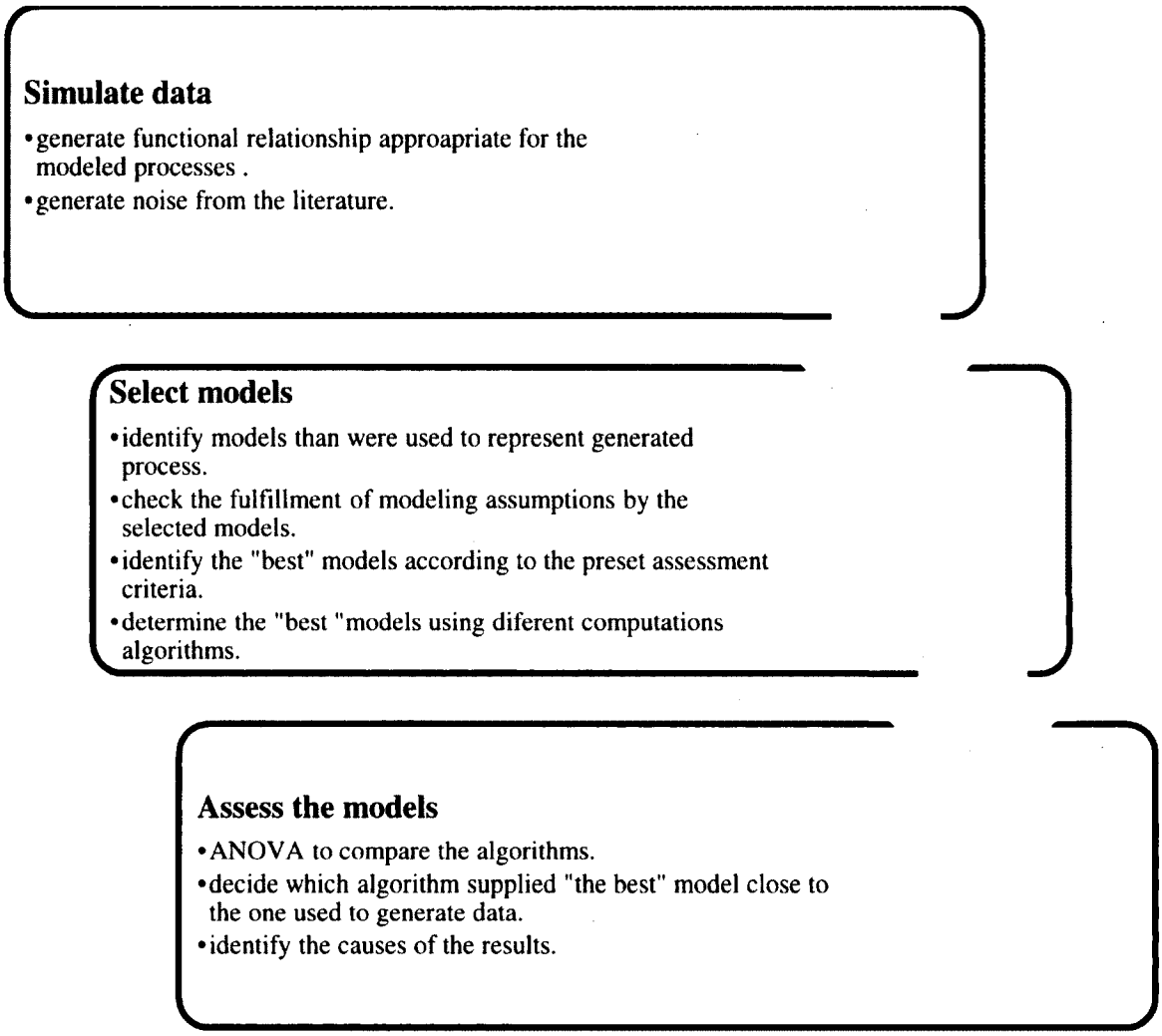


Figure 4-1: Framework for identifying the impact of the algorithms on model selection.

4.2.1 Data Description

The process selected for generating data was the tree height growth. A plethora of models describe height growth, but only Schumacher equation (Schumacher 1939) and 4^{th} degree polynomial were selected as: 1) both functions have horizontal asymptotes, 2) are simple and easy to understand and made inferences, and 3) are differentiable, which makes the computation of the Jacobian and Hessian matrices easy. Schumacher type model **Equation 4-1** has an asymptote at e^{k0}, and the polynomial function **Equation 4-2**

at $e^{k1} \times 10^{k2}$, where $H_{Schumchaer}$ and $H_{Polynomial}$ are the height computed with Schumacher and polynomial equations, respectively, *k0, k1* and *k2* are parameters.

$$H_{Schumacher} = e^{(5-\frac{Sch\_exp}{age})} \quad \textbf{Equation 4-1}$$

$$H_{Polynomial} = 10^{(-29.468*(-0.02+\frac{1}{age})+938.97*(-0.02^2+\frac{1}{age^2})-16102*(-0.02^3+\frac{1}{age^3})+88775*(-0.02^4+\frac{1}{age^4}))} \quad \textbf{Equation 4-2}$$

The two equations are customary used in the southern region of the USA for pine plantations, the first one is an adjustment from A very and Burkhart (2001) while the second one is from Farrar (1981). Considering that height growth is a function of site productivity, an additional term was included in the model, which took values from 90, 81.4 and 73.7. The three values, which represent the site index (Avery and Burkhart 2001), multiplies each of the two functions representing the change in height with age. The height was predicted from age 20 to 120 with a time-step of 5 years, similar to Farrar (1981).

To mimic reality, noise was generated by using a Gamma distribution **Equation 4-3.** Considering that natural variability can be represented by a multitude of distribution, ranging from exponential to unimodal symmetric, a set of distributions were generated. The distributions were defined by the two parameters of the Gamma distribution, alpha and beta parameters. Alpha ranged from 0.5 to 5 in increments of 0.5, while beta increased from 1 to 4 and are in steps of 1.

$$noise = \Gamma(\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}(\frac{\alpha}{\beta})^{\alpha-1}e^{-\beta(\frac{\alpha}{\beta})} = \frac{\beta}{\Gamma(\alpha)}(\alpha)^{\alpha-1}e^{-\alpha} = \beta * \Gamma(\alpha) \quad \textbf{Equation 4-3}$$

Since there were 10 values for alpha and 4 values for beta, forty pairs (10×4) of $\Gamma(\alpha, \beta)$ were produced. To increase even more the generality of the simulated data 10 replications were run for each pair alpha-beta, which lead to a total of 50,400 observations (21 values of age × 2 height models × 40 alpha and beta cases× 3 site indices × 10 replication) were completed; 25,200 for each model. The relationship between age and height for each model is shown in **Figure 4-2**.

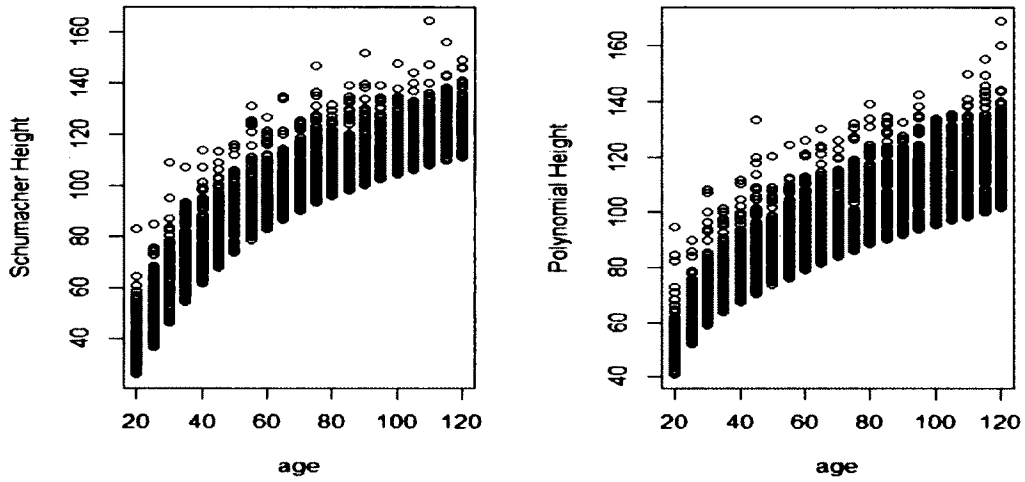


Figure 4-2: Synthetically generated data for loblolly pine using Schumacher exponential relationship **Equation 4-1** and polynomial equation of Farrar **Equation 4-2**.

4.2.2 Models and Algorithms Description (Phase I)

The generated data will be treated as if the appropriate models used to build it are unknown. Then the inference is made by executing two height models (i.e., Schumacher and Polynomial type), the same models used to generate the data. In essence, the exact same models used to construct the data will be tested on the respective data, in anticipation of an excellent goodness of fit.

Polynomial model: $height = b_0 + SI \times (b_1/\text{age} + b_2/\text{age}^2 + b_3/\text{age}^3 + b_4/\text{age}^4)$ **Equation 4-4**

Schumacher model: $\text{height} = b_0 \times SI \times \exp((5 - b_1)/\text{age})$ **Equation 4-5**

where b_i are parameters to be estimated.

To ensure validity and completeness of the findings that describe change in forest height the modeling will be executed in two phases. The first phase will be fitting the selected height models to the data, aiming at estimating the parameters of each model. The assessment measures were mean square error (MSE), standard error (SE), and coefficient of determination (R^2). In the second phase, the autoregressive process (Brockwell & Davis, 1996) was applied on the residuals supplied by models in the first phase, as measurements used to represent tree growth are executed on the same entities, which lead to a correlated process.

The final models were obtained by applying four algorithms to the autoregressive equations describing the temporal dependencies between residuals. Each algorithm was applied to the differenced errors resulting from both height models. The last step compared the Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (SBC) and MSE supplied by each algorithm, based on alpha and beta values.

Two models were used to estimate tree height, and each model studied the relation between height versus age and SI. SI variable has been chosen to be a part of the model since it provides important information on tree development, as it quantifies site productivity, which is a key on the height. To identify the models applied to a particular generated data a two symbols abbreviation was employed: the first letter represent the

fitted model and the second letter the model used to generate the heights (e.g., P-P stands for polynomial estimated model on polynomial generated data)

In the first stage the analysis was executed using the nonlinear regression procedure NLIN of SAS 9.3 (SAS Institute, Inc., 2012). Estimation of the parameters of each model was assessed using three statistics, namely means square error (MSE), standard error (SE) and coefficient of determination (R^2):

$$MSE = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n - p} \qquad \textbf{Equation 4-6}$$

$$SE = \frac{S}{\sqrt{n}} \qquad \textbf{Equation 4-7}$$

$$R^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2 - \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \qquad \textbf{Equation 4-8}$$

where n is the number of observations, p is the number of the parameters, S is the standard deviation, Y_i , $\hat{Y}_i$,and $\bar{Y}$ are the actual, the predicted, and the average values for the dependent variable, respectively. n is the total number of observations, and p is the number of parameters in each model.

The first modeling phase is completed when adequate nonlinear models are obtained, even that the residuals are serial correlated. In the second modeling phase the serial correlation is removed by differencing, starting with lag 1, as following:

$$residual_{\,1}(\varepsilon) = \varepsilon_t - \varepsilon_{t-1} \qquad \textbf{Equation 4-9}$$

4.2.3 Algorithms Evaluation by using Autoreg (Phase II)

After the time series had been differencing, the next step is fitting the autoregressive model to evaluate the algorithm's performance. In SAS 9.3 (SAS Institute, Inc., 2012), autoregressive procedure is used to estimate linear regression models of a time series data which is in the following formula:

$$Y_t = X_t'\beta + V_t \qquad \textbf{Equation 4-10}$$

where Y_t is the response variable, X_t' is the regressor variable with β slope, and V_t is the error term. As errors are correlated they are represented by **Equation 4-11**:

$$V_t = -\varphi_1 V_{t-1} - \cdots - \varphi_m V_{t-m} + \varepsilon_t \qquad \textbf{Equation 4-11}$$

After considering the differenced errors of height resulted from the first phase, the obtained residuals should be white noise a_t, with zero mean, constant variance, and zero covariance.

To assess the significance of the relationship between residuals, autocorrelation function (ACF), partial autocorrelation function (PACF) and inverse autocorrelation function (IACF) were used, as suggested by Brockwell & Davis (1996). In eventuality that significant correlations were noticed for some lags, the residuals were differenced in various configurations to remove the relationships, and to ensure stationarity. Several differencing schemes were considered to reach the white noise status, but the start was the difference of two consecutive residuals, specifically $\Delta\varepsilon_i = \varepsilon_i - \varepsilon_{i-1}$ (where ε_i is the residual at the i^{th} age).

In eventuality that white noise was not obtained by differencing with a lag of 1 then an additional differencing was executed, but now on the $\Delta\varepsilon_i$. The new difference was

of the form $\Delta_k(\Delta\varepsilon_i)_j$, where Δ_k is the lag k difference. The lag considered was not larger than 6, as values larger than 6 would contradict the development of the stem from physiological perspective by implying that events occurring more than six years apart, approximately 20% of rotation, are significantly related. Therefore, the models considered in analysis were **Equations 4-12** to **4-16** where e is white noise.

- Difference at lag 2:

$$lag_2(\Delta\varepsilon) = \Delta_2(\Delta\varepsilon) + e = \Delta_2(\varepsilon_i - \varepsilon_{i-1}) + e = \Delta_2(\varepsilon_i) - \Delta_2(\varepsilon_{i-1}) + e = (\varepsilon_i - \varepsilon_{i-2}) - (\varepsilon_{i-1} - \varepsilon_{i-3}) + e \Rightarrow \varepsilon_i - \varepsilon_{i-2} - \varepsilon_{i-1} + \varepsilon_{i-3} + e$$

$$\hat{y} = f(age, si) + \varepsilon_i - \varepsilon_{i-1} - \varepsilon_{i-2} + \varepsilon_{i-3} + e \qquad \textbf{Equation 4-12}$$

- Difference at lag 3:

$$lag_3(\Delta\varepsilon) = \Delta_3(\Delta\varepsilon) + e = \Delta_3(\varepsilon_i - \varepsilon_{i-1}) + e = \Delta_3(\varepsilon_i) - \Delta_3(\varepsilon_{i-1}) + e = (\varepsilon_i - \varepsilon_{i-3}) - (\varepsilon_{i-1} - \varepsilon_{i-4}) + e \Rightarrow \varepsilon_i - \varepsilon_{i-3} - \varepsilon_{i-1} + \varepsilon_{i-4} + e$$

$$\hat{y} = f(age, si) + \varepsilon_i - \varepsilon_{i-1} - \varepsilon_{i-3} + \varepsilon_{i-4} + e \qquad \textbf{Equation 4-13}$$

- Difference at lag 4:

$$lag_4(\Delta\varepsilon) = \Delta_4(\Delta\varepsilon) + e = \Delta_4(\varepsilon_i - \varepsilon_{i-1}) + e = \Delta_4(\varepsilon_i) - \Delta_4(\varepsilon_{i-1}) + e = (\varepsilon_i - \varepsilon_{i-4}) - (\varepsilon_{i-1} - \varepsilon_{i-5}) + e \Rightarrow \varepsilon_i - \varepsilon_{i-4} - \varepsilon_{i-1} + \varepsilon_{i-5} + e$$

$$\hat{y} = f(age, si) + \varepsilon_i - \varepsilon_{i-1} - \varepsilon_{i-4} + \varepsilon_{i-5} + e \qquad \textbf{Equation 4-14}$$

- Difference at lag 5:

$$lag_5(\Delta\varepsilon) = \Delta_5(\Delta\varepsilon) + e = \Delta_5(\varepsilon_i - \varepsilon_{i-1}) + e = \Delta_5(\varepsilon_i) - \Delta_5(\varepsilon_{i-1}) + e = (\varepsilon_i - \varepsilon_{i-5}) - (\varepsilon_{i-1} - \varepsilon_{i-6}) + e \Rightarrow \varepsilon_i - \varepsilon_{i-5} - \varepsilon_{i-1} + \varepsilon_{i-6} + e$$

- Difference at lag 6: **Equation 4-15**

$$lag_6(\Delta\varepsilon) = \Delta_6(\Delta\varepsilon) + e = \Delta_6(\varepsilon_i - \varepsilon_{i-1}) + e = \Delta_6(\varepsilon_i) - \Delta_6(\varepsilon_{i-1}) + e = (\varepsilon_i - \varepsilon_{i-6}) - (\varepsilon_{i-1} - \varepsilon_{i-7}) + e \Rightarrow \varepsilon_i - \varepsilon_{i-6} - \varepsilon_{i-1} + \varepsilon_{i-7} + e$$

$$lag_6(\Delta\varepsilon) = \Delta_6(\Delta\varepsilon) + e = \Delta_6(\varepsilon_i - \varepsilon_{i-1}) + e = \Delta_6(\varepsilon_i) - \Delta_6(\varepsilon_{i-1}) + e = (\varepsilon_i - \varepsilon_{i-6}) - (\varepsilon_{i-1} - \varepsilon_{i-7}) + e \Rightarrow \varepsilon_i - \varepsilon_{i-6} - \varepsilon_{i-1} + \varepsilon_{i-7} + e$$

$$\hat{y} = f(age, si) + \varepsilon_i - \varepsilon_{i-1} - \varepsilon_{i-6} + \varepsilon_{i-7} + e \qquad \textbf{Equation 4-16}$$

The impact of each algorithm with each of two models was assessed for every alpha-beta combination, which supplied information in each group of the resulted trees. Moreover, the larger amount of information obtainable for each alpha-beta pair drove to more precise models (Neter *et al.*, 1996). It is important to declare that by doing this analysis, the hypothesis which claims that the default option algorithm under any statistical software procedure is not always the perfect choice for a predefined model.

