SECOND ORDER OPTIMAL ROTATABLE DESIGN IN FOUR DIMENSIONS AND ITS APPLICATION IN OPTIMIZATION OF BIOETHANOL YIELD USING PINEAPPLE PEELS.

BY

KABUE TIMOTHY GICHUKI

A THESIS SUBMITTED TO THE SCHOOL OF SCIENCES AND AEROSPACE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN BIOSTATISTICS

MOI UNIVERSITY

2021

DECLARATION

Declaration by Candidate

This Thesis is my original work and has not been presented for a degree in any other University. No part of this thesis may be reproduced without prior permission of the author and /or Moi University.

Sign: _____ Date: _____

Kabue Timothy Gichuki

DPS/PHD/14/2013

Declaration by the Supervisors

This thesis has been submitted for examination with our approval as University Supervisors.

Sign: _____ Date: _____

Professor Joseph K. Koske

Department of Mathematics, Physics and Computing,

Moi University.

Sign: _____ Date: _____

Professor John M. Mutiso

Department of Mathematics, Physics and Computing,

Moi University.

DEDICATION

To my Wife Wanjiru and daughters Njeri and Wangari and son Kabue, friends and colleagues for their unfailing love and support which enabled me to realize this goal.

ACKNOWLEDGMENT

I thank the Almighty God for the gift of life, wisdom and strength during the entire period of this work. I am grateful to Moi University for the opportunity granted to me to undertake this course and my supervisors and mentors, Professor Joseph Koske and Professor John Mutiso, for their continuous guidance, advice and valuable comments throughout the course work, development and finalization of this Thesis. To Dr Eric Kuria for his technical support on the reaction equations and Nelson Elias Mandela for experimental set-ups and reagents preparations in the laboratory, my classmates and friends who encouraged me during that period not forgetting my late Dad who laid the foundations of education in me and finally to my family for the patience and endurance they went through during those challenging times and many hours of my physical absence from home. Finally but not least Mount Kenya University for the provision of free and limitless internet services and their well- equipped chemistry laboratories where the experiments were conducted without which it would not have been possible to accomplish this research.

ABSTRACT

An eco-friendly bio-ethanol from biomass waste is an alternative to petroleum products which are becoming scarce due to increased demand and depletion of source besides their environmental pollution effects. Most studies on optimization of process variables using Response Surface Methodology exploit Central Composite Design yet other designs exist. Optimal designs have fewer trials employed with the aim of obtaining efficient designs for fitting reduced quadratic or higher order models. Efficiency as discriminating criteria allows comparison between any design and the best design. This study sought to apply a second order optimal rotatable design in four dimensions to the ethanol production through fermentation of pineapples peels using yeast. The D-, A-, E- and T-Optimal values of the general design with their corresponding relative efficiencies were obtained. The design was used to carry out an experiment in order to determine the effects of time, pH, temperature and concentration of substrate on ethanol yield during fermentation of pineapple peels. Finally the optimum settings of time, temperature, initial pH and substrate concentration that led to optimal yield of ethanol were determined. The coded values of a design constructed by Das and Narasimham were used to obtain a design matrix X which was then used to construct a moment matrix $X^{/}$ X. The moment matrix was then utilized to determine the optimal values and the required efficiencies. The ratio of the optimal value of the general design to the corresponding optimal variance of the optimal design was used as a measure of relative efficiency of the design. A second order model was fit into experimental data in order to study the effects of the process variables, and the optimization through response surface plots and analytical method. The D-, A-, E- and T-Optimal values were found to be 0.6796529, 0.04104631, 0.002856958 and 1.135448 respectively, and the corresponding relative efficiencies were 98%,71.7%, 87.5% and 1% respectively. Normal probability plots and R-squared of 0.95 and Adjusted R-squared of 0.911 for E-optimal (most efficient with only 32 runs) design indicated the model fit the data well. An optimal yield of 12.35g/L of ethanol realized after 54.35 hours at 4.96 levels of pH, 34.67^{0}C temperature and 28.03g/L of substrate concentration translated to 0.441g of ethanol per gram of substrate (roughly 86% of the theoretical yield of 0.511g ethanol per g of substrate) which compares well with findings from similar studies. The yield by E-optimal design was slightly lower than that of general design by 0.040g/L. The design was found reliable in modeling, optimizing and studying effects of the factors to the processes of fermentation of pineapples peels for ethanol production. A comparison of these results and the result of rotatable design with four factors constructed using balanced incomplete block design when replications are more than three the number of times pairs of treatments occur together in the design is suggested in studying the effects of the four process variables on ethanol yield and optimization of the process using pineapple peels as feedstock. Study established the factor settings for optimal ethanol yield from pineapple peels. These wastes if not properly disposed can be a major source of pollution. A cheaper fuel than fossil fuel is provided while managing wastes.