The white noise is obtained by using an autoregressive model on every alpha and beta pair using four algorithms, namely Yule Walker (YW), Iterative Yule Walker (ITYW), Maximum Likelihood (ML), and Unconditional Least Square (ULS).

4.2.3.1 *Yule Walker (YW)*

This algorithm is the default option in the autoregressive procedure in SAS. In a large sample from an AR (p) process, YW algorithm is represented by **Equation 4-17**:

$$C_t = \sum_{1}^{p} \phi_p C_{t-p} + \sigma_\varepsilon^2 \delta_{t.0} \qquad \textbf{Equation 4-17}$$

where C_t is the autocovariance function of the series, ϕ is the vector of autoregressive parameters which is the estimation of β, p is the autoregressive lag order, σ_ε^2 is the standard deviation of the residual, and $\delta_{t.0}$ is the Kronecker delta function.

4.2.3.2 *Iterative Yule Walker (ITYW)*

This algorithm uses the resulted residual that comes out of YW algorithm to create new estimators of ϕ and V (the error vector).

4.2.3.3 *Maximum Likelihood (ML)*

The Maximum Likelihood algorithm is efficient but is in need of a good starting point and cannot be computed for some data. Thusly, it is maximized by minimizing the objective function:

$$|L|^{1/N} e'e|L|^{1/N} \qquad \textbf{Equation 4-18}$$

where e is the residual vector and L is the likelihood function which represented as

$$-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|V|) - \frac{S}{2\sigma^2} \qquad \textbf{Equation 4-19}$$

where σ^2 is the variance, V is variance matrix of the error vector, $|V|$ is the determinant of V, N the number of the observations, and $S = e'.e$ (unconditional sum of squares of the model) where e is the transformed error.

4.2.3.4 *Unconditional Least Squares (ULS)*

The last algorithm that can be used to solve the autoregressive equation is the ULS, which is computed by minimizing S with respect to the parameters β and ϕ_i (SAS, Inc., 2012). Assessment of the models was executed using MSE on e.g., lowest value MSE or AIC), or the maximum coefficient of determination (R^2) (Dayton, 2003; Cameron & Windmeijer, 1995). In this research, pairwise comparison has been used, which means selecting the model based on three statistics: AIC, SBC and MSE. These methodologies are mostly used in time series and regression model analysis. AIC and SBC measure goodness of fit of the model and complexity of the model, and are calculated with the following formulas:

$$AIC = -2\ln(L) + 2k \qquad \textbf{Equation 4-20}$$

$$SBC = -2\ln(L) + k\ln(n) \qquad \textbf{Equation 4-21}$$

where L is the maximized value of the likelihood function for the estimated model and K is the number of estimated parameters and n is the total number of observations. To assess the average performance of the algorithms one-way ANOVA was performed using AIC as dependent variable. The model was:

$$AIC = \text{Algorithm} \qquad \textbf{Equation 4-22}$$

where algorithm had four values: YW,ITYW, ML and ULS.

4.3 Results

4.3.1 Data Generation

A set of 38,400 models were generated from a factorial combination of two models to generate heights, two fitted height models on generated data, four autoregressive numerical algorithms, forty alpha-beta pairs, 20 replications and three site productivities. Inference was based on **Equations 4-12** to **4-16**, which modeled the residual for several orders of the autoregressive process (i.e., 2, 3, 4, 5 and 6).

4.3.2 Phase I Models

The parameters of the fitted nonlinear models (i.e., Schumacher and Polynomial) are presented in **Table 4-1**, as well as the associated assessment statistics. All models and all parameters were significant (p-value<0.001).

According to **Table 4-1** Schumacher model has high R^2 (larger than 0.98). The parameters for polynomial models reflect the magnitude of data and their sign depends on the fit of the model. For Schumacher model to ensure conformity with development

processes the parameter in the exponential term should be negative (i.e., there is an asymptote of tree highlight), value presented to both type of generated data. The results from **Table 4-1** are mirrored by the graphs of plotting the equation used to generate the data with the fitted models.

Table 4-1: Parameter estimates and fit statistics of the Height-Age equations. The heading symbols were explained in text and P-S stands for polynomial model on Schumacher generated heights, P-P for polynomial model on polynomial generated heights, S-S for Schumacher model on Schumacher generated heights, and S-P for Schumacher model on polynomial generated heights

Model - Data	Parameter	Parameter Estimation	SE	MSE	R^2
P-S	b_0	52.1094	0.2741	75.8809	0.8697
	b_1	186.9	0.6676		
	b_2	-14396.0	46.7743		
	b_3	375579	1326.5		
	b_4	-3282375	12838.2		
P-P	b_0	39.2989	0.2826	80.6748	0.7973
	b_1	205.3	0.6883		
	b_2	-15561.4	48.2292		
	b_3	416357	1367.7		
	b_4	-3735830	13237.6		
S-S	b_0	1.8633	0.000950	58.9702	0.9941
	b_1	31.8967	0.0322		
S-P	b_0	1.6660	0.000766	51.0007	0.9947
	b_1	25.0731	0.0274		

Figure 4-3 explains the relationship between the tree heights and their age for both models end every single site index.

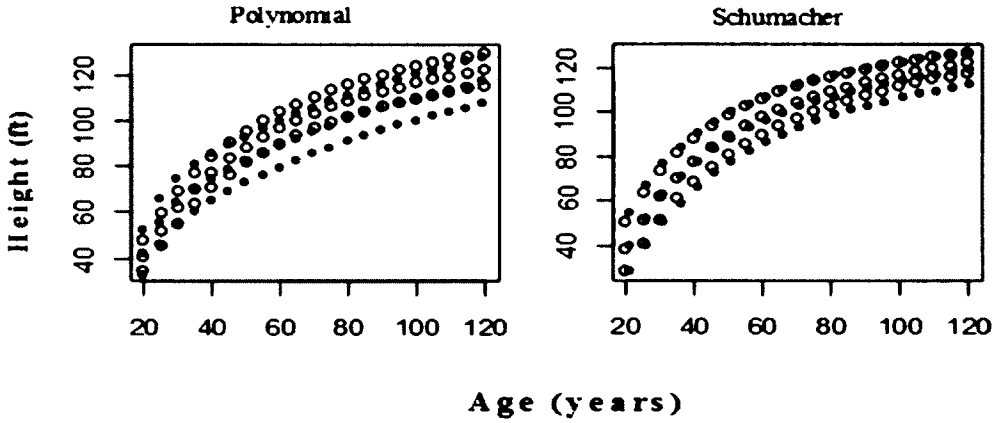


Figure 4-3: Increase in height with age described by the Schumacher and Polynomial models for the three site indices (90, 81.4, and 73.7). The curve with the empty circles represents Schumacher generated heights and the solid circles represent Polynomial generated heights.

According to **Table 4-1**, the lowest recorded MSE are present to the Schumacher model, meaning the accuracy of the Polynomial model performs better than Schumacher model, irrespective the generated data. Polynomial height is less accurate than Schumacher height, as growth is more pronounced at early ages. Polynomial model places the two heights curves closely, irrespective the site index, while Schumacher model made a clear distinction among site productivities. **Figure 4-4** represents the mean standard error (MSE) values for Polynomial and Schumacher models for both heights.

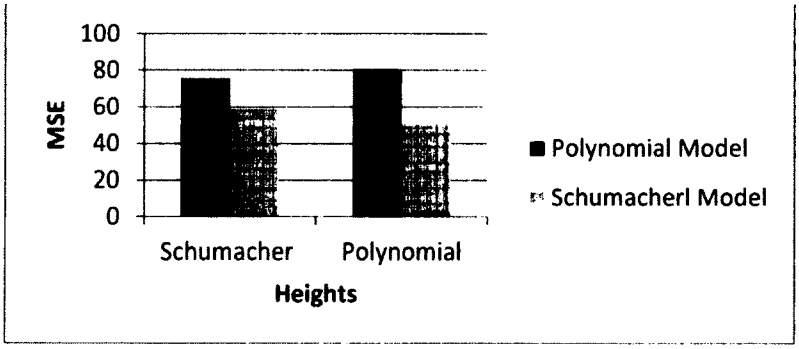


Figure 4-4: The mean standard error (MSE) bar plots of the estimated height of every growth model for both heights type. The left group of bar for the Schumacher height and the left group for the Polynomial height.

Figure 4-5 represents the relationship between the generated heights (actual values) and the predicted height for both models.

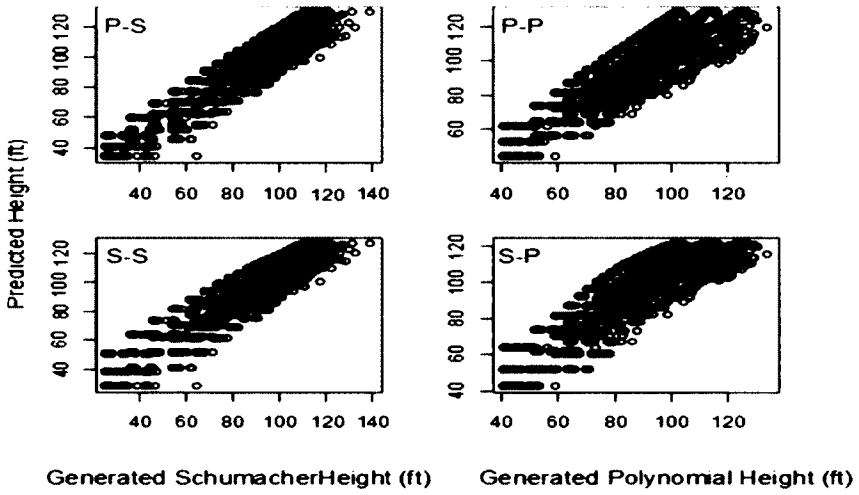


Figure 4-5: Relationships between generated heights and predicted height.

4.3.3 Phase II algorithms

To every data generated model, fitted model, SI, replication, noise (i.e., alpha and beta) the autoregressive procedure was executed, each execution being implemented by four algorithms. Therefore, the autoregressive procedure was fitted on each individual 38,400 models. Furthermore, to reach the white noise of the errors, which conclude the modeling exercise, for each autoregressive model several lags were tried, if the lag 1 difference did not work.

The residuals resulted from the nonlinear models exhibited a pronounced pattern, as the ACF, PACF and IACF revealed. Several patterns were noticed but all presented a significant ACF, PACF and IACF at lag 1 **Figure 4-6**.

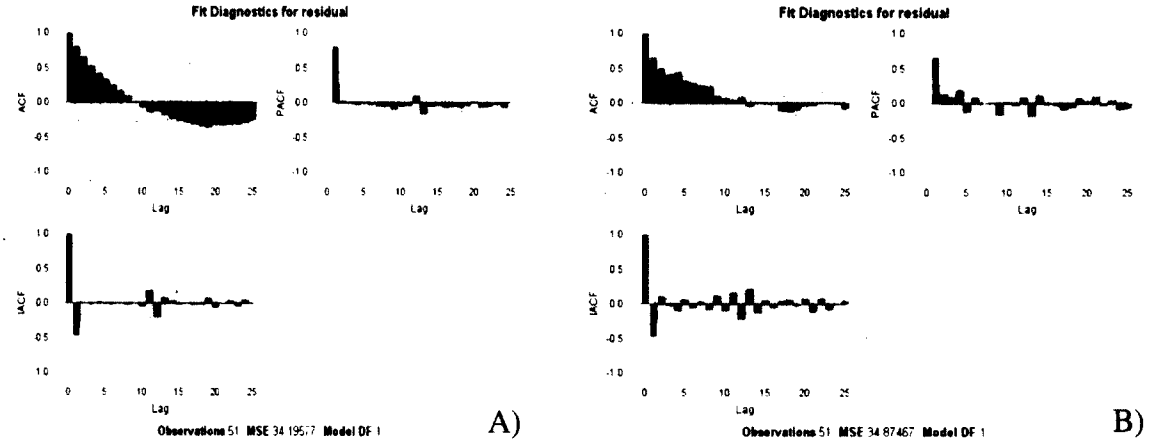


Figure 4-6: ACF, PACF and IACF of the models residuals for two cases from the 38400: A) P-P, alpha=0.5, beta=1, YW algorithm and B) S-P, alpha =5, beta=2, YW algorithm.

Therefore, an initial differencing of the residuals was executed with lag 1, trying to remove the autocorrelation. However, differencing with lag 1 did not produce the expected results; in fact in some cases the results worsen **Figure 4-7**. Hence for each data generated model, height model, alpha and beta pairs, and site productivity the less parsimonious models of the **Equations 4-12** to **4-16** were selected, such that the error term is distributed as white noise.

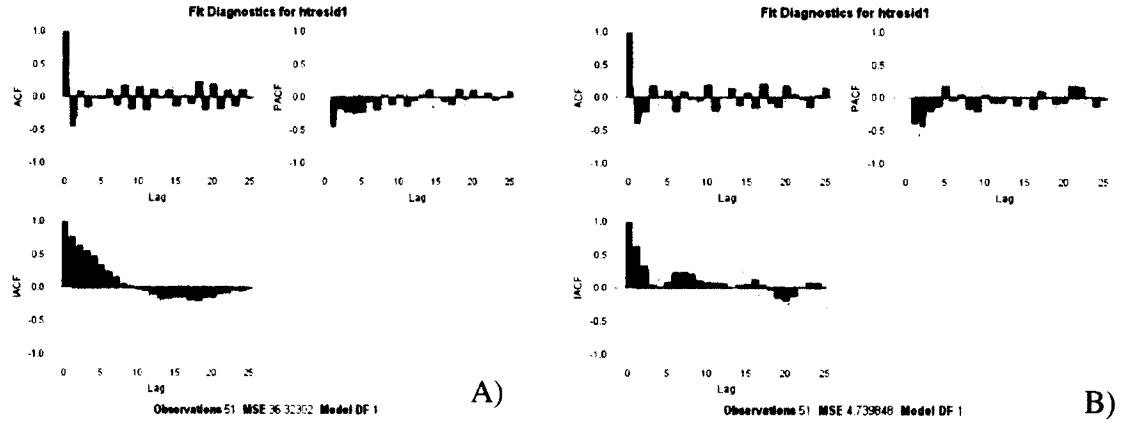


Figure 4-7: ACF, PACF and IACF of the lag 1 differenced residuals for two cases from the 38400: A) S-S, alpha=0.5, beta=2, YW algorithm and B) S-S, alpha =5, beta=2, YW algorithm

To evaluate the overall performance of each algorithm, irrespective the model complexity, a box and wishers plot was produced for the two types of generated heights **Figure 4-8**. Durbin Watson test was applied on the residuals and results showed no autocorrelation between them.

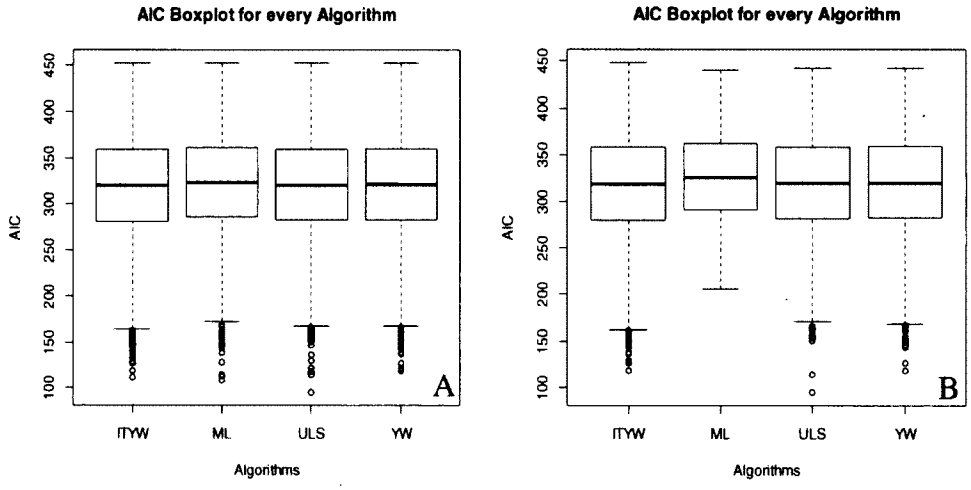


Figure 4-8: Boxplots showing the performance of each algorithm in respect with the generated model (A) is Polynomial height and B) is Schumacher height), using AIC as criterion.

Depending on the autoregressive computation algorithm the structure of the residuals differed, but across al all fitted models and all data generated models the difference at lag 3 of the differenced residuals **Equation 4-13** performed the best **Table 4-2**. The next most popular residual model was for **Equation 4-12**, but with less than 100 cases.

Table 4-2: The number of equations by type that produced white noise.

Model	Height	Algorithm	Equation	# equations
Polynomial	Polynomial	ITYW	4-12	61
			4-13	2314
			4-14	20
			4-15	3
			4-16	1
		ML	4-12	61
			4-13	2314
			4-14	20
			4-15	3
			4-16	1
		ULS	4-12	61
			4-13	2314
			4-14	20
			4-15	3
			4-16	1
		YW	4-12	61
			4-13	2314
			4-14	20
			4-15	3
			4-16	1
	Schumacher	ITYW	4-12	71
			4-13	2303
			4-14	20
			4-15	2
			4-16	2
		ML	4-12	81
			4-13	2293
			4-14	15
			4-15	4
			4-16	5
		ULS	4-12	77
			4-13	2303
			4-14	12
			4-15	4
			4-16	0
		YW	4-12	77
			4-13	2303
			4-14	12
			4-15	0
			4-16	0
Schumacher	Polynomial	ITYW	4-12	83
			4-13	2267
			4-14	31
			4-15	12
			4-16	3
		ML	4-12	90
			4-13	2266
			4-14	34
			4-15	8

			4-16	1
		ULS	4-12	97
			4-13	2264
			4-14	34
			4-15	1
			4-16	3
		YW	4-12	114
			4-13	2245
			4-14	36
			4-15	4
			4-16	1
	Schumacher	ITYW	4-12	121
			4-13	2229
			4-14	43
			4-15	4
			4-16	2
		ML	4-12	104
			4-13	2247
			4-14	34
			4-15	8
			4-16	5
		ULS	4-12	91
			4-13	2266
			4-14	41
			4-15	1
			4-16	1
		YW	4-12	98
			4-13	2253
			4-14	40
			4-15	5
			4-16	4

ANOVA confirms the boxplot summary, by revealing that there is no difference among the results supplied by all algorithms in **Table 4-3**, as for both data generated models the p-value >0.9. Tukey test showed that on average the algorithms are undistinguishable (p-value>0.99).

Table 4-3: ANOVA

Height generated using polynomial model			Height generated using Schumacher model		
Algorithm	**AIC Mean**	**Welch's ANOVA for AIC (Pr>F)**	**Algorithm**	**AIC Mean**	**Welch's ANOVA for AIC (Pr>F)**
ITYW	319.866	0.9913	ITYW	317.036	<.0001
ML	319.562		ML	326.913	
ULS	319.613		ULS	319.355	
YW	319.780		YW	319.541	
Tukey Pairwise comparison		**p-value**	**Tukey Pairwise comparison**		**p-value**
ML-ITYW		0.999	ML-ITYW		0.999
ULS-ITYW		0.999	ULS-ITYW		0.194
YW-ITYW		0.999	YW-ITYW		0.271
ULS-ML		0.999	ULS-ML		0.317
YW-ML		0.999	YW-ML		0.405
YW-ULS		0.999	YW-ULS		0.998

From the 38,400 cases the best models for each algorithm, fitted model, generated model, site productivity, and alpha were selected according to the lowest AIC, SBC and MSE in **Table 4-4**. To represent the type of noise generated only alpha was used, as it described the shape of the distribution, which ranges from open to 0 (therefore the skewness was labeled as "open" in the table), to almost symmetric unimodal (therefore the skewness was labeled as "reduced").

Table 4-4: Best cases for each algorithm, generated data, fitted model, site productivity, and alpha. The alpha describe the level of skewness of the generated noise, which was open (similar to an exponential distribution when alpha=0.5 and 1), extreme skewed (alpha=1.5, 2 and 2.5), moderately skewed (alpha=3, 3.5) and reduced skewed (alpha=4, 4.5 and 5).

Model	Height	Algorithm	Site Index	Alpha	skewness	AIC	RMSE	Lag
Schumacher	Sch	ULS	81.4	0.5	open	95.71668	0.60659	2
Schumacher	Poly	ML	90	0.5	open	108.6282	0.70156	4
Schumacher	Poly	ITYW	90	0.5	open	111.5912	0.72334	3
Schumacher	Poly	ULS	81.4	0.5	open	117.9193	0.75409	2
Schumacher	Sch	ITYW	81.4	0.5	open	118.6499	0.75939	3
Schumacher	Sch	YW	81.4	0.5	open	118.6499	0.75939	3
Schumacher	Poly	YW	81.4	0.5	open	121.4686	0.7796	3
Schumacher	Sch	ML	90	0.5	open	218.2155	2.01447	2
Polynomial	Sch	ML	81.4	1	open	211.7804	1.87145	3
Polynomial	Sch	ULS	81.4	1	open	220.7776	2.03117	3
Polynomial	Sch	YW	81.4	1	open	223.1135	2.11134	3
Polynomial	Sch	ITYW	81.4	1	open	230.6519	2.27531	3
Polynomial	Poly	ML	73.7	1	open	233.5917	2.33032	3
Polynomial	Poly	ULS	73.7	1	open	233.754	2.32722	3
Polynomial	Poly	ITYW	73.7	1	open	235.0604	2.37607	3
Polynomial	Poly	YW	73.7	1	open	235.0755	2.37645	3
Schumacher	Sch	ULS	90	1	open	150.547	1.03817	3
Schumacher	Poly	ML	90	1	open	154.6066	1.08036	3
Schumacher	Poly	ITWY	90	1	open	144.7931	0.98082	3
Schumacher	Poly	ULS	81.4	1	open	152.5464	1.05213	3
Schumacher	Sch	ITWY	90	1	open	146.1344	0.99449	3
Schumacher	Sch	YW	90	1	open	145.1467	0.98486	3
Schumacher	Poly	YW	90	1	open	147.9092	1.01044	3
Schumacher	Sch	ML	90	1	open	228.5976	2.22911	2
Polynomial	Sch	ML	81.4	0.5	open	230.9733	2.27666	3
Polynomial	Sch	ULS	81.4	0.5	open	236.4885	2.40439	3
Polynomial	Sch	YW	81.4	0.5	open	236.7286	2.41439	3
Polynomial	Sch	ITWY	90	0.5	open	240.1235	2.61875	3
Polynomial	Poly	ML	81.4	0.5	open	240.2294	2.49396	3
Polynomial	Poly	ULS	81.4	0.5	open	240.303	2.49234	3
Polynomial	Poly	ITWY	81.4	0.5	open	241.0178	2.52002	3
Polynomial	Poly	YW	81.4	0.5	open	241.0233	2.52017	3
Schumacher	Poly	ITWY	73.7	1.5	extreme	177.5197	1.35276	3
Schumacher	Poly	ML	90	1.5	extreme	185.25	1.45864	3

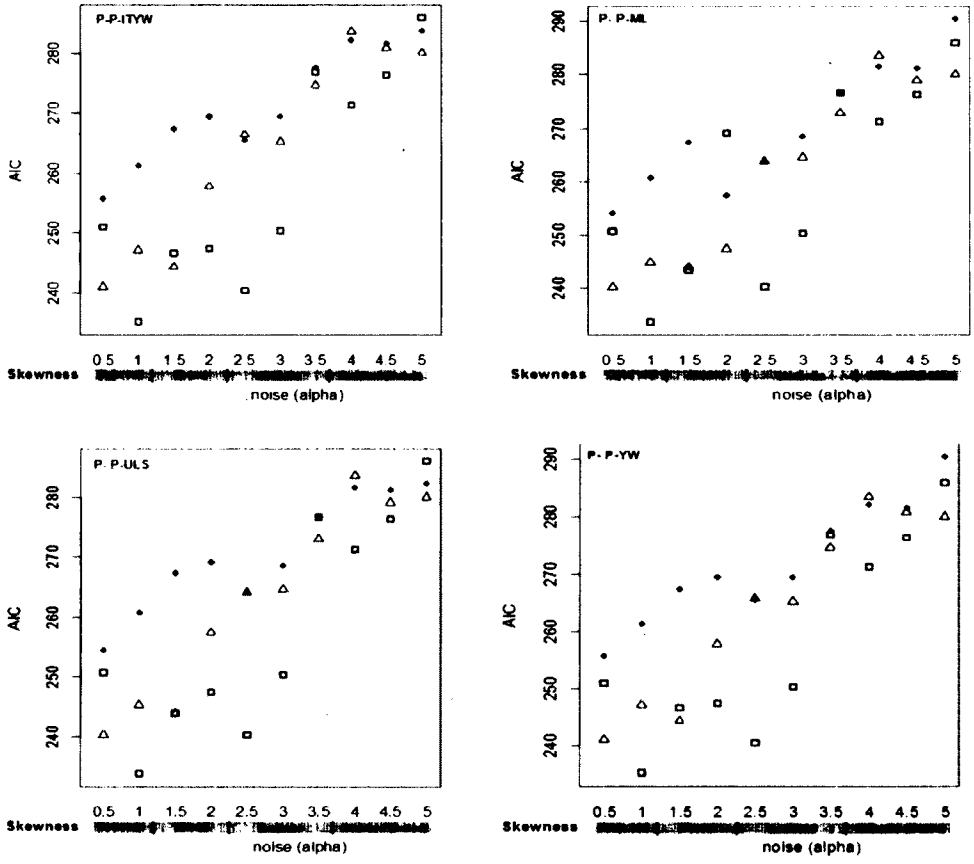
Schumacher	Poly	ULS	73.7	1.5	extreme	173.2033	1.2968	3
Schumacher	Poly	YW	81.4	1.5	extreme	168.4272	1.23705	3
Schumacher	Sch	ITWY	90	1.5	extreme	152.4949	1.05736	3
Schumacher	Sch	ML	90	1.5	extreme	205.0801	1.76415	6
Schumacher	Sch	ULS	90	1.5	extreme	156.8992	1.10306	3
Schumacher	Sch	YW	90	1.5	extreme	165.8711	1.20552	3
Polynomial	Poly	ML	73.7	1.5	extreme	243.4122	2.54727	3
Polynomial	Sch	ML	81.4	1.5	extreme	244.4816	2.6038	3
Polynomial	Poly	ULS	81.4	1.5	extreme	243.9169	2.58727	3
Polynomial	Sch	ULS	81.4	1.5	extreme	230.8678	2.25873	3
Polynomial	Poly	ITWY	81.4	1.5	extreme	244.3181	2.60302	3
Polynomial	Sch	ITWY	81.4	1.5	extreme	237.9337	2.44299	3
Polynomial	Poly	YW	81.4	1.5	extreme	244.3275	2.60329	3
Polynomial	Sch	YW	81.4	1.5	extreme	232.3837	2.31419	3
Schumacher	Poly	ITWY	90	2	extreme	178.3657	1.36318	3
Schumacher	Poly	ML	90	2	extreme	184.6608	1.44871	3
Schumacher	Poly	ULS	90	2	extreme	189.0899	1.51471	3
Schumacher	Poly	YW	90	2	extreme	197.5388	1.64583	3
Schumacher	Sch	ITWY	90	2	extreme	174.1339	1.30625	3
Schumacher	Sch	ML	90	2	extreme	218.657	2.02371	2
Schumacher	Sch	ULS	81.4	2	extreme	182.7596	1.42345	3
Schumacher	Sch	YW	90	2	extreme	174.1339	1.30625	3
Polynomial	Sch	YW	90	2	extreme	241.7463	2.53553	3
Polynomial	Poly	YW	73.7	2	extreme	247.4318	2.6849	3
Polynomial	Sch	ULS	90	2	extreme	241.0956	2.50971	3
Polynomial	Poly	ULS	73.7	2	extreme	247.4297	2.6848	3
Polynomial	Sch	ML	90	2	extreme	246.8983	2.66996	3
Polynomial	Poly	ML	73.7	2	extreme	247.4296	2.68481	3
Polynomial	Sch	ITWY	81.4	2	extreme	235.8651	2.39332	3
Polynomial	Poly	ITWY	73.7	2	extreme	247.4317	2.68489	3
Schumacher	Poly	ITWY	90	2.5	extreme	198.3901	1.65652	3
Schumacher	Poly	ML	81.4	2.5	extreme	206.0051	1.78635	3
Schumacher	Poly	ULS	73.7	2.5	extreme	199.9743	1.68593	3
Schumacher	Poly	YW	81.4	2.5	extreme	187.692	1.49309	3
Schumacher	Sch	ITWY	81.4	2.5	extreme	198.7665	1.66566	3
Schumacher	Sch	ML	81.4	2.5	extreme	222.515	2.10073	2
Schumacher	Sch	ULS	90	2.5	extreme	209.2069	1.84575	3
Schumacher	Sch	YW	81.4	2.5	extreme	188.725	1.50966	3
Polynomial	Sch	ITWY	81.4	2.5	extreme	239.5882	2.48558	3
Polynomial	Poly	ITWY	73.7	2.5	extreme	240.4288	2.50565	3
Polynomial	Poly	ML	73.7	2.5	extreme	240.2948	2.50092	3

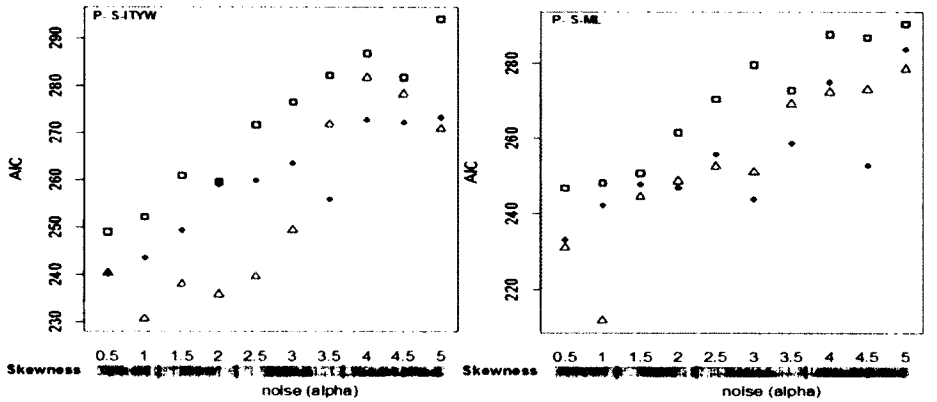
Polynomial	Sch	ML	81.4	2.5	extreme	252.4842	2.81938	3
Polynomial	Poly	ULS	73.7	2.5	extreme	240.306	2.50065	3
Polynomial	Sch	ULS	90	2.5	extreme	252.1029	2.80556	3
Polynomial	Poly	YW	73.7	2.5	extreme	240.4318	2.50573	3
Polynomial	Sch	YW	90	2.5	extreme	252.3373	2.81573	3
Schumacher	Sch	ITWY	90	3	moderate	191.1918	1.54693	3
Schumacher	Sch	ML	90	3	moderate	291.7542	4.14423	2
Schumacher	Sch	ULS	90	3	moderate	216.6789	1.9809	4
Schumacher	Sch	YW	90	3	moderate	214.1264	1.93676	3
Schumacher	Poly	ITWY	73.7	3	moderate	213.0452	1.91393	3
Schumacher	Poly	ULS	90	3	moderate	206.648	1.79994	3
Schumacher	Poly	YW	90	3	moderate	204.903	1.76854	3
Polynomial	Poly	ITWY	73.7	3	moderate	246.4597	2.65945	3
Polynomial	Poly	ML	73.7	3	moderate	246.4595	2.65944	3
Polynomial	Poly	ULS	73.7	3	moderate	246.4595	2.65944	3
Polynomial	Poly	YW	73.7	3	moderate	246.4597	2.65945	3
Polynomial	Sch	ITWY	90	3	moderate	304.9554	4.71825	3
Polynomial	Sch	ML	90	3	moderate	273.2415	3.44894	3
Polynomial	Sch	ULS	90	3	moderate	282.0401	3.76582	3
Polynomial	Sch	YW	90	3	moderate	282.1867	3.77412	3
Schumacher	Poly	ITWY	81.4	3.5	moderate	210.2794	1.86418	3
Schumacher	Poly	ML	73.7	3.5	moderate	221.0246	2.0693	3
Schumacher	Poly	ULS	81.4	3.5	moderate	217.318	1.99745	3
Schumacher	Poly	YW	90	3.5	moderate	219.8709	2.04796	3
Schumacher	Sch	ITWY	81.4	3.5	moderate	217.9715	2.0108	3
Schumacher	Sch	ML	81.4	3.5	moderate	257.1294	2.95076	2
Schumacher	Sch	ULS	90	3.5	moderate	227.3935	2.20576	3
Schumacher	Sch	YW	90	3.5	moderate	216.5875	1.9842	3
Polynomial	Sch	YW	90	3.5	moderate	288.6282	4.02086	3
Polynomial	Sch	ULS	90	3.5	moderate	288.5519	4.0162	3
Polynomial	Sch	ML	90	3.5	moderate	282.372	3.78178	3
Polynomial	Sch	ITWY	90	3.5	moderate	276.2694	3.55947	3
Polynomial	Poly	ITWY	90	3.5	moderate	303.5108	4.65246	3
Polynomial	Poly	ML	90	3.5	moderate	303.4675	4.6498	3
Polynomial	Poly	ULS	90	3.5	moderate	303.4701	4.64965	3
Polynomial	Poly	YW	90	3.5	moderate	303.512	4.65252	3
Schumacher	Sch	ITWY	90	4	reduced	222.6445	2.10516	3
Schumacher	Sch	ML	81.4	4	reduced	307.6033	4.83552	4
Schumacher	Sch	ULS	90	4	reduced	247.1194	2.67643	3
Schumacher	Sch	YW	90	4	reduced	239.67	2.48817	2
Schumacher	Poly	ITWY	73.7	4	reduced	224.3936	2.14005	3