TABLE OF CONTENTS

DECLARATION ii
DEDICATION iii
ACKNOWLEDGMENT iv
ABSTRACT v
TABLE OF CONTENTS vi
LIST OF TABLES ix
LIST OF FIGURES x
ABBREVIATIONS xi
CHAPTER ONE 1
INTRODUCTION 1
1.0 Overview 1
1.1 Background of the Study 1
1.1.1 Ethanol Production 1
1.1.2 Optimal Designs for Fitting Second Order Model 4
1.2 Statement of the Problem 6
1.3 Justification of the Study 6
1.4 Scope of the Study 7
1.5 Objectives of the Study 8
1.5.1 General Objective 8
1.5.2 Specific Objectives 8
1.6 Significance of the Study 9
CHAPTER TWO 10
LITERATURE REVIEW 10
2.0 Introduction 10
2.1 Optimality Criterion 10
2.1.1 D-Optimality 13
2.1.2 E-Optimality 14
2.1.3 A-Optimality 14
2.1.4 T-Optimality 14
2.2 Efficiency of a Design 15
2.3 Process Variables for Ethanol Production from Pineapple Peels 15

CHAPTER THREE **25**
METHODOLOGY **25**
3.0 Introduction 25
3.1 Optimality Criterion 25
3.1.1 D-Optimality Criterion 30
3.1.2 E-Optimality 31
3.1.3 A-Optimality 32
3.1.4 T-Optimality 32
3.2 Relative Efficiency and Optimal Designs 33
3.2.1 D-Optimal Efficiency 34
3.2.2. E −Optimal Efficiency 35
3.2.3.A −Optimal Efficiency 36
3.2.4. T −Optimal Efficiency 37
3.3 Substrate and Effects of Factors on Fermentation. 38
3.3.1 Materials and Substrate Preparations. 39
3.3.2 Experimental Set-Up for Ethanol Determination 41
3.3.3 Modelling Ethanol Yield 44
3.3.4 The Adequacy of the Fitted Model 45
3.4 Optimization of Bioethanol Yield 48
3.4.1 Response Surface Plots 49
3.4.2 Optimization Analytically 51
3.4.3 Modelling and Optimization using E-Optimal Design 52
CHAPTER FOUR **53**
RESULTS AND DISCUSSION **53**
4.0 Introduction 53
4.1 The $D-, A-, E-$ and T −Optimal Values of the General Design. 53
4.1.1 D −Optimal Value for the General Design 54
4.1.2 E −Optimal Value for the General Design 54
4.1.3 A −Optimal Value for the General Design 54
4.1.4 T −Optimal Value for the General Design 55
4.2 Relative Efficiencies of the Optimal Designs 55
4.2.1. D −Optimal Design and its Relative Efficiency 56
4.2.2. E − Optimal Design and its Efficiency 56

4.2.3 A −Optimal Design and its Efficiency 57
4.2.4 $\boldsymbol{T}$-Optimal Design and its Efficiency 58
4.3 Process Variables of Bioethanol Yield 59
4.3.1 Ethanol Content Determination 59
4.3.2 Modelling Ethanol Yield 60
4.3.3 Adequacy of the Model 61
4.4 Optimization of Ethanol Yield 66
4.4.1 Graphical Analysis 66
4.4.2 Analytical Determination of the Optimum Settings 71
4.4.3. Optimization using the $\boldsymbol{E}$ −Optimal Design 76
CHAPTER FIVE 84
SUMMARY, CONCLUSION AND RECOMMENDATIONS 84
5.0 Introduction 84
5.1 Conclusion 84
5.2 Recommendations 85
REFERENCES 87
APPENDICES 92
Appendix I: R-Code for Selected Runs 92