Schumacher	Poly	ML	90	4	reduced	228.4635	2.22801	3
Schumacher	Poly	ULS	73.7	4	reduced	213.6872	1.92768	3
Schumacher	Poly	YW	73.7	4	reduced	220.1505	2.05362	3
Polynomial	Poly	YW	90	4	reduced	316.4521	5.28164	3
Polynomial	Poly	ULS	90	4	reduced	316.36	5.27443	3
Polynomial	Poly	ML	90	4	reduced	316.3534	5.27482	3
Polynomial	Poly	ITWY	90	4	reduced	316.4492	5.28148	3
Polynomial	Sch	ITWY	90	4	reduced	276.8242	3.58154	3
Polynomial	Sch	ML	90	4	reduced	279.8231	3.6869	3
Polynomial	Sch	ULS	90	4	reduced	285.9714	3.90943	3
Polynomial	Sch	YW	90	4	reduced	286.1333	3.92038	3
Schumacher	Poly	ITWY	81.4	4.5	reduced	234.4843	2.36485	3
Schumacher	Poly	ML	81.4	4.5	reduced	231.9641	2.30694	3
Schumacher	Poly	ULS	81.4	4.5	reduced	228.2672	2.22227	3
Schumacher	Poly	YW	73.7	4.5	reduced	221.0452	2.07176	3
Schumacher	Sch	ITWY	90	4.5	reduced	240.7219	2.51174	3
Schumacher	Sch	ML	90	4.5	reduced	321.7263	5.55891	2
Schumacher	Sch	ULS	90	4.5	reduced	245.3393	2.62681	3
Schumacher	Sch	YW	90	4.5	reduced	231.3613	2.2928	3
Polynomial	Sch	YW	90	4.5	reduced	283.6265	3.82504	3
Polynomial	Sch	ULS	90	4.5	reduced	283.5904	3.82144	3
Polynomial	Sch	ML	90	4.5	reduced	288.1191	4.00058	3
Polynomial	Sch	ITWY	90	4.5	reduced	271.2638	3.3906	3
Polynomial	Poly	ITWY	90	4.5	reduced	295.9005	4.3164	3
Polynomial	Poly	ML	90	4.5	reduced	295.7991	4.31011	3
Polynomial	Poly	ULS	90	4.5	reduced	295.8083	4.30973	3
Polynomial	Poly	YW	90	4.5	reduced	295.9052	4.31663	3
Schumacher	Sch	ITWY	90	5	reduced	229.5177	2.24908	3
Schumacher	Sch	ML	90	5	reduced	268.7227	3.30133	2
Schumacher	Sch	ULS	90	5	reduced	239.7764	2.48933	3
Schumacher	Sch	YW	90	5	reduced	252.606	2.82458	3
Schumacher	Poly	ITWY	90	5	reduced	238.609	2.46141	3
Schumacher	Poly	ML	73.7	5	reduced	227.3696	2.20377	3
Schumacher	Poly	ULS	90	5	reduced	232.2543	2.31347	3
Schumacher	Poly	YW	90	5	reduced	243.1071	2.57339	3
Polynomial	Poly	YW	90	5	reduced	296.9522	4.36192	3
Polynomial	Poly	ULS	90	5	reduced	296.8179	4.3531	3
Polynomial	Poly	ML	90	5	reduced	296.8083	4.35352	3
Polynomial	Poly	ITWY	90	5	reduced	296.9483	4.36174	3
Polynomial	Sch	ITWY	81.4	5	reduced	290.7355	4.10178	3
Polynomial	Sch	ML	81.4	5	reduced	290.9006	4.11159	3

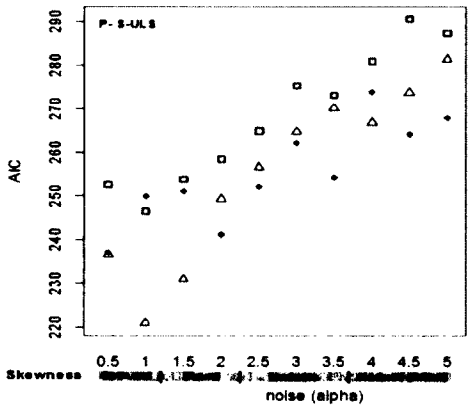
Polynomial	Sch	ULS	81.4	5	reduced	287.7327	3.98582	3
Polynomial	Sch	YW	81.4	5	reduced	287.7348	3.98595	3

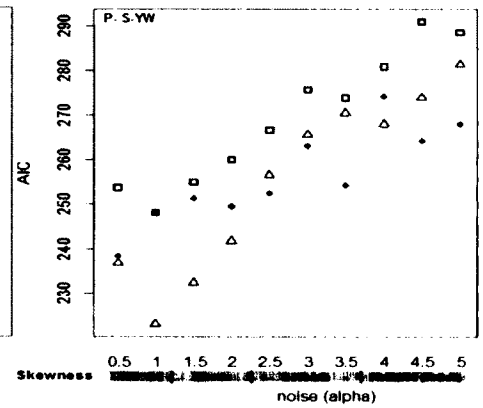
The AIC for the best cases for each algorithm, generated data, fitted model, site productivity, and alpha showed that the fit decreases according to the noise, in the sense that the larger the noise (i.e., alpha) the weaker the fit **Figure 4-9**. This comes as no surprise, considering that generated noise is one level of magnitude less than height values, therefore smaller the noise better the fit.

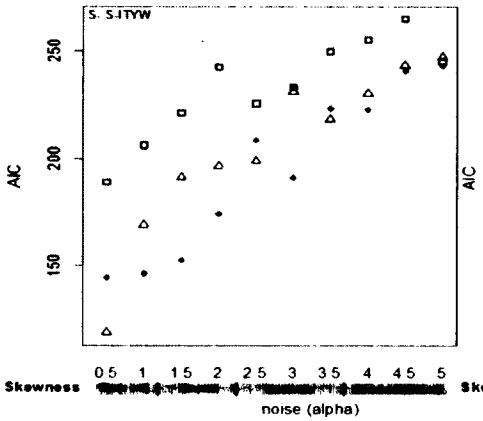


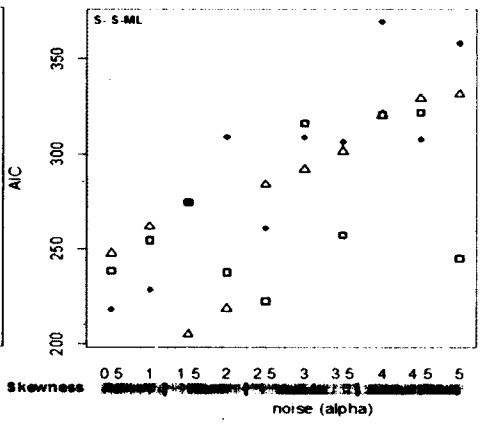

P- S-ITYW
AIC
Skewness
noise (alpha)

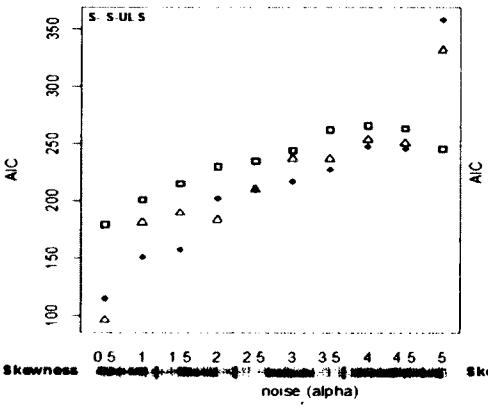
P- S-ML
AIC
Skewness
noise (alpha)

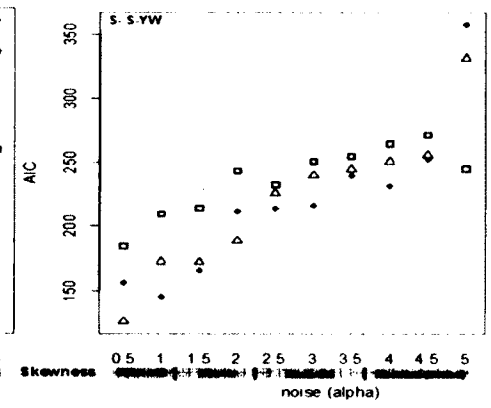

P- S-ULS
AIC
Skewness
noise (alpha)


P- S-YW
AIC
Skewness
noise (alpha)


S- S-ITYW
AIC
Skewness
noise (alpha)


S- S-ML
AIC
Skewness
noise (alpha)


S- S-ULS
AIC
Skewness
noise (alpha)


S- S-YW
AIC
Skewness
noise (alpha)

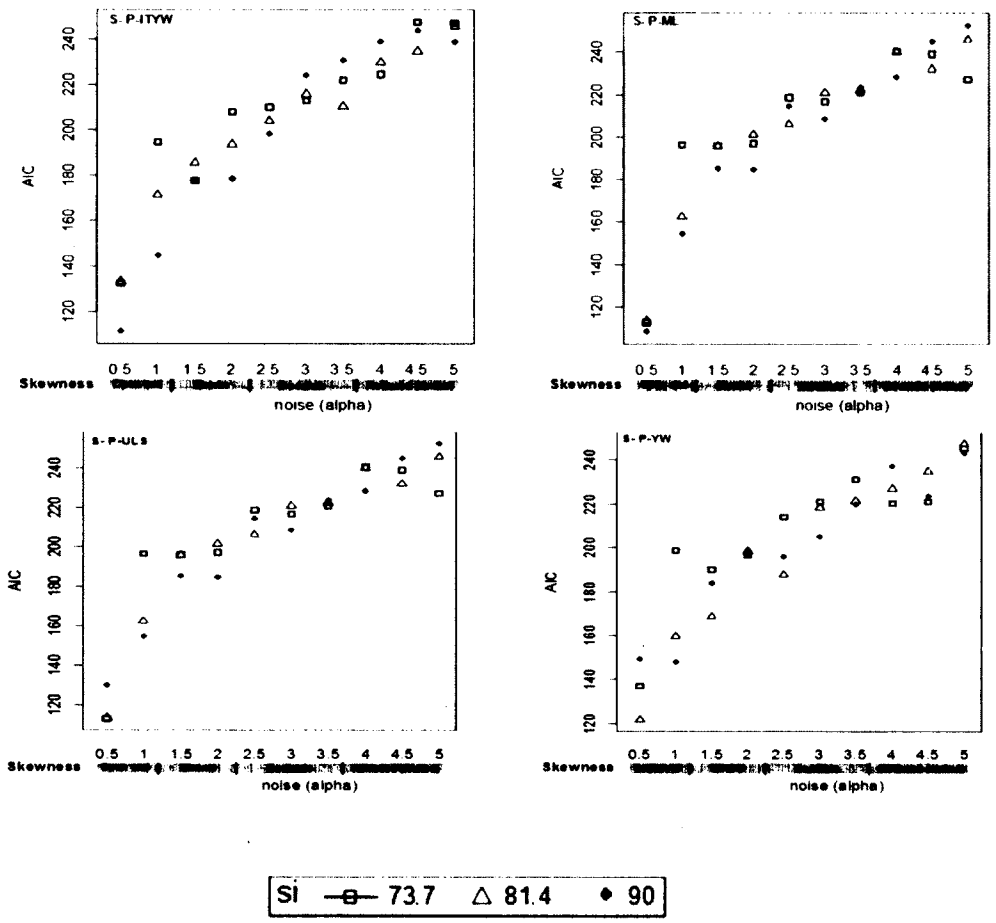


Figure 4-9: Ranking of models for each model, height, and algorithm according to AIC for every site index. The heading symbols were explained in text (i.e., P-S-ML stands for polynomial model on Schumacher generated height with respect to ML algorithm)

Mimicking the strategy of Strimbu and Paun (2013), comparisons were made on the best 10 models according to AIC, SBC and MSE from to the model used to generate the data, fitted model and alpha. Therefore, for each data generated model 200 cases were selected (10 alpha x 2 fitted models). To assess the performance of the algorithms the number of models leading to good results were counted for each algorithm **Figure 4-10**.

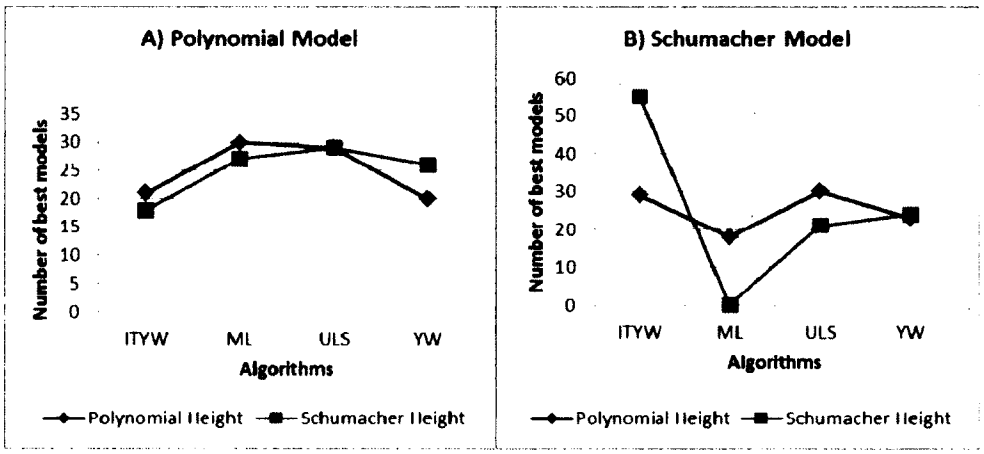


Figure 4-10: Number of models that lead to white noise produced by every algorithm for each data generated model and fitted model.

Various results according to **Figure 4-10** were obtained from fitting every algorithm as the following: for Polynomial model and Polynomial height, 15% of the effective impact was produced by applying the ML algorithm, 10.5% by the ITYW, 14.5% by the ULS, and finally 10% was produced by applying YW. At that time, for Polynomial model and Schumacher height, ULS and ML were the most effective algorithms with 14.5% and 13.55, respectively. The least significance algorithm this time is ITYW with 9%. Additionally, for Schumacher model and Polynomial height ULS and ITYW were the most effective algorithms by 15% and 14.5 % respectively. The least significance algorithm was ML with 9%. Finally for Schumacher model and Schumacher height, ML was the worst algorithms with have no number of best models at all, and the most significance algorithm was ITYW with 27.5% of the best models achieved by using it.

4.4 Discussion

The results support the decision of software developer of having the least complex algorithm as default in solving autocorrelated models, as on average the algorithms are not different, conclusion supported by ANOVA. However, each analysis is an individual case; therefore a data analyst should be aware that the section of the algorithm could lead to wrong results. In fact YW algorithm, which is the default option in most software packages, proved to be a bad choice. YW was the worse algorithm in properly identifying the actual model (i.e., match the fitted model with data generation model) when polynomial function was used. The situation did not changed significantly for Schumacher equation, as YW was the third algorithm but only when the generated noise was very small, situation when the departure from deterministic relationship was minimal.

The most effective algorithms were ML and ULS algorithms. The best result for the Schumacher model was the noise has a small magnitude, which indicates that the selection of the proper nonlinear model is pivotal in further analyses. This means that the wrong model cannot be improved by choosing the correct residual autoregressive model.

4.5 Conclusion

The tree's Height-Age relationships have been widely discussed in forest growth and yield through modeling by using linear and nonlinear functions. Frequently, the best model assessment is performed based upon selection criterion measurement as a mean square error, AIC, and SBC. Furthermore, the judgment of the best model or algorithm should be based upon more than one test or investigation. More significantly, several numerical algorithms when associated with a model can behave better and have a higher

impact than other algorithms associated with the same model when applied on the same dataset. To illustrate the approach of this study, assumption that if the true model is known was used. After that by expending the appropriate modeling one should be able to reach a formula similar, if not identical, to the true model.

To mimic reality, noise was generated using Gamma distribution which defined by the two parameters of the Gamma distribution, alpha and beta parameters. Alpha ranged from 0.5 to 5 in increments of 0.5, while beta increased from 1 to 4 and are in steps of 1. The data was generated by specific models, and then used as if the true models used to build it are unknown. Then the implication is made by executing two height models (i.e., Schumacher and Polynomial type), the same models used to generate the data, then they used for testing the goodness of fit. The white noise of the generated data was obtained by using an autoregressive model on every alpha and beta pair using four numerical algorithms: Yule Walker (YW), Iterative Yule Walker (ITYW), Maximum Likelihood (ML), and Unconditional Least Square (ULS).

The final outcomes were achieved in this paper supported the decision of software developer of having the least complex algorithm as default in solving autocorrelated models, as on average the algorithms are not different. However, each analysis is an individual case; therefore a data analyst should be aware that the section of the algorithm could lead to wrong results. In fact YW algorithm, which is the default option in most software packages, showed to be a bad choice. The most effective algorithms were ML and ULS algorithms.

CHAPTER 5

ANALYTICAL ASSESSMENT FOR AUTOREGRESSIVE ALGORITHM'S IMPACT ON DIAMETER-HEIGHT (TAPER) MODELS OF LONGLEAF PINE TREES

5.1 Introduction

Software applications have been used to solve very complicated problems in all research fields. As well, any procedure in most software can be executed through different numerical methods (algorithms). The key point is which algorithm is the best choice for solving a model as a means of generating the best results. In the analytical interpretation of modeling analyses and their performance comparison, discussing embedded algorithms under software procedures is a very important step in the comprehension of the effects on the model's results. Many researchers take on diverse statistical mechanisms to have a clear decision about the difference between the algorithms and the depiction of general conclusions (Demsar, 2006).

Analysis and the modeling of time series data is a significant topic of research in many fields (Meek, Chickering, & Heckerman, 2002). Therefore, time series is more critical in defeat stated assumptions than modeling the independent data (Gujarati, 1995). This analysis helps in studying any natural incident occurring repeatedly in periods of time and predicts how the future will look based upon these periods.

This study focuses on the implementation of autoregressive process to a time series model, and compares the four possible results out of this process based on the noise (residuals) of the data. Concepts applied aids in learning the data efficiently while making the results easier to be interpreted and compared. The mentioned four achievable results are produced by numerical algorithms used to solve the autoregressive equations.

Delivering the research ideology is necessary of a situation inclusive of the time series in which the observations are repeatedly recorded in periods of time. The tree Diameter-Height relationship is an important element of growth and yield models in forestry, which is a good example to use in applying the concept of this research. This relationship is a very important indicator of any tree growth, and can be affected by human intervention such as burning, industrial pollution, pruning, or thinning; or by natural causes such as insects, diseases, animal damaging, or weather conditions like lightening and flooding.

One of the main significant characteristics in studying any tree growth model is called taper, which is a decreasing relationship between providing tree diameter and its correspondence of height, thusly estimating the volume and total tree height (VanderSchaaf, 2008).

Taper equation is used to illustrate tree volume at any predefined diameter and height by giving precise unbiased predictions; usually, volume equations make negative predictions for larger trees (Weiskittle, Hann, Kershaw, & Vanclay, 2011). In other words, taper is thinning in diameter according to increasing heights.

Tree stem taper has been studied since 100 years ago by developing taper functions to describe the connection between diameter and height (Newnham, 1992).

Some of these well-known taper equations which were proposed to study the Diameter-Height relevance are: Munro (1966), Bruce (1968), Kozak (1969), Demaerschalk (1972), Cervera (1973), Ormerod (1973), Max and Burkart (1976), Coffre (1982), and Jimenez (1994) (Rojo *et al.*, 2005).

In the wake of European settlement, Longleaf pine trees, which comprise the data for this research, covered an estimated 36 million (ha) of southeastern United States (Van Lear, Carroll, Kapeluck, & Johnson, 2005). After many years, 1.2 million ha remain, and these remaining trees spread from southeastern Virginia to eastern Texas.

In recent years, different forestry organizations started encouraging the creation of natural forests by planting Longleaf pine trees (Gonzalez-Benecke *et al.*, 2013), which will help foresters to research this type of tree and predict future aspects of tree growth based on current attributes such as age, height, site index, and diameter (Avery & Burkhart, 2002).

An abundance of qualitative data and literature reviews have been performed on Longleaf pine trees as a means of studying their stem volume, height, diameter at breast height (DBH), and diameters; however, no one has yet compared the numerical algorithms leading to autoregressive process results. The goal of this presented study is to examine the relationship between the Longleaf pine tree taper in contrast to height, age and site index (SI). This research will be assessed through analyzing various impacts of algorithms under different taper models. The algorithms discussed and implied throughout the qualitative and quantitative methodologies include Yule Walker (YW), Iterative Yule Walker (YW), Maximum Likelihood (ML), and Unconditional Least Square (ULS) algorithms. Discussing how the data is sensitive in comparison to the four

algorithms was completed by applying a set of key input variables and parameter values. To test how the data reacts to different algorithms, AIC and R^2 were the factors chosen as the guideline in building the decisions. Beside the mentioned factors, model errors (residuals) are very important benchmarks utilized to test the models performance.

5.2 Method

5.2.1 Data Description

The Longleaf data used to study the impact of diverse algorithms were recorded from seven studies and two hundred and twenty-nine persistent plots from Pineville, Louisiana and Auburn, Alabama. The trees were measured and collected by the U.S. Forest Service's Laboratory at Pineville, LA to investigate the consequences of spacing and thinning Longleaf pine tree which spread throughout the Western Gulf Coastal Plains, from Santa Rosa County, Florida to Jasper County, Texas (Leduc & Goelz, 2009).

The trees were parceled into plots, representing a group located close to one another in one area. Tree plotting and taper models are two significant factors in assessing the impact of algorithms to solve the autoregressive process.

Tree height and stem diameter at 1.37 meters (m) in height, also classified as DBH, are very important tree characteristics for helping any researcher study and understand tree behavior, and predicting forest stock (Gonzalez-Benecke *et al.*, 2013).

Twenty-five thousand, six hundred and fifty-nine trees were distributed (without human intervention) in southern Alabama, Mississippi, Louisiana, and eastern Texas. These trees were repeatedly measured as follows: diameter for the outside bark at 1.37 m height (DBH, cm) was measured to the nearest 0.254 cm, the total height (TH, m) was measured to the nearest 0.328 m, and lastly the specific tree stem outer bark diameter (D)

to height (H) at 5.08 cm were measured. Those tree measurements were collected at different ages for each targeted tree. The Longleaf Pine trees were divided into two categories: fertilized which is represented by type 1 and unfertilized which is represented by type zero. Fertilized tree type one was removed and only unfertilized trees were selected for this study. One observation was excluded from the study since it has the total height of the tree less than a height relative to specific diameter of the tree, which is impossible to happen (it could be a human entry mistake). Factor F was created as $F = TH / DBH$ ($m.cm^{-1}$) and used as an indicator to remove the damaged trees. In addition, when $F < 0.54 (m.cm^{-1})$ or $F > 13.5 (m.cm^{-1})$, related trees are ineligible for inclusion in the data investigation. As a result of the previous step, 13 plots were eliminated out of 115, causing elimination for study 4.10 (which is the T.R. Miller plots from Auburn). Furthermore, 91 plots were found to be elite out of the initial 225 plots.

Study 2.29 and 4.10 were removed since they consist of the plots which were excluded by the factor F. After study 2.29 and 4.10 were eliminated, five studies remained inclusive: 2.03, 3.02, 3.12, 3.13, and 3.29 were all from Pineville. These five studies are defined as the following: study 3.02 represents the burned tree portion of this study, including four planting spacings and four thinning treatments, with study plots located in J.K. Johnson Forest (about 17 miles southeast of Alexandria, Louisiana). Study 2.03 is the unburned tree portion of the study with four pruning treatments and this study poses some problems for trees at specific ages. Study 2.29 has three basal areas and two fertilizer treatments, in addition to study 3.12 having seven thinning treatments with plots located on a plantation in Perry County, Mississippi (about 5 miles northwest of Richton). Study 3.13 also has seven thinning treatments and its plots located in Texas in

the Angelina district. Finally, study 3.29 plots are located in Jasper and Newton counties in Texas and in Beauregard, Rapides and Sabine Parishes of Louisiana.

Table 5-1 shows the number of observations in each plot, revealing the lowest number of observations being 58 in plot 447 and the maximum was 1701 in plot 14.

Table 5-1: Observations in 91 plots of Longleaf pine trees.

Plot	Obs.	Plot	Obs.	Plot	Obs.	Plot	Obs.	Plot	Obs.	Plot	Obs.	Plot	Obs.
1	934	**14**	1701	**27**	677	**40**	546	**53**	139	**146**	203	**338**	212
2	1116	**15**	914	**28**	539	**41**	1570	**54**	143	**148**	221	**344**	123
3	960	**16**	945	**29**	645	**42**	1657	**55**	146	**232**	148	**345**	195
4	1001	**17**	954	**30**	564	**43**	1060	**56**	90	**233**	179	**346**	143
5	1036	**18**	919	**31**	1702	**44**	1296	**57**	217	**237**	151	**347**	133
6	940	**19**	960	**32**	1376	**45**	198	**69**	154	**238**	164	**431**	209
7	953	**20**	841	**33**	1638	**46**	223	**70**	143	**241**	178	**433**	113
8	931	**21**	1762	**34**	1495	**47**	141	**131**	114	**244**	216	**436**	84
9	883	**22**	1440	**35**	565	**48**	129	**132**	211	**245**	89	**438**	217
10	814	**23**	1453	**36**	461	**49**	174	**135**	183	**246**	81	**442**	167
11	1956	**24**	1491	**37**	532	**50**	151	**137**	180	**332**	131	**444**	165
12	1882	**25**	631	**38**	427	**51**	167	**143**	149	**332**	192	**445**	240
13	1668	**26**	627	**39**	560	**52**	127	**144**	106	**333**	250	**447**	58

Site index (SI, m) is defined as the total height (TH) at age 50, predetermined for 48 plots, but is not straight for the other 43, therefore being generated by using the following Schumacher equation: $SI = 1.3 + a.\exp(b - \frac{c}{50})$. The parameters a, b, and c were chosen to be 1, 25, 10, respectively, and the equation was solved by using Jacobian and Hessian matrix. The final outcomes yield 91 plots, 6 studies, age, height, year of measurement, tree number, states, and counties of three locations totaling a population size of 45,675. Standard statistical characteristics of the fitting dataset are shown in

Table 5-2. Since trees are grouped by plots and every tree measured repeatedly at different ages, taper models of Longleaf Pine data are considered as mixed models.

Table 5-2: Summary of Standard Statistical characteristics for 91 plots of Longleaf.

	Taper equations dataset N=45675			
Variable	Mean	SD	Min	Max
Age	39	8.649407	16	50
DBH	26.162	8.269708	2.032	50.038
TH	15.3924	3.76244	1.8288	30.7848
H	13.2588	6.557018	0.3048	27.7368
SI	26.29922	1.486754	22.87372	30.7848
D	5.08	8.791802	5.08	50.8

The distribution of total observations (N = 54855 h^{-1}) were used for taper model fittings and model evaluations by age and site index. The final 91 Longleaf Pine plots were the following: 5.49% of plots had N < 100 h^{-1}, at the same time 20.88% of plots had N > 1000 h^{-1}, while 73.63% had observations between 100 and 1000 h^{-1}. With respect to site index, most places (60.44%) had SI between 25 and 28 m, and only 6.59% of the data had SI > 28 m, and lastly the results reflected about 32.97% had SI < 25 m. Finally, by analyzing the age findings conclude that 1.27% of data lie under age 20 years, 47.64% had ages between 20 and less than 40 years, and 51.09% of the data lie between 40 and 50 years.

The relationship between DBH and TH, and between H and D shows clearly in **Figure 5-1** that the diameter and the height have a negative relationship, for as long as the height increases the diameter decreases. In contrast, DBH and TH have a positive relationship in which trees with higher total height have wider DBH.

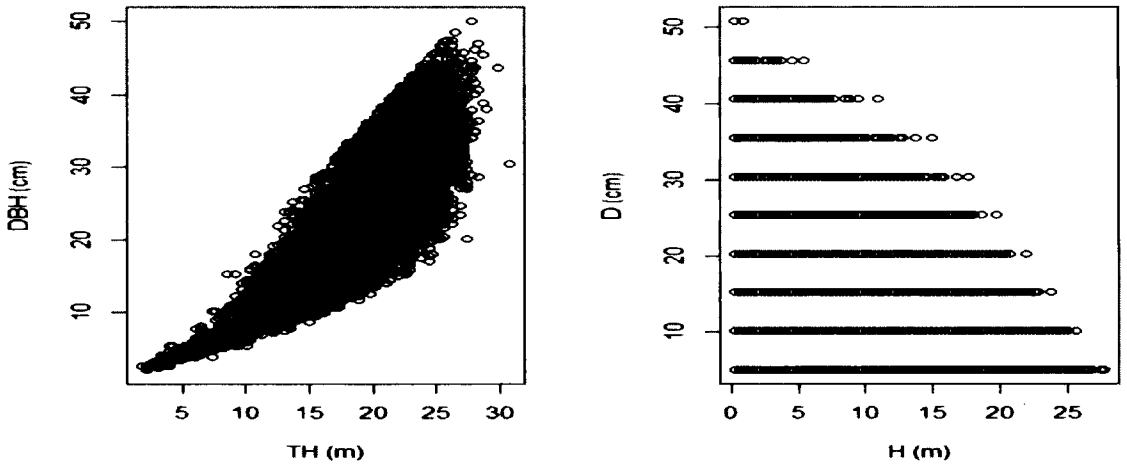


Figure 5-1: Relationships between DBH and TH and between H and D.

5.2.2 Models and Algorithms Description

To reach the goal of investigating and assessing the impact of different algorithms on Longleaf Pine data and select the best, two sets of runs were executed on the entire dataset. The reason is because the autoregressive analysis based upon the errors resulted from the taper equations.

For the first phase, I applied five famous taper models on the data (45,675 observations) to obtain parameter estimations for each model and model errors (residuals). These five taper models are referenced as: Bruce, Coffre, Jimenez, Kozak, and Munro models (Rojo, PeaStem, & Sanchez-Rodriguez, 2005).

The second phase involves the autoregressive process which, according to Brockwell and Davis (1996), is the most familiar approach in describing the time series data. Autoregressive process is applied on the achieved errors from the taper equations in the first phase: thus, these errors were differenced to remove the correlation among them since the Longleaf Pine data is a time series. The differenced errors are applied in a time series regression equation by using the autoregressive procedure, which is solved by four

different algorithms as was mentioned; one after another the four algorithms were applied to the differenced errors which resulted from each taper model. The last step entails comparing the results of each plot separately based upon Akaike Information Criterion (AIC) and Coefficient of Determination (R^2), which are the two guide methods for selecting the best and the worst models. Choosing two selection criterion methods to compare the algorithms' performance is called multivariate pairwise tests. In such a case, two vectors are used for comparisons (Yildiz, Aslan, & Alpaydin, 2011). In both phases no data splitting are used because of three reasons: (1) the model selection methods are simple and straightforward, (2) the whole dataset will be applied various times for fitting different models, and (3) according to Hirsch, "splitting a sample into two pieces cannot substitute for true attempts at replication." The data in this study has true replication many times for every tree as previously mentioned.

5.2.2.1 *Taper Models Evaluations (Phase I)*

Five models were used to estimate tree taper, and each model studies the relation between taper and age: SI, H, TH, and some models add the DBH variable of the tree. The previous variables had been chosen since they provide pertinent tree feature information such as productivity, which has a very important impact on the taper. For all trees, X was calculated by dividing the difference between the total height and the height over the difference of the total height and 1.3716, which is the breast height of a tree, $X = \frac{TH - H}{TH - 1.3717}$. T is representative of the taper and defined by the squared of diameter divided by DBH, $T = (\frac{D}{DBH})^2$. Age, SI, DBH, TH, and H are given variables (already measured for each tree).

Bruce was the first applied taper model reflected as :

$$T = a_1 \cdot X^{1.5} + a_2 \cdot (X^{1.5} - X^3) \cdot DBH + a_3 \cdot (X^{1.5} - X^3) \cdot TH + a_4 (X^{1.5} - X^{32}) \cdot TH \cdot DBH + a_5 \cdot (X^{1.5} - X^{32}) \cdot TH^{0.5} + a_6 \cdot (X^{1.5} - X^{40}) \cdot TH^2 + a_7 \cdot e^{(age^{-1})} \cdot SI + \varepsilon_1$$

Equation 5-1

where a_1 to a_7 are curve fit parameters and ε_1 is the Bruce model error term. The second taper model used is Coffre which has the formula of:

$$T = b_1 \cdot X + b_2 \cdot X^2 + b_3 \cdot X^3 + b_4 \cdot e^{(age^{-1})} \cdot SI + \varepsilon_2 \qquad \textbf{Equation 5-2}$$

where b_1 to b_4 are curve fit parameters and ε_2 is the Coffre model error term. The

$$T = c_1 + c_2 \cdot (H/TH) + c_3 \cdot (H/TH)^2 + c_4 \cdot (H/TH)^3 + c_5 \cdot (H/TH)^4 + c_6 \cdot (H/TH)^5 + c_7 \cdot e^{(age^{-1})} \cdot SI + \varepsilon_3$$

Equation 5-3

third model used is Jimenez and fitted as the:

where c_1 to c_7 are curve fit parameters and ε_3 is the Jimenez model error term. The fourth taper model was Kozak and applied as:

$$T = d_1 \cdot ((H/TH) - 1) + d_2 \cdot ((H/TH)^2 - 1) + d_3 \cdot e^{(age^{-1})} \cdot SI + \varepsilon_4 \qquad \textbf{Equation 5-4}$$

where d_1 to d_3 are curve fit parameters and ε_4 is the Kozak model error term. The last taper model used is called Munro which has the formula of:

$$T = g_1 - g_2 \cdot (H/(TH - 1.3)) + g_3 \cdot e^{(age^{-1})} \cdot SI + \varepsilon_5 \qquad \textbf{Equation 5-5}$$

where d_1 to d_3 are curve fit parameters and ε_4 is the Kozak model error term.

The parameter estimations for each taper model were accomplished by using nonlinear regression procedure PROC NLIN, in SAS 9.3 (SAS Institute, Inc., 2012). For every model numerous initial values were used until two requirements were completed: (1) the model converges; and (2) the lowest mean square error (MSE) and coefficient of determination (R^2) were achieved (Kobayashi & Salam, 2000). MSE and R^2 were calculated by the following equations:

Mean Square Error:
$$MSE = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n-p} \qquad \textbf{Equation 5-6}$$

Coefficient of determination:
$$R^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \qquad \textbf{Equation 5-7}$$

where $Y_i, \hat{Y}_i, \bar{Y}$ are the actual, the predicted, and the average values for the dependent variable, respectively. In addition, n is the total number of observations and p is the number of parameters in each model.

By the time statistical analyses had been completed for all five taper models, individual model residuals were prepared for the second phase. Since the Longleaf Pine tree is a time series data, the error is dependent over time. In essence, errors serially correlated, which means (ε_t) are correlated with residuals in earlier periods such at ε_{t-1}. To remove the correlation between residuals, differencing was accomplished and applied at lag1 as the following equations:

$$residual_1(\varepsilon) = \varepsilon_t - \varepsilon_{t-1} \qquad \textbf{Equation 5-8}$$

The produced differenced residual ($residual_1$) of each one of the five taper models will be used in the autoregressive procedure with four different algorithms in phase II.