LIST OF TABLES

Table 2.1: Composition of Treated and Un-Treated Pineapple Peel 22
Table 2.2: Effects of Pretreatment on Reducing Sugar Yield from Pineapple Peel 22
Table 3.1: Factors Settings in Coded and Natural Levels.....44
Table 3.2:Anova table.....46
Table 4.1: Moment Matrix of the General Second Order Design.....54
Table 4.2: The Four Optimality Criteria for the General Design 55
Table 4.3: Moment Matrix for the D-Optimal Design.....56
Table 4.4: Moment Matrix for the E-Optimal Design 57
Table 4.5: Moment Matrix for A-Optimal Design.....58
Table 4.6: Experimental, Estimated and Residuals Values Ethanol.....60
Table 4.7: Call: Rsm (Formula = Y ~ So(X1, X2, X3, X4), Data = D3).....63
Table 4.8: Analysis of Variance (Anova) 65
Table 4.9: The Stationary Point of Response Surface: 71
Table 4.10: Path of Steepest Ascent: 74
Table 4.11: Regression Analysis for the E −Optimal Design 76
Table 4.12: Analysis of Variance Table 77
Table 4.13: Ridge Analysis E-Optimal.....78
Table A: The Design Matrix X for the Full Second Order Model.....98
Table B: The Matrix $X'X$ 99
Table C: The Matrix $X'X$ Inverse.....100
Table D: Monosaccharide Compositions of Pineapple Waste Cell Walls.....101
Table E:Dry Matter (% Fwt), Fiber and Soluble Sugars (% Dry Matter).....102

LIST OF FIGURES

Figure 3.1: Experimental set-up.....42
Figure 3.2: Response Surface Plot.....49
Figure 3.3: Contour Plot 50
Figure 4.1: Plots of Residuals.....62
Figure 4.2: The Contour Plots for Bioethanol Yield.....67
Figure 4.3: Images of Bioethanol Yield.....69
Figure 4.4: Response Surface Plots General Design 70
Figure 4.5: Path of Steepest Ascent Plot.....75
Figure 4.6: Images of Yield for E-Optimal Design 79
Figure 4. 7: Contour Plots from the E-Optimal 80
Figure 4.8: Response Surface Plot E-Optimal Design.....82
Figure A: Fresh Pineapple Peels Left: Right Sun Dried Peels after three Days At $250C$.....103
Figure B: Left; Milled Peels after Oven Drying: Right Dissolved and Pre-Treated Peels.....103
Figure C: Ccontours Plot Figure 4.2f Expanded 104

ABBREVIATIONS

BOD	-	Biochemical Oxygen Demand
CCD	-	Central Composite Design
CCFD	-	Central Composite Face Centred Design
COD	-	Chemical Oxygen Demand
DOE	-	Design of Experiments
EY	-	Ethanol Yield
Rpm	-	Rotations per minute.
RSM	-	Response Surface Methodology
TY	-	Theoretical Yield

CHAPTER ONE

INTRODUCTION

1.0 Overview

This chapter contains background to the study which introduces the subject area of the study and the current situation, optimal designs for fitting second order models and rotatable designs constructed using balanced incomplete block designs. Statement of the problem, justification of the study, Scope of the study, objectives of the study constituting general objective and specific objectives and the Significance of the Study.

1.1 Background of the Study

1.1.1 Ethanol Production

Ethanol is an attractive alternative fuel to fossil fuels since it is a renewable bio-based resource which is oxygenated, thereby providing the potentiality of reducing emissions in compression–ignition engines. Hansen et al., (2005), observed that the global fuel crises in the 1970s brought awareness amongst many Nations of their vulnerability to oil embargoes and shortages shifting focus on the development of alternative fuel sources with particular reference to the alcohols where a blend of 10% dry ethanol and unleaded gasoline known as E10 were commercially introduced into the US and continues to be marketed in the Midwestern states. The use of ethanol blended with diesel was a subject of research in the 1980s where studies showed that ethanol–diesel blends were technically acceptable for existing diesel engines. The relatively high cost of ethanol production at that time meant that the fuel could only be considered in cases of fuel shortages. But recently the economics have become much more favourable in ethanol production and it is able to compete with standard diesel. Consequently, there has been renewed interest in the ethanol–diesel blends

with particular emphasis on emissions reductions. According to Rattanapan et al., (2011) ethanol is easily adaptable to existing engines and it is a cleaner fuel with high octane rating than gasoline. Isaias et al., (2004) observed that ethanol is known to reduce green-house gases by between 86% to 90% .While Goettemoeller and Goettemoeller, (2007) reported that bioethanol is easily produced from cheap raw waste by products known as molasses of sugar industries from sugarcane and sugar beet. Wang et al., (2001) demonstrated that ethanol can be produced from Glycerol which is a major by-product of both soap manufacturing and biodiesel production industries by microbial fermentation and chemical syntheses. Molasses is a widely used substrate of bioethanol production since it is cheap and readily available and ready for conversion with little pre-treatments as compared to other starchy materials as all sugars are present in fermentable form Razmovski and Vucurovic, (2011). The most widely used sugar for ethanol fermentation is blackstrap molasses which contains about 35 – 40 wt.% sucrose,15 – 20 wt.% invert sugar such as glucose and fructose, and 28 – 35 wt.% of non-sugar solids Osunkoya and Okwudinka, (2011). Basically, five steps are involved in ethanol production namely grinding, cooking, fermentation, distillation, and hydration. In each step, there are several ways to improve the quantity and quality of ethanol produced. Tropea et al., (2014) records that Pineapple wastes, comprise of fruit trimmings produced in huge amounts by canning industries and markets throughout the world and that Costa Rica is the main producer and exporter, with around 110,000 acres of land under pineapples plants. They further noted that 25% of the fresh pineapples harvested in Costa Rica is processed to make added value products such as pineapple juice, jelly and canned pineapple. Nearly 75% of the fruit processed in canneries results in peeled skin, core, and crown as the end waste products, which are not utilized and generally

discharged as a waste. The dry matter content in pineapple waste is around 10%, and is composed of about 96% organic and 4% inorganic matter. These wastes have a potential for recycling to get raw materials or for conversion into useful product of higher value, or even as raw material for other industries after biological treatment Abdullah, (2007).These materials exhibit both high biochemical oxygen demand (BOD) and chemical oxygen demand (COD) values Ban-Koffi & Han, (1990)giving rise to serious pollution problems if not properly disposed of, according to Tropea et al., (2014) Pineapple wastes are rich in intracellular sugars and plant cell walls which are composed mainly of cellulose, peptic substances and hemicelluloses. Generally, to obtain high quality and high yield of ethanol in ethanol industry, selection of fermentative yeast is very important. The most well-known and commercially significant yeasts that have been primarily used for bioethanol production are the related species and strains of *Saccharomyces cerevisiae* Chandel et al., (2007a).These organisms have long been utilized to ferment the sugars of rice, wheat, barley, and corn to produce alcoholic beverages and in the baking industry. Tsuyoshi et al., (2005), observed that one yeast cell can ferment approximately its own weight of glucose per hour and that Sugars from sugar cane, sugar beets, molasses, and fruits can be converted to ethanol directly. A number of factors like incubation temperature, molasses concentration and sugar tolerance of the yeast strain used in fermentation and incubation period are known to limit the production of ethanol in quality and quantity. The use of concentrated sugar substrate is one way of obtaining high ethanol yield during fermentation process. However, high substrate concentrations are inhibitory to fermentation due to the osmotic stress. The pH of the fermentation medium significantly affects the fermentation process according to Ergun and Ferda Mutlu, (2000) and Jones et al., (1981).The purpose of this study was to investigate the

potential of transforming such residues of pineapple wastes in particular the peels into ethanol after physical pre-treatment and fermentation of the resulting simple sugars using the *Saccharomyces cerevisiae.* Optimization of process variables which are known to affect the quantity of ethanol produced during fermentation was studied using response surface methodology (RSM) based on a second order optimal rotatable design constructed using balanced incomplete block design in four dimensions to estimate the number of runs and optimum conditions for four independent variables namely incubation time, temperature, initial pH and substrate concentration of the pineapple peels using yeast.