5.2.2.2 *Algorithms Evaluation by using Autoreg (Phase II)*

Evaluating algorithmic performance is based on applying autoregressive procedure in SAS 9.3 (SAS Institute, Inc., 2012). This procedure is used to estimate the linear regression models of a time series data. For example, the formula of the regression in this case is

$$Y_t = X_t'\beta + V_t \qquad \textbf{Equation 5-9}$$

where Y_t is the response variable, X_t' is the regressor variable with β slope, and V_t is the error term. As it was mentioned, errors are correlated and represented by the following equation:

$$V_t = -\varphi_1 V_{t-1} - \cdots - \varphi_m V_{t-m} + \varepsilon_t \qquad \textbf{Equation 5-10}$$

After considering the differenced errors of Longleaf Pine which resulted from the previous stage, and since these residuals are a consequence of uncorrelated random variables, the outcome residuals ($residual_1$) are called the white noise process (a_t) with zero mean, constant variance, and zero covariance.

Suggested by Cleveland (1972), a statistical method for analyzing the autocorrelation, inverse autocorrelation and partial autocorrelation functions were applied to determine the autoregressive model order; three produced the best lag order for the given data. By stating three to be the order of the autoregressive error process for fitting, the errors' autoregression equation was in the form of:

$$(1-\phi_3 B^3)Z_t = a_t \qquad \textbf{Equation 5-11}$$

where a_t is $residual_t$ and the autoregressive part is the residual before differencing, the final equation was applied in SAS autoreg procedure:

$$(1-\phi_3 B^3)\cdot residual_t = residual_t \qquad \textbf{Equation 5-12}$$

This stage of the research focuses on the residuals and how they are critical to the four numerical algorithms utilized for solving autoregressive procedure, helping to comprehend the sensitivity of tree taper to tree age, SI, H, TH, and DBH. Studying the impact of the algorithms with each model was processed for each plot to get detached analyses on observations in each group of trees. Moreover, increased obtainable information for each plot drove data to fit more precise models (Neter, Kutner, Nachtsheim, & Wasserman, 1996).

It is important to declare that by doing this analysis, the hypothesis, which claims that the default option algorithm under any statistical software procedure is not always the perfect choice for a predefined model, will be realized.

Reaching the point of comparing the data reacting occurs through applying four estimation methods, which are implemented for the autoregressive approach of residual models on every tree plot, and algorithms are defined as the following: Yule Walker (YW), Iterative Yule Walker (ITYW), Maximum Likelihood (ML), and Unconditional Least Square (ULS).

5.2.2.2.1 Yule Walker (YW)

This algorithm is the default option in the autoregressive procedure in SAS. In a large sample from an AR (p) process, YW algorithm is represented by **Equation 5-13**:

$$C_t = \sum_{1}^{p} \phi_p C_{t-p} + \sigma_\varepsilon^2 \delta_{t,0} \qquad \textbf{Equation 5-13}$$

where C_t is the autocovariance function of the series, ϕ is the vector of autoregressive parameters which is the estimation of β, p is the autoregressive lag order, σ_ε^2 is the standard deviation of the residual, and $\delta_{t,0}$ is the Kronecker delta function.

5.2.2.2.2 Iterative Yule Walker (ITYW)

This algorithm uses the resulted residuals that come out of YW algorithm to create new estimators of ϕ and V (the error vector).

5.2.2.2.3 Maximum Likelihood (ML)

The Maximum Likelihood algorithm is efficient but is in need of a good starting point and cannot be computed for some data. Thusly, it is maximized by minimizing the objective function:

$$|L|^{1/N} e'e|L|^{1/N} \qquad \textbf{Equation 5-14}$$

where e is the residual vector and L is the likelihood function which represented as

$$-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|V|) - \frac{S}{2\sigma^2} \qquad \textbf{Equation 5-15}$$

where σ^2 is the variance, V is variance matrix of the error vector, $|V|$ is the determinant of V, N is the number of the observations, and $S = e'.e$ (unconditional sum of squares of the model)where e is the transformed error.

5.2.2.2.4 Unconditional Least Squares (ULS)

The last algorithm that can be used to solve the autoregressive equation is the ULS, which is computed by minimizing S with respect to the parameters β and ϕ_i (SAS, Inc., 2012).

Since there are five models, four algorithms, and ninety-one plots, 1820 candidate models (5 models* 4 algorithms* 91 plots = 1928) resulted from studying the best outcome for every algorithm associated with each model based upon selection criteria. Statistical analyses were estimated for the third order of autoregressive process AR(3) of every model.

The strong implication test considered by many researchers is the best method for contrasting models (Platt, 1964). Moreover, the supreme model of any research is the remarkable model for making inference from the data. In any research, the execution of any model or algorithm is measured by choosing extreme statistics. In essence, either the minimum number of errors (e.g., lowest value MSE or AIC), or the maximum number of coefficient of determination (R^2) can be implied (Dayton, 2003; Cameron & Windmeijer, 1995). Pairwise comparison have been used by selecting the main proficient models based on the lowest AIC and the highest coefficient of determination R^2 (was mentioned in the first phase) as it is mostly used in time series and regression model analysis. AIC is a measure of the goodness of the fit and the complexity of the model, calculated as:

$$AIC = -2\ln(L) + 2k \qquad \textbf{Equation 5-16}$$

where L is the maximized value of the likelihood function for the estimated model and K is the number of estimated parameters.

5.3 Results

Models in general are roughly sensitive to the input variables and the parameter values. The measure of sensitivity differs not only between models but among different groups (plots) in the same model (Weiskittle, Hann, Kershaw, & Vanclay, 2011). Applying numerical methods to complicated nonlinear models make them possessive of

diagnostic solutions. The correct solution is unknown, but the embedded algorithms are able to find trusted results (Seppelt & Richter 2005). Accordingly, four autoregressive algorithms were applied to the taper models.

By checking time effects on five taper model residuals, **Figure 5-2** proves the Bruce model variance is constant. In other words, the residuals are unbiased and show homoscedasticity or residuals having homogeneity of variance (Bera & Jarque, 1981). Only the Bruce model was shown in **Figure 5-2** as an example of the five taper models since the case of the residuals for the remaining four models are exactly the same.

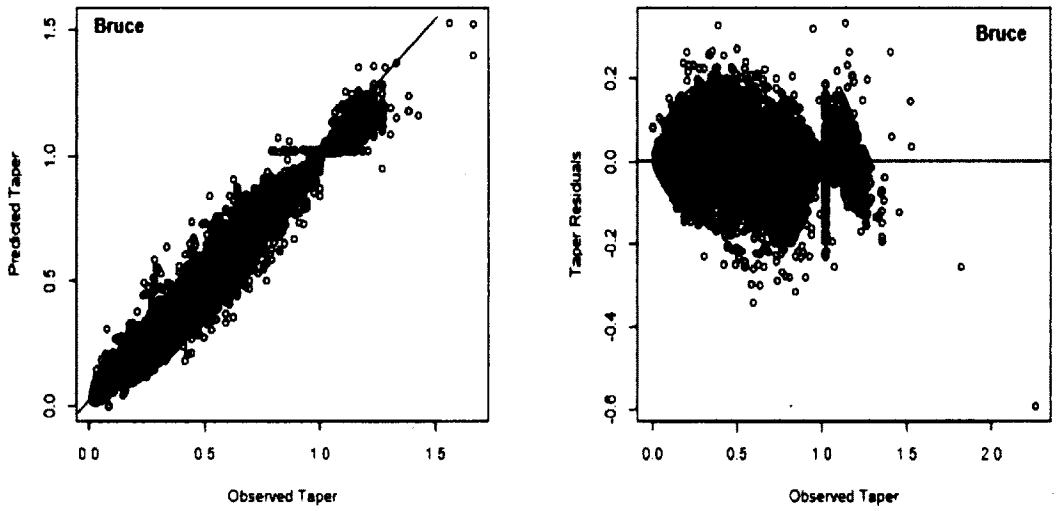


Figure 5-2: The relationship between observed and predicted values for Bruce model (left) and the time's effect on residuals of Bruce taper model (right).

Parameter estimations of five taper models related to the Longleaf Pine trees growing in Pineville located in the southern portion of the United States are recorded in **Table 5-3**. For all five model parameters, estimations were significant at $P < 0.0001$ and $n = 45675$.

Reflective of **Table 5-3** we notice that Bruce and Coffre models have the highest R^2 equaling 0.9828 and 0.9827, despite the Bruce model having DBH in the equation and

Coffre does not. However, both models have the lowest MSE among the five models. The other three models, Jimenez, Kozak, and Munro, share the same variables with different equation forms, but Kozak shows more significance than the other two, with R^2 equaling 0.9697. The entire five models share the last term which has age and SI variables. The parameter estimation for that term was negative for all equations, except in the Bruce model which was positive. As such, age and SI have negative effects on the taper in four models.

Table 5-3: Parameter estimates and fit statistics of the Longleaf Pine tree taper equations.

Model	Parameter	Parameter Estimate	SE	MSE	R^2
Bruce	a_1	1.0349	0.000959	0.00603	0.9827
	a_2	-0.00451	0.000393		
	a_3	0.0119	0.000540		
	a_4	1.42E-12	6.9E-14		
	a_5	-542E-14	2.31E-13		
	a_6	1.44E-17	6.44E-19		
	a_7	0.000824	0.000026		
Coffre	b_1	0.7514	0.00630	0.00602	0.9828
	b_2	0.2834	0.00860		
	b_3	0.0546	0.00357		
	b_4	-0.00120	0.000044		
Jimenez	c_1	1.3165	0.00821	0.0102	0.9257
	c_2	-3.0040	0.0624		
	c_3	6.2952	0.4092		
	c_4	-8.7238	1.0827		
	c_5	4.3748	1.2439		
	c_6	-0.1751	0.5168		
	c_7	-0.00282	0.000289		
Kozak	d_1	-1.5649	0.00631	0.0106	0.9697
	d_2	0.3578	0.00637		
	d_3	-0.00157	0.000052		
Munro	g_1	1.2046	0.00853	0.0120	0.9130
	g_2	1.1413	0.00150		
	g_3	-0.00336	0.000313		

The combination of the differenced residuals result from the five taper equations in autoregressive procedure, four numerical algorithms, and ninety-one plots led to generating 1820 models with 364 models for every taper equation. As was mentioned, selecting the best models from the candidate of 1820 was based upon two important criterions: akaike information criterion (AIC) and coefficient of determination (R^2). Prior to starting the deep exploration for the best models, general statistical analyses were done on AIC and R^2 by using the 1820 models **Table 5-4**.

Table 5-4: Five measurements summary of AIC and R^2 statistics

	Attribute	AIC	R^2
	Mean	-1532.24	0.350231
	Stand Deviation	1438.844	0.078243
Quartiles	**0% Min**	-5403.43250	0.18650
	25% Q1	-2249.16885	0.29790
	50% Median	-734.53468	0.33550
	75% Q3	-442.28818	0.38675
	100% Max	-48.04991	0.79340

As it is well known the low AIC and the height R^2 values are the best in the selection criteria properties (Dayton, 2003). By observing **Figure 5-3** the AIC values for every model, algorithm, and for the 91 plots, it clearly appears the lowest AIC values are in plots 11 to 44. In addition, the plots over 44 have higher AIC values which make the models less significant in these plots. If we consider only AIC as a selection criterion for the best model, the most significant models for describing the strong relationship between the taper and the other variables would have been in plots 11, 14, and 31. At the same time, plots from 1 to 45 are more significant than the other plots.

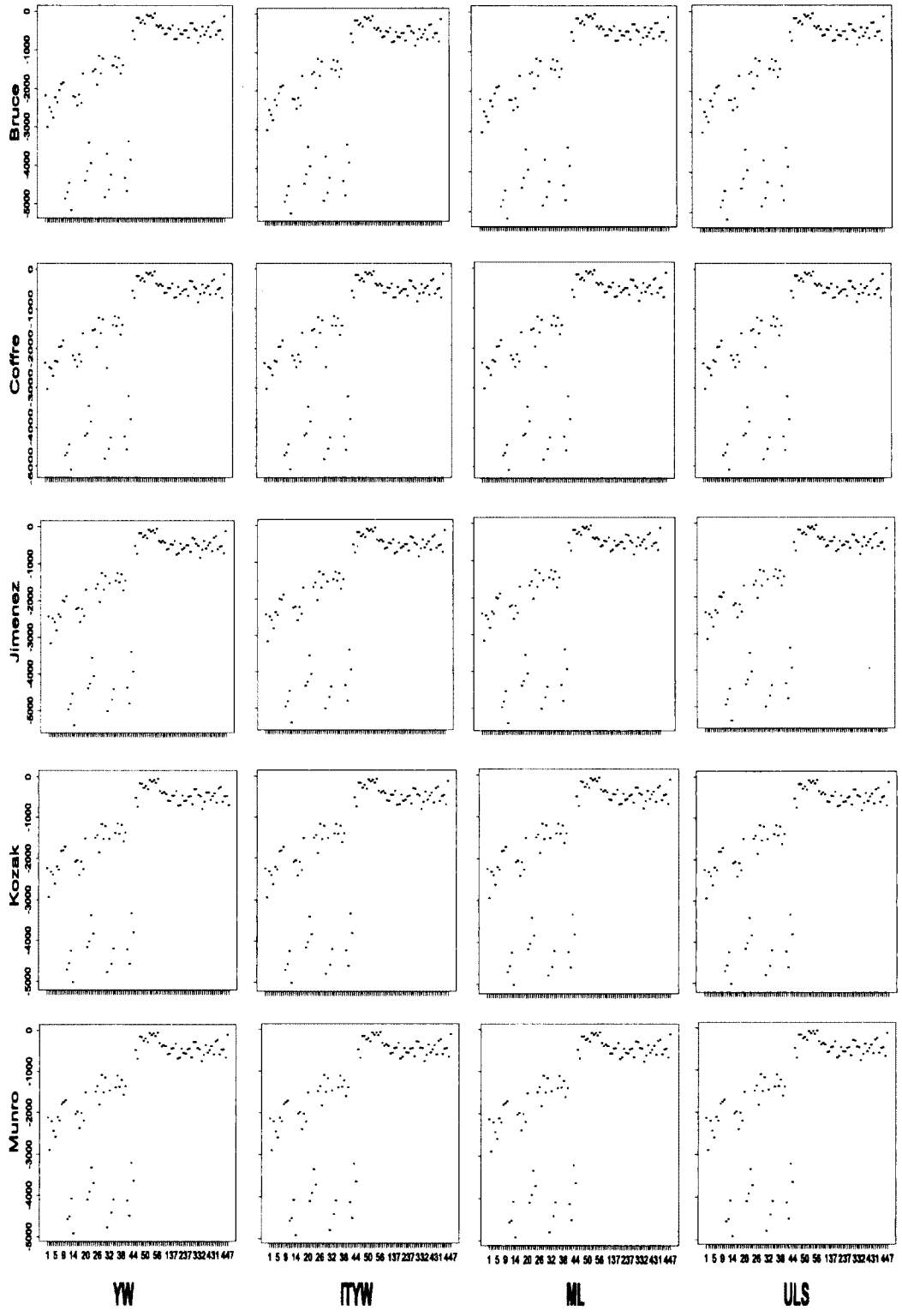


Figure 5-3: AIC values for every model, algorithm, and 91 plots from 1820 models. Models from top to bottom are: Bruce, Coffre, Jimenez, Kozak, and Munro. Algorithms from left to right are: YW, ITYW, ML, and ULS.

Figure 5-4 shows R^2 possess higher values in plots 447, 56, 53, 52, 51, and 47, which means these plots are very significant for representing the relationship between the taper and the variables. Simultaneously, plots 436, 433,131, 39, and 36 have the lowest R^2 values making them less significant for explaining the same relationship.