1.1.2 Optimal Designs for Fitting Second Order Model

Optimal designs are a class of experimental designs that are optimal with respect to some statistical criterion. These designs allow parameters to be estimated without bias and with minimum variance while a non-optimal design on the other hand requires a greater number of runs to estimate parameters with the same precision as an optimal design. Thus optimal designs reduce time and costs of carrying out experiments due reduced number of runs as well as increasing the precision with which the parameters are estimated. These designs also accommodate multiple type of factors such as process mixtures and discrete factors in addition to allowing design optimization when the design space is constrained, for example, when the mathematical process space contains factor settings that are practically infeasible e.g. due safety concerns. The most popular designs for fitting second order models is the Central Composite designs (CCD) which comprise of a 2^k factorial or fractional factorial of resolution V with n_F runs, $2k$ axial or star runs and n_c center runs. The application of CCD arises through sequential experimentation when the 2^k factorial has been used to fit a first order model and it exhibits lack of fit, axial runs are added to allow the quadratic

terms to be incorporated in the model. The CCD has two parameters that must be specified, namely the distance α of the axial runs from the design center and the number of center points n_c.Since rotatability is a spherical property when the region of interest is a sphere (although it is not important to have exact rotatability to have a good design), the best choice of α from a prediction variance viewpoint for CCD is to set $\alpha = \sqrt{k}$. This design is a spherical CCD since it puts all the factorial and axial points on the surface of a sphere of radius $\sqrt{k}$. When the region of interest is cuboidal rather than spherical, a useful variation of the CCD is the face centered central composite design (CCFD) in which $\alpha = 1$. This design locates the star or axial points on the centers of the face of the cube. The design is used since only three levels of each factor is required and in practice it is difficult to change factors levels however the design is not rotatable. The face centered central composite design does not require as many center points as the CCD although sometimes more centers are required to give a reasonable estimate of experimental error Montgomery, (2005). Optimal designs are the alternative when the experimental region is irregular due to factor levels constraints or when the experimenter has prior knowledge about the process being studied which may suggest a non-standard model where some higher order terms or some interaction terms between factors may not be included in the model or even when the process factors are categorical or an unusual sample size may be of importance due to cost or time considerations. A second order rotatable design in four dimensions constructed using balanced incomplete block designs when the number of replications (r) are less than three the number of times (λ) pairs of treatments occur together in the design as put forward by Das and Narasimham, (1962) was applied.

1.2 Statement of the Problem

The need for a cleaner environment in urban areas and the high cost of petroleum products which are becoming scarce due to unbalanced relation between supply and demand besides air pollution of sources has led to the research for other fuels to replace fossil fuels. An eco-friendly bio-ethanol produced from biomass waste is one such alternative fuel that can be used in petrol engines without modification and with the current fuelling infrastructure and it is easily applicable in present day combustion engine, as mixing with gasoline Hansen et al., (2005).Combustion of ethanol results in relatively low emission of volatile organic compounds, carbon monoxide and nitrogen oxides. The emission and toxicity of ethanol are lower than those of fossil fuels such as petroleum and diesel Wyman and Hinman, (1990).Little in literature has been done in the use of second order rotatable design constructed using incomplete block designs in the optimization of process variables in almost all fields of study and their properties of optimality have not been documented yet this study has shown that this design is equally effective in modelling the effects of process variables considered as well as optimizing ethanol yield from pineapple peels.

1.3 Justification of the Study

Large scale farming of Pineapples by firms such as Kakuzi and Delmonte and small scale farming by surrounding communities in Thika and its environs results in huge quantities of Pineapple wastes produced by canning industries and market places as waste in the town posing serious environmental pollution problems and waste management challenge to the county government since these wastes exhibit both high biochemical oxygen demand (BOD) and chemical oxygen demand (COD) values which gives rise to serious pollution problems if not properly disposed. There is need to exploit these wastes fully for other good use. Choonut et al., (2014) observed that

Pineapple peel, a by-product of the pineapple processing industry, account for 29-40% (w/w) of total pineapple weight and that after pretreatment with water and heat at 100^0C for 4 hours, 36.25±2.87% of cellulose was achieved from the peels therefore there is a great need to exploit these wastes in producing bioethanol since they are rich in cellulosic and non-cellulosic sugars for a cleaner environment and alternative fuel sources due to depletion of fossil fuel sources. Data from experiments with levels of combinations of one or more factors as treatments are normally investigated to compare levels effects of the factors and also their interactions. But these investigations do not give information regarding the possible effects of the intervening levels of the factors or their combinations at treatment combinations not tried in the experiment. Hence there is need to carry out investigations with two aims: namely, determination and quantification of the relationship between the response and the settings of group experimental factors as well as finding the settings of the experimental factors that optimize response. Optimal designs are the alternative when the experimental region is irregular due to factor levels constraints or when the experimenter has prior knowledge about the process being studied which may suggest a non-standard model where some higher order terms or some interaction terms between factors may not be included in the model or even when the process factors are categorical or an unusual sample size may be of importance due to cost or time considerations.

1.4 Scope of the Study

A second order optimal rotatable design in four dimensions constructed using balanced incomplete block designs was applied in this study. The $D-$, $A-, E-$and $T-$optimal criteria for a general design were obtained. The weights corresponding to the $D-, A-, E-$ and $T-$optimal designs as put forward by Pukelsheim, (2006) were used to obtain optimal designs by varying the proportions each regression vector was

to be run. The respective $D-, A-, E-$ and $T-$optimal criteria for the general designs were compared to the optimal variances of the optimal designs to obtain the relative efficiencies of the designs. An application to a four factor experiment where incubation time, initial pH levels of the fermentation broth, incubation temperature during fermentation and initial substrate concentration were investigated in relation to bioethanol yield. A second order model was fit to the experimental data. Model adequacy checking was carried out using the normal probability plots, residual plots, regression analysis and analysis of variance (anova). F-ratios and t-statistics to test various hypotheses on significance of model parameters i.e linear, two-term interactions and quadratic terms were employed. Factor settings which optimized ethanol yield were determined through the path of steepest ascent analysis method using R programming as well as the determination of optimum ethanol yield using Response surface plots. Canonical analysis was used to analytically determine the point of maximum yield and the corresponding ethanol yield.

1.5 Objectives of the Study

1.5.1 General Objective

Apply a second order optimal rotatable design in four dimensions constructed using balanced incomplete block design to develop a statistical model that describes ethanol production through fermentation process of pineapples peels using Saccharomyces Cerevisiae (yeast).

1.5.2 Specific Objectives

The specific objectives for this study was to.

i. Determine $D-, E-, A-and\ T-$Optimal values of the general rotatable second order design in four dimensions constructed using balanced incomplete block designs.

ii. Derive Relative efficiencies of the $D-, E-, A-$ and $T-$Optimal designs over the general design

iii. Determine the effects of incubation time, incubation temperature, initial PH and substrate concentration on pineapple peels during fermentation using yeast on ethanol yield.

iv. Obtain Optimum settings of time, temperature, initial PH and substrate concentration of pineapple peels that led to optimal yield of ethanol using the design.

1.6 Significance of the Study

This study aimed at exploring prudent waste management method and at the same time explore cheap alternative sources of fuel to fossil fuels. Ethanol is a renewable bio-based resource which is oxygenated potentially reducing emissions in compression–ignition engines and at the same tackling urban pollution caused by the by-products of pineapple processing firms and markets which are known to exhibit both high biochemical oxygen demand (BOD) and chemical oxygen demand (COD) values which give rise to serious pollution problems if not properly disposed. Most studies in optimization of process variables in response surface methodology have concentrated on Box-Behnken designs, CCD and its hybrids CCDF yet rotatable second order designs constructed using balanced incomplete block designs have been successfully used in this work to optimize ethanol yield from pineapple peels. The D-,A-,E- and T-optimal values of the general design and the relative efficiencies of D-,A-,E- and T-optimal designs have been determined providing knowledge to scholars. To determine the levels of ethanol produced per experimental run with higher precision, the method of fractional distillation is recommended though it is tedious, slow and expensive to undertake for each and every other sample.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

In this chapter, response surface methodology (RSM) in relation to D-,E-,A-and T-optimality criteria is reviewed as well as optimal design and the relative efficiency of the general design to optimal designs. Literature on ethanol production, literature on suitability of pineapple peels as substrate for ethanol production using Saccharomyces Cerevisiae (yeast) as a fermentation agent and the process variables that influence ethanol production and its optimization using RSM. The research gaps on designs used in optimization method was also identified for study.