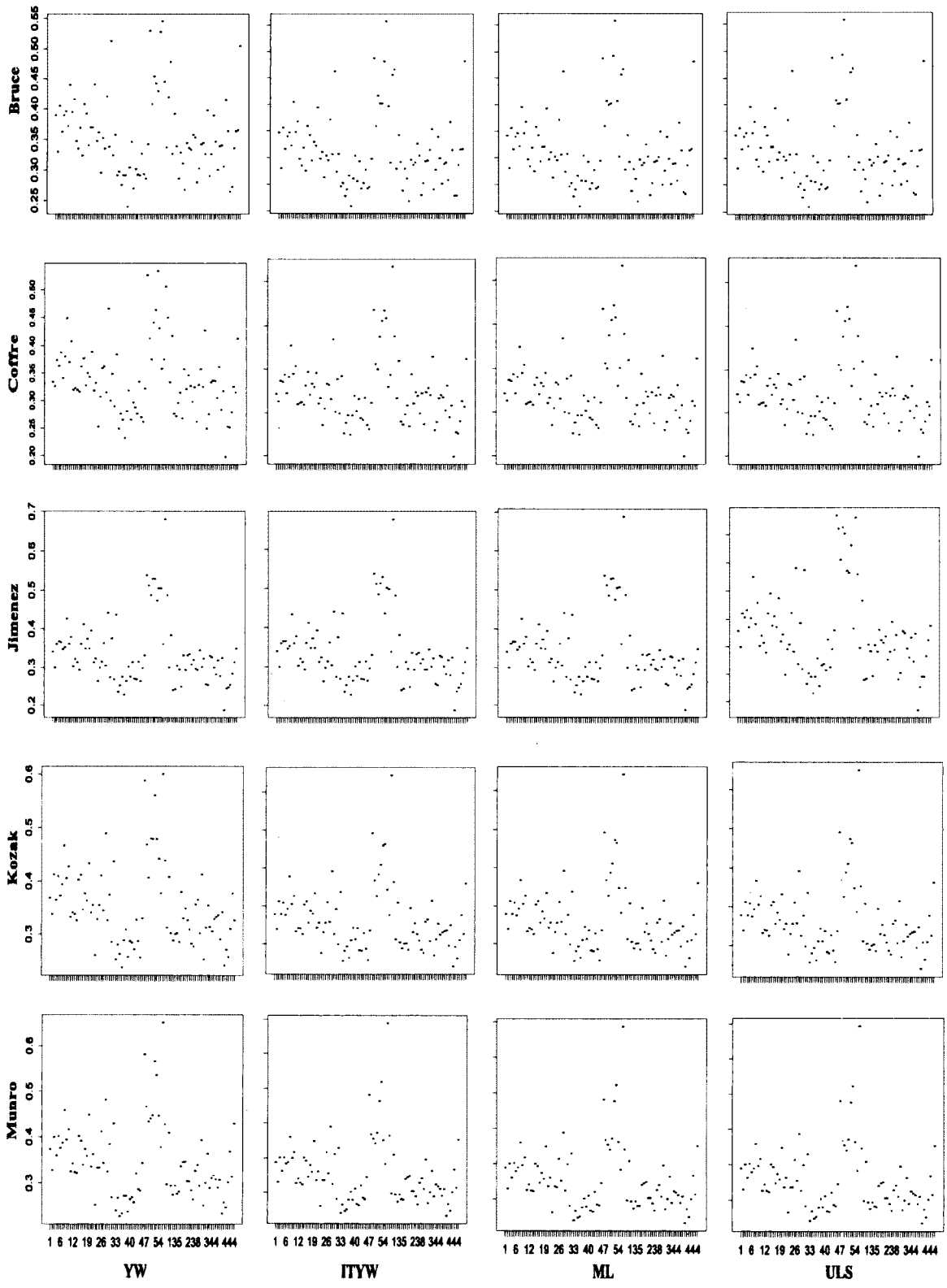


Figure 5-4: R^2 values for every model, algorithm, and 91 plots from 1820 models. Models from top to bottom are: Bruce, Coffre, Jimenez, Kozak, and Munro. Algorithms from left to right are: YW, ITYW, ML, and ULS.

By comparing the results from **Figure 5-3** and **Figure 5-4**, it is clear that the hypothesis that states model selection tasks yields different results achieved. For this reason and for satisfying the fairness in assessing error sensitivity to the algorithms' performance and behavior under every model, pairwise selection method was chosen. This selection method is inclusive of the 1820 resulted models (364 models were produced by every taper model) and were sorted by ascending order with respect to AIC and descending with respect to R^2 for each plot.

Duly mentioned, autoregressive procedure was used to solve the linear errors process in the form of $residual=(1-\phi_3 B^3)\cdot residua$ that resulted from the 1820 models. To select the best models, ranking was assigned from 1 to 1820 for every candidate model by ordering the data according to every model, algorithm, and plot. Through mimicking the strategy of Strimbu and Paun (2013), comparisons were made and the best initial models according to the fitting of the pairwise method were selected and shown in **Table 5-5.** Twenty models were counted as the best models for every taper model and algorithm. The table includes predictions for the third order of the autoregressive process AR (3), AIC, R^2, and P values for the Durbin Watson test.

Table 5-5: Autoregressive error model results for the top 20 best models according to lowest AIC and highest R^2 for every model and algorithm All the 20 models from plot 14.

Model	Algorithm	AR(3) ε_{t-3}	MSE var(ε_t)	AIC	R^2	Durbin Watson at lag 3	
						DW	Pr>DW
Jimenez	ITYW	0.0382	0.00244	-5403.4325	0.2929	1.995	0.5285
Jimenez	ULS	0.1219	0.00244	-5403.4128	0.2929	1.993	0.5381
Jimenez	ML	0.1212	0.00244	-5403.4049	0.2929	1.993	0.5372
Jimenez	YW	0.0398	0.00245	-5395.5983	0.2896	1.995	0.5252
Bruce	ITYW	0.0398	0.0028	-5166.2851	0.3253	1.991	0.5527
Bruce	ULS	0.07	0.0028	-5165.3009	0.3249	1.985	0.5122
Bruce	ML	0.0695	0.0028	-5165.2733	0.3249	1.985	0.6052
Bruce	YW	0.0398	0.0028	-5163.1712	0.324	1.960	0.5618
Coffre	ITYW	0.0257	0.00294	-5085.2256	0.3179	1.996	0.5144
Coffre	ULS	0.0718	0.00294	-5084.0714	0.3174	1.987	0.5928
Coffre	ML	0.0711	0.00294	-5084.0286	0.3174	1.987	0.5921
Coffre	YW	0.0257	0.00295	-5080.2371	0.3159	1.993	0.5364
Kozak	ITYW	0.0019	0.00306	-5013.1983	0.3263	1.996	0.514
Kozak	ULS	0.0168	0.00307	-5010.6335	0.3253	1.993	0.5395
Kozak	ML	0.0166	0.00307	-5010.6224	0.3253	1.993	0.5388
Kozak	YW	0.0019	0.00307	-5009.8482	0.3253	1.999	0.4847
Munro	ITYW	-0.0303	0.00326	-4909.1543	0.3224	2.002	0.4632
Munro	ULS	0.00326	-0.0304	-4909.1533	0.3224	2.002	0.4617
Munro	ML	-0.0304	0.00326	-4909.1533	0.3224	2.002	0.4626
Munro	YW	-0.0303	0.00326	-4909.1532	0.3224	2.002	0.4637

From **Table 5-5**, all the given models come out of plot 14, which is the plot with the largest number of observations as shown in **Table 5-1**. Therefore, the top four models were produced by using ITYW, ULS, ML, and YW algorithms with the Jimenez taper equation. The top model for every taper equation was obtained by using ITYW algorithm.

It is clear that by fitting the Jimenez model, the best results were produced and ITYW, ULS, and ML algorithms have the same effect on the Jimenez model with very little difference with YW when they were applied. All four algorithms have the same effect on the Munro model producing the least significant results. By looking at the

comparison criteria AIC and R^2, it clearly appears that ML and ULS have the exact effects when used with Bruce and Coffre models. At the same time, ITYW algorithm is more significant than YW when used with the Bruce model. Furthermore, ULS, ML, and YW algorithms have the same effects when applied with the Kozak model, and collectively, ITYW was the most sensitive algorithm when used with the same model.

By using the Durbin Watson test (Watson, 1951) for testing serial correlation, we notice $DW > DW_{U}$ for all the plots in **Table 5-5**, that means we do not reject H_0. Stating all the serial auto correlation equals to zero means there are no serial correlation (all the serial correlation $\rho_s = 0$). The previous results were expected since we already have differenced the residuals to remove the correlations among them. We also notice that all AIC values are negative since the part -2 ln (L) in **Equation 5-16** is greater than 2 k.

Since the inference from previous analysis concludes that plots with the highest number of observations always produce the best models, analysis is done from different points of view. By choosing the top ten highest models (according to the pairwise method) for every plot, the inference is constructed based on 910 models (91 plots*10 top best models). This represents 50% of the original 1820 models, summarizing the best 910 models which describe the number of models produced for every taper model and numerical algorithms shown in **Figure 5-5**. It is clear that taper residuals were less sensitive to all four algorithms as a result of fitting the Kozak and Munro models. Moreover, Jimenez was the superior taper model for describing the relationship and YW algorithm was the least significant when used with it. Even though Yule Walker (YW) is the default option, this algorithm was not a good choice for some models like Jimenez,

Bruce, and Coffre; hence, Unconditional Least Square (ULS) algorithm performs better with them.

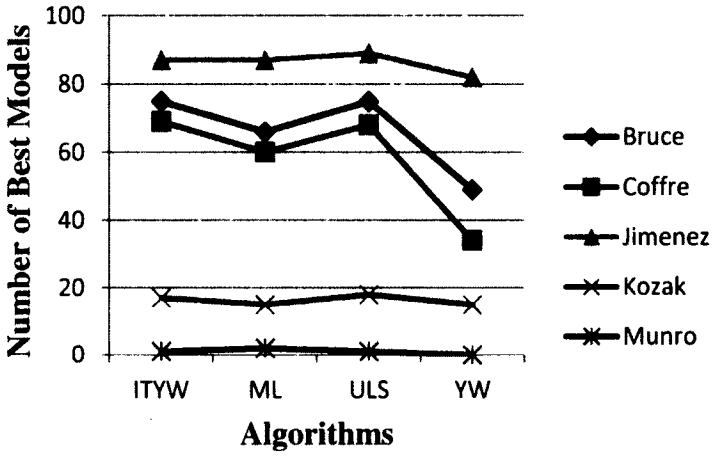


Figure 5-5: Number of best models produced by every taper model and numerical algorithm, it is summarizing for 910 best models

Despite which taper model or tree plot was used, the four algorithms had an undistinguished impact. For each taper model, diverse results according to autoregressive process were obtained as the following: 37.91% of the 910 best models were produced by fitting the Jimenez taper model which was the highest percentage, 29.12% were produced by the Bruce model, 25.38% by the Coffre model, 7.14% by the Kozak model, and finally the lowest number of good models were produced by fitting the Munro model with 0.44%. On the other hand, if the number of the good produced models were counted according to the algorithms' effect, the following is found: 27.36% of the 910 effective impact models were produced by applying ITYW algorithm, 25.27% by ML, 27.58% by ULS, and finally 19.78% were produced by applying YW algorithm. While the four algorithms seem to have the same number of best models when applied with the Munro model (with 1% and lower), ITYW and ULS algorithms have the highest effects with the

Jimenez model by 9.56 and 9.78% for each one, respectively. The number of best models produced by using Bruce and Coffre estimated 256 in comparison to Coffre at 231.

According to the selection criteria, the lowest models produced were in plots 49 and 47. These two plots contain trees from study 329, and the trees in these plots were measured at three age levels, which are 16 years, 21 years, and 26 years during the periods of 1973, 1978, and 1983.

The most effective way to assess any model performance is through residuals by plotting the prediction errors for all trees' grouping and their variables (Weiskittle, Hann, Kershaw, & Vanclay, 2011). Residuals plots are extensively used to discover a model shortage in regression analysis (Anscombe, 1961; Chatterjee, & Hadi, 1980). In addition, graphical analyses of residuals are very useful in investigating the goodness of fit for any regression model (Montgomery, Peck, & Vining, 2006). **Figure 5-6** shows how the differenced residuals of Jimenez model are normally distributed.

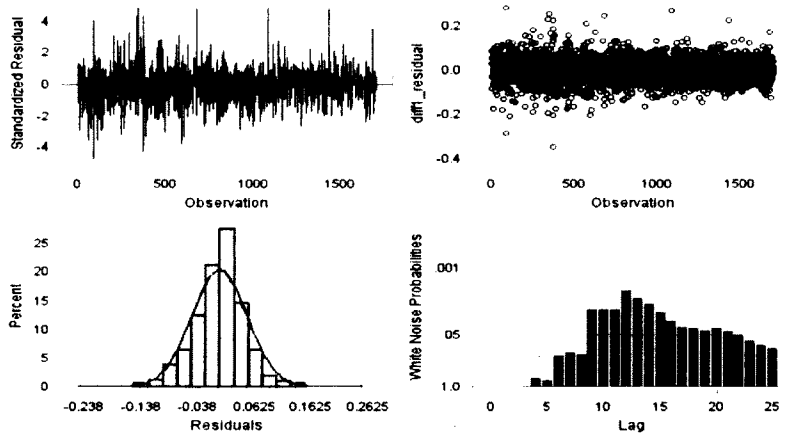


Figure 5-6: Complete information about plot 14 that resulted from Jimenez with ITYW algorithm which is the best produced model. The studentize residuals for the models with the nlag = 3, standardized, and histogram of the residuals and plots the white noise probabilities.

Figure 5-7 shows Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and Inverse Autocorrelation Function (IACF) which are three important functions for any time series model. ACF is a bar chart used to test the correlation between the time series coefficients and time lag. If any pattern is found, this indicates a correlation (dependency) exists. PACF is a bar chart reflecting the amount of correlation between variables and the time lag. IACF is the reverse of the calculated ACF of any time series. By investigating the partial autocorrelation function (PACF) plot in **Figure 5-7**. All partial autocorrelations are within the 5% significance error limits curves and there is no pattern observed, which means AR(3) model was the suitable one to fit the data.

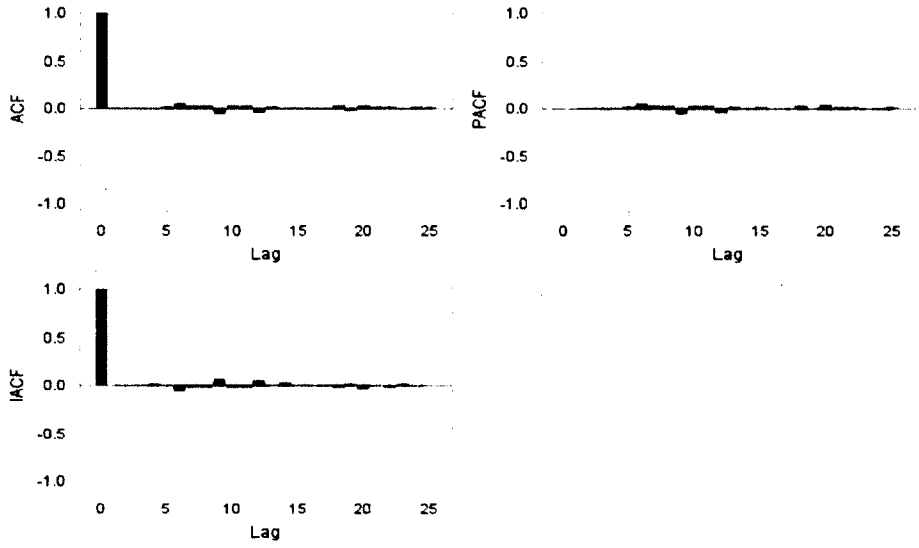


Figure 5-7: Autocorrelation, inverse autocorrelation, and partial autocorrelation function for the residuals for plot 14 that resulted from Jimenez with ITYW algorithm which is the best produced model.

Figure 5-6 and **Figure 5-7** show the situation of the residuals for plot 14 that resulted from Jimenez with ITYW algorithm which describes the general look for the best produced model in **Table 5-5**. It is obvious from **Figure 5-6** that the residuals are

independent and identically normally distributed and the white noise probabilities' plot point out that the resulting residuals are white noise.

5.4 Discussion

Taper is one of the main significant characteristics in studying any tree growth model. The first phase of this study reported about five taper prediction equations of Longleaf Pine trees. These taper models provided a useful explanation about the relationship between the taper and height, DBH, age, and SI. Age and site index can have major participation on the precision of any model's estimation (Weiskittle, Hann, Kershaw, & Vanclay, 2011). In this study, SI was set to be a height at age 50, and age had an affirmative impact on the tree taper only in the Bruce model. The DBH existence in the Bruce model can be the most proficient cause, since it was not present in the other four models. That means four out of five taper models show how the age increasing affects the tree taper in the opposite way. The Kozak taper model was the only model with negative intercept among the five equations. According to the mean square error (MSE) and the coefficient of determination (R^2), the Jimenez model behaved the best when fitted on the data to predict tree taper. The Bruce and Coffre models have close results, as the reason could be the existence of variable X in those two models and with its absence throughout the remaining three models.

The second phase was a general approach for algorithm comparisons on one dataset which described tree taper. The mentioned numerical algorithms are used to solve autoregressive process of the residuals as a resulted of fitting the five taper models. The four algorithms behaved in a different way when they were applied with different taper models for each plot of trees. The best results according to AIC and R^2 selection criteria

were for plot 14, which has the highest number of observations. Even though the Yule Walker (YW) algorithm is the default option when using the autoregressive procedure, the Iterative Yule Walker (ITYW) algorithm behaved better. ML and ULS algorithms behave the same when applied with Bruce and Coffre models in plot 14, but ITYW was the best with all five models. Kozak and Munro were the least significant taper models according to AIC and R^2. The Durbin Watson test was applied on the residuals and the results showed no autocorrelation between them. According to the models, the highest percentage of the best models was 37.91% produced by applying the Jimenez model. In lieu of previous findings, the lowest percentage of the most favorable models was produced by the Munro model at 0.44%. According to algorithms, the highest percentages of the best models were 27.36% and 27.58% produced by ITYW and ULS algorithms, respectively; the lowest was 19.78% by the YW algorithm. Selection criteria played a significant role for producing the final results, and it was obvious in this study that the model selection task gives different results when used separately. The result was based on combining both conditions of AIC and R^2 to select the best model and show how the residuals from Longleaf Pine tree data are affected by selecting different algorithms.

5.5 Conclusion

Using the mixed models helps the researcher to consider all factors that contributes to effecting data. Tree diameter-height (taper) relationships have been widely discussed in forest growth and yield through modeling them using nonlinear equations. The best model assessment and selection are based upon statistical characteristics. In this paper, selection criterion measurements were AIC and R^2. Furthermore, the judgment for

the best model or algorithm should be based upon more than one test or investigation. More significantly, several numerical algorithms when associated with a model can behave better and have a higher impact than other algorithms associated with the same model when fitted on the same dataset. To illustrate the approach of this study, 1820 candidate models were used to describe tree taper. Three hundred and sixty-four models for every taper equation were assessed to identify the potential impact of different variables on tree taper within 91 tree plots. Two taper models discussed in this study did not perform as expected when applied with the four numerical algorithms. They could be replaced by other taper models in future work. Longleaf Pine trees performed well in delivering the idea of this study and future researchers could repeat this research while taking into account the fertilized trees and the study type as variables in their taper model, which could create more important results. All the analyses and results were obtained based on Longleaf Pine trees' data. How models and algorithms will perform when used with other data is uncertain to be the same. In essence, the basic idea of this study could be applied on any time series data.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

Sensitivity of mixed models focuses on numerical algorithm comparisons used to solve the time series models. Three different projects were used to test the methodology of this research by using different mathematical and statistical models, depending on the project type. Two of the most popular time series models are ARIMA and the autoregressive models (Brockwell & Davis, 1996; Wei, 2006). The previous two methodologies have advantages in solving any time series, and provide accurate results for the relationship between data variables. Each dataset had relations between different variables of that environment. The important consideration in any type of comparison is the selection criteria the decision will be based upon, and that decision investigator takes leads to make strong inference about the data. In any research, the performance measurement of the algorithm, which is computed by the selection criteria, is represented by selecting excessive statistics, either the minimum or the maximum. The analyses of each data were based on residuals which resulted from the parametric models, and these analyses are very useful in investigating the goodness of the fit for any regression model (Montgomery, Peck, & Vining, 2006). The most significant discovery in this research is using whatever numerical algorithm is embedded in a statistical software procedure to

solve the time series models leading to the results and solutions; unless the user is performing manual comparisons of these algorithms, there will be no notifications for the best algorithm under the investigation model. According to previous information, applying scientists should be aware of the numerical algorithms' differences when using the built-in software procedure, and not dependent on the default option algorithm of that procedure, which sometimes leads to contradicting results, of what has been predicted which causes reversal of conclusions. In conjunction with frequently proven research, the default option algorithm is not the best choice for solving the model, and sometimes proves it is the worst algorithm to select and solve the model. Three datasets were used in this thesis, as a means of illustrating the idea of algorithm comparisons and their behavior evaluation.

The thesis results supported the decision of software developer of having the least complex algorithm as default in solving autocorrelated models, as on average the algorithms are not different. However, each analysis is an individual case; therefore a data analyst should be aware that the section of the algorithm could lead to wrong results.

In Chapter 3, Ozone data which was suggested by Box and Tiao (1957), was used to test seven different algorithms, three for the ARIMA model and four for the autoregressive model. The first assessment was for the ARIMA model with three numerical algorithms: CLS, ML, and ULS. The previous thee algorithms were used to solve the model and the worst result was achieved by CLS, which is the default option for that model, and at the same time, the best result was achieved by ML algorithm. By using the ARIMA model, studies possessed the capability of predicting the future year 1973, with 95% confidence interval by using the same three algorithms. The predicted

observations for every month in that year were closed by using CLS, ML, and ULS algorithms, but the squared error (SE) for ULS were the lowest, which means the best prediction was achieved by the ULS algorithm.

The second assessment for the Ozone data was achieved by using the autoregressive model with four numerical algorithms. The best result for this model was obtained by using the ML algorithm, which is also not a default option for this model. The ULS algorithm showed significant results and was close to the ML algorithm, but still not the default option which is the YW algorithm. One very important piece of information has to be mentioned about Ozone data which is, all regressors in this data are dummy variables. In addition, all of them were significant in describing the model except the winter variable. For the significant variables p value was less than 0.0001 which indicates that there is less than 0.01% chance that the observed measurements are a meaningful addition to the model, because the change in the predicted variables are related to the change in the response variable being the Ozone.

In Chapter 4, the data of this chapter differs from the Ozone data in (Chapter 3), and the Longleaf Pine data in (Chapter 5) because it was generated different manner than the other two which were collected real data. Since the basic idea of this thesis is analyzing the data based on its residuals, the noise in this data was generated by using the gamma distribution, inclusive of alpha and beta parameters which are very important factors in it; investigating the autoregressive model in this chapter was grouped by alpha and beta pairs. After generating data with two height types, Height-Age relationship was studied by using two different height models, the same two had been used to generate the data (Schumacher and Polynomial). The residuals of all models were tested through the

autoregressive model and its four algorithms. The best results were obtained by applying ML and ULS algorithms which are not the default option for the autoregressive model, and it was noticeable that the best results were always found at lower alpha and beta pairs such as (1,1). For Polynomial model and Polynomial height, 15% of the effective impact was produced by applying the ML algorithm, 10.5% by the ITYW, 14.5% by the ULS, and finally 10% was produced by applying YW. At that time, for Polynomial model and Schumacher height, ULS and ML were the most effective algorithms with 14.5% and 13.55, respectively. The least significance algorithm this time is ITYW with 9%. Additionally, for Schumacher model and Polynomial height ULS and ITYW were the most effective algorithms by 15% and 14.5 % respectively. The least significance algorithm was ML with 9%. Finally for Schumacher model and Schumacher height, ML was the worst algorithms with have no number of best models at all, and the most significance algorithm was ITYW with 27.5% of the best models achieved by using it.

The same generated data can be used with different models such as Larsen and Hann (1987) or Curtis (1967), comparing their results with different autoregressive algorithms. The investigation was based on models consisting of height as a predictor with age and site index as regressors, so that this data could be used repeatedly with the same strategy, which includes adding more regressors such as thinning or fertilization as dummy variables.

In Chapter 5, Longleaf Pine trees were used to study the taper models, under the ideology of numerical algorithms comparisons. This data was gathered from southeastern Virginia to eastern Texas. Five famous taper nonlinear models were used to test the relationship between age, height, total height, dbh, and site index with Height-Diameter

(taper). Three of the independent variables showed very significant results, and even the best numerical algorithm performances were under these three models which are Jimenez, Bruce, and Coffre. The ULS algorithm behaved the best with the Jimenez model and was reflective with the plots with the most number of observations, yielding more significant models produced. The Jimenez model was the best with 37.91% of producing the best models out of the 910 candidate models. This dataset was inclusive of a factor created to remove damaged trees, and it was later discovered that the damaged trees are in plots the model could not converge. Besides the ULS algorithm, ITYW showed very important results with 27.36% of the 910 candidate models which were produced. Furthermore, Age and site index were very important variables in describing the taper model, and it has been said that these variables can have major participation on the precision in any growth model estimation (Weiskittle, Hann, Kershaw, & Vanclay, 2011). Akaike Information Criteria was negative for all models, because the lack of fit component was very large compared to the number of parameters. In this dataset, we excluded the fertilized trees and omitted the study type factor, which can possibly be used in future studies, making important changes in the results. Numerous researches have been conducted on this data, but not one study reflects the residuals' reactions towards models.

6.2 Status of the Hypotheses and the Research Significance

The present research is based on two assumptions:

1- The future is unknown, but it is predictable.

2- Residuals of any Time Series data, has at least one organized pattern.

The previous two assumptions are supplied as a foundation for testing the following two hypotheses in studying the time series reaction to different algorithms:

1- The default option algorithm under any statistical software procedure is not always the perfect choice for a predefined model.

2- Model Selection tasks produce different results for the best models among a set of candidate models, depending on the selection criteria chosen.

During three chapters of this research (Chapters 3, 4, 5), assumptions and hypotheses stated at the beginning of this thesis have been proven.

The first assumption was investigated by using Ozone data in Chapter 3. When the Ozone from 1973 was not given it was supposed to be a future according to the dataset which ended in 1972. By using the ARIMA model, the Ozone from 1973 was predicted, proving the assumption that states the future is unknown, but it is however predictable. The second assumption clearly appeared in the Ozone data where there was a downward trend obviously shown, as well as in Height-Age and taper data where an increasing pattern was observed.

With respect to the hypotheses, the first was proving through the three datasets, that the default option was not a good choice for any predefined model in obtaining the best results. The second hypothesis was about model selection tasks, which was proven by using AIC and R^2 methods, both of them were important in our research. When performed separately, they yielded different results. Accordingly, we depended on a method called multivariate pairwise tests to compare algorithm performances (Yildiz, Aslan, & Alpaydin, 2011) by combining them (AIC and R^2) to get the best model.

6.3 Future Work

Future work is suggested based on the investigation of this research which was proposed independently in each chapter. One suggestion in conducting future research is to complete analyses using another time series procedures, while comparing embedded numerical algorithms results. The other suggestion is expanding the models of each data explored by including variables ignored in our study and examining their effects on the model, or how the dependent variable reacts to their changes. Furthermore, studying added variables contribute in providing the best model by observing changes when the numerical algorithm of the procedure is altered.

REFERENCES

1. Akaike, H. (1974). A new look at the statistical model identification. Institute of Statistical Mathematics 19:716-23.
2. Anscombe, F. J. (1961). Examination of residuals. Proc. Berkeley Symposium on Mathematical Statistics and Probability 1: 1-36.
3. Avery, T.R. and Burkhart, H.E. (2002). Forest Measurement. MCGraw-Hill Inc.
4. Baayan, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language 59(4): 390 – 412.
5. Bartlett, M. S. (1978). An Introduction to Stochastic Processes. Cambridge University Press, Cambridge.
6. Bera, A. and Jarque, C. (1981). Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals: Monte Carlo evidence. Econometric Letters 7: 313–318.
7. Box, G. E. P. & Tiao, G. C. (1975). Intervention Analysis with Application to Economic and Environmental Problems. Journal of American Statistical Association 70: 70-79.
8. Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. (1994). Time Series Analysis: Forecasting and Control. Prentice Hall, Englewood Cliffs NJ.

9. Bozdogan, H. (2000). Akaike's Information Criterion and Recent Developments in Information Complexity. Journal of Mathematical Psychology 44: 62-91.

10. Brockwell, P. J. & Davis, R. A. (1996). Introduction to time series and forecasting. New York: Springer.

11. Burnham, K. P. & Anderson, D. R. (2004). Multimodel inference Understanding AIC and BIC in Model Selecting. Sociological Methods Research 33: 261-304.

12. Cameron, Colin A., and Windmeijer, Frank A.G., (1995). R-Squared Measures for Count Data Regression Models With Applications to Health Care Utilization. Journal of Business and Economic Statistics 14: 209-220.

13. Chapman, D.G. (1961). Statistical problems in population dynamics. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California. Press, Berkeley, CA, pp. 153–168.

14. Chatterjee, S. and Hadi, A. S. (1988). Sensitivity Analysis in Linear Regression. John Wiley& Sons 5(1): 93-96.

15. Chuvieco, E., P. Englefield, A. P. Trishchenko, and Y. Luo. (2008). Generation of long time series of burn area maps of the boreal forest from NOAA–AVHRR composite data. Remote Sensing of Environment 112:2381-2396.

16. Cleveland, W.S. (1972). Inverse Autocorrelations of A Time Series and Their applications. Technometrics 14(2): 277-293.

17. Coffin, M, & Saltzman, M. (2000). Statistical Analysis of Computational tests of Algorithms and Heuristics. INFORMS Journal on Computing 12: 24- 44.

18. Cormen, T. H., Leiserson, C. H., Rivest, R. L., & Stein, C. (2009). Introduction to algorithms. Cambridge, Ma: MIT Press.

19. Curtis, R. O. (1967). Height-diameter and height-diameter-age equations for second-growth Douglas-fir. Forest Science 13, 365-375.

20. Dayton, M. C. (2003). Model Comparisons Using Information Measures. Journal of Modern Applied Statistics Methods 2(2): 281-292.

21. Demsar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research 7:1-30.

22. Draper, N. R., & Smith, H. (1998), Applied Regression Analysis, Third Edition, New York: John Wiley& Sons, Inc.

23. Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. Biometrika 38:159–178.

24. Farrar, R. M. (1981). A Site-Index Function for Naturally Regenerated Longleaf Pine in the East Gulf Area. Southern Journal of Applied Forestry **5**:150-153.

25. Fennessey, J. & D'Amico, R. (1980). Collinearity, Ridge Regression, and Investigator Judgment. Sociological Methods Research, 8, 309-340.

26. Gelman, A. (2005). Analysis Of Variance-Why It Is More Important Than Ever. The Annals of Statistics 33: 1–53

27. Gentleman, R. and R. Ihaka. (2014). R. University of Auckland, New Zealand, Auckland, New Zealand.

28. Gonzalez-Benecke, C. A., Gezan, S. A., Martin, T. A., Cropper Jr., W. P., Samuelson, L. J., and Leduc, D. J. (2013). Individual Tree Diameter, Height, and Volume Functions for Longleaf Pine. Society of American Foresters 60(1): 43-56.

29. Gujarati, D. N. (1995). Basic Econometris. McGraw-Hill Companies.

30. Hirsch, R. P. (1991). Validation samples. Biometrics 47(3): 1193-1194.

31. IBM Corporation. (2014). IBM SPSS Statistics. IBM, Armonk, New York.

32. Kobayashi, K., and SALAM, M.U. (2000). Comparing simulated and measured values using mean squared deviation and its components. AGRONOMY Journal 92:345–352.

33. Korzukhin, M. D., M. T. TerMikaelian, and R. G. Wagner. (1996). Process versus empirical models: Which approach for forest ecosystem management? Canadian Journal of Forest Research-Revue Canadienne de Recherche Forestiere 26:879-887.

34. Larsen, D. R., and Hann, D. W. (1987). Height-diameter equations for seventeen tree species in southwest Oregon. Research Paper 49. Oregon State University Forest Research Laboratory, Corvallis, OR.

35. Leduc, D. and Goelz, J. (2009). A Height–Diameter Curve for Longleaf Pine Plantations in the Gulf Coastal Plain. Society of American Forester 3(4): 164- 170.

36. McCullagh, P. (2002). What Is A Statistical Model. The annuals of statistics 30: 1225- 1310.

37. McGeoch , C. C. (1996). Toward an Experimental Method for Algorithm Simulation. INFORMS Journal on Computing 8(1): 1-15. Retrieved April 21, 2014, from http://pubsonline.informs.org/doi/abs/10.1287/ijoc.8.1.1.

38. Meek, C., Chickering, D. M., and Heckerman, D. (2002). Autoregressive Tree Models for Time-Series Analysis. Society of Industrial and Applied Mathematics Journal 229-244.

39. Montgomery, D. C., Peck, E. A., & Vining, G. G.(2006), Introduction To Linear Regression Analysis. New York: John Wiley & Sons, Inc.

40. Moser, E. B. "Repeated Measures Modeling With PROC MIXED." SAS Institute Inc., Cary, NC, USA. [2004], http://www2.sas.com/proceedings/sugi29/188-29.pdf. (2013). Louisiana State University, Baton Rouge, LA.

41. Neter, J., M.H. Kutner, C. J. Nachtsheim, and W. Wasserman. (1996). Applied linear statistical models. WCB McGraw-Hill.

42. Newnham, R.M. (1992). Variable-form taper functions for four Alberta tree species. Canadian Journal of Forest Research 22(2): 210-223.

43. Platt, John. R. (1964). Strong Inference. Science, 146: 347-353.

44. Pretzsch, H. (2009). Forest dynamics, growth and yield. Springer, Berlin.

45. Rojo, A. , Perales, X., Sánchez-Rodríguez, F., Álvarez-González, J. G., and Gadow, K. V. (2005). Stem Taper Functions for maritime Pine (Pinus pinaster Ait.) in Galicia (Northeastren Spain). 124(3):177-186.

46. Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. Behavioral Ecology 20 (2): 416-420.

47. Seppelt, R., & Richter, O. (2005). "It was an artifact not the result": A note on systems dynamic model development tools. Environmental Modelling & Software 20: 1543-1548.

48. Stockburger, D. W. (1996). Introductory statistics concepts, models, and applications. Monument, CO: David Stockburger.

49. SAS Institute Inc. (2012) SAS Software 9.1.3 Cary NC USA.

50. Shumway, R. H. and Stoffer, S. S. (2011). Time series Analysis and Its Applications. Springer New York Dordrecht Heidelberg London.

51. Strimbu, B. M. and Paun, M. (2013). Sensitivity of forest plan value to parameters of simulated annealing. NRC Research Press 43(1): 28-38.

52. Tabachnick, B. G. and L. S. Fidell. (2001). Using multivariate statistics. Allyn and Bacon, Needham Heights.

53. Tsai, C. L., Cai, Z., & Wu, X. (1998). The Examination of Residual Plots. Statistica Sinica, 8, 445-465.

54. Tukey, J. W. (1977). Exploratory data analysis. Reading, Ma: Addison-Wesley Pub. Co.

55. VanderSchaaf, C. (2008). Compatible Stem Taper and Total Tree Volume Equations for Loblolly Pine Plantations in Southeastern Arkansas. Journal of the Arkansas Academy of Science 62: 106.

56. Van Lear, D. H., Carroll, W.D., Kapeluck, P.R., and Johnson, R. (2005). History and restoration of the longleaf pine-grassland ecosystem: Implications for species at risk. Forest Ecology and Management 211: 150–165.

57. Wang, Y. M. (2000). Model Comparisons and Model Selections Based on Generalization Criterion Methodology. Journal of Mathematical Psychology 44: 171-189.

58. Wang, M., B. E. Borders, and D. Zhao. (2008). An empirical comparison of two subject-specific approaches to dominant heights modeling: The dummy variable method and the mixed model method. Forest Ecology and Management 255:2659-2669.

59. Wei, W. S. (2006). Time series analysis univariate and multivariate methods. Redwood City, Ca: Addison-Wesley Pub.

60. Weiskittle, A. R, Hann, D. W., Kershaw, J. A., and Vanclay, J. K. (2011). Forest Growth and Yield Modeling. John Wiley & Sons, Inc.

61. Yildiz, O., T., Aslan, O., and Alpaydin, E. (2011). Multivariate Statistical Tests for Comparing Classification Algorithms. Springer-Verlag Berlin Heidelberg 1-15.