2.1 Optimality Criterion

The main goal of RSM is to use a sequence of designed experiments to obtain an optimal response as introduced by Box and Wilson, (1951).There are many situations for which RSM has proved to be a very useful tool. Hill and Hunter, (1966) illustrated chemical and processing applications of canonical analysis and use of multiple responses. As an important subject in the statistical design of experiments, (RSM) is a collection of mathematical and statistical techniques useful for the modelling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize this response Montgomery, (2005). The main advantage of RSM is the reduced number of experimental runs needed to provide sufficient information for statistically acceptable results, its suitability for multiple factor experiments and exploration of common relationship between various factors towards finding the most appropriate production conditions for the bioprocess and forecast response, Isaias et al., (2004)(Isaias et al., 2004)(Isaias et al., 2004). Response Surface Methodology is widely used in many disciplines such as Manufacturing Industry,

Biological, Clinical, Social, Food processing, Engineering, Agricultural sciences, amongst others. It is a tool in statistical analysis of experiments where the yield is believed to be influenced or determined by one or more controllable factors. Mead and Pike, (1975) investigated the extent to which RSM had been used in applied research and gave examples from biological applications which led to optimization of some conditions for good quality leads of the outputs. RSM has successfully been used by Zhang et al., (2015) to optimize the cell density and fermentation process. Myers et al., (1989) summarized the developments in RSM that had occurred since the review of Hill and Hunter, (1966) while a more recent summary by Khuri, (2017) provided an overview of response surface methodology including the modelling of a response function, the corresponding choice of design, and the determination of optimum conditions as well as an overview of the use of RSM in agricultural and food sciences with several examples taken from a variety of applied journals. Box and Hunter, (1957) introduced rotatable designs for the exploration of response surface. They constructed these designs through geometrical configurations and obtained several second order designs. Afterwards Gardiner et al., (1985) obtained some third order designs through the same technique for two and three factors and a third design for four factors. Bose and Draper, (1959) obtained some second order designs by using the transformation group in three dimensions and its generated points sets and the formation of rotatable arrangements and rotatable designs by combination of several points sets generated. Box and Behnken, (1960) obtained some second order rotatable designs from those of first order. Box and Hunter, (1957) showed that the number of Centre points in the rotatable central composite designs could be chosen to provide a design with uniform precision for the estimated surface within one unit of the design center co-ordinates on the coded scale. They reasoned that the investigator

is mostly interested in the response surface near the Centre of the design. Designs have been developed to have as close to the minimum number of points as possible to estimate the second order response surface. Tables of these designs or methods to construct them can be found in Draper, (1985). Most of these designs are based on 2^{n-p} fractional factorial designs augmented with design center points to estimate second-order response surface models. A class of three-level designs to estimate second-order response surfaces was proposed by Box and Behnken, (1960), these designs are rotatable or nearly so with a reduction in the number of experimental units by the 3^n designs. The designs are formed by combining 2^n designs with incomplete block designs. Box and Hunter, (1957) gave the conditions for blocking second order response surface designs so that the block effects do not affect the estimates of the parameters for the response surface equation. Das and Narasimham, (1962) gave a method using properties of balanced incomplete block designs (BIBD) of obtaining second order rotatable designs with any number of factors. The designs were observed to have reasonably small number of points and by extending the method, third order rotatable designs both sequential and non-sequential of up to fifteen factors were obtained using doubly balanced incomplete block designs and complementary B.I.B designs. Koske et al., (2011) developed a third order rotatable design in five dimensions with 320 points through balanced incomplete block design. Box and Wilson, (1951) observed that the circumstances of spheres are such that their exact rotatability is unattainable, although it is still a good idea to make the design nearly rotatable. Optimal designs are experimental designs that are generated based on a particular optimality criterion and are generally optimal only for a specific statistical model. Smith, (1918) was the first to state a criterion and obtain optimal designs for regression problems. Many years later, Kiefer, (1959) developed useful computational

procedures for finding optimum designs in regression problems of statistical inference. There are many optimality criteria, sometimes called alphabetical optimality criteria for designs. These are single number criteria where each one intends to capture an aspect of the 'goodness' of a design and can be classified into either information-based criteria, distance-based criteria, compound design criteria and other criteria. Information-based criteria are related to the moment matrix $X'X$ of the design. This matrix is important because it is proportional to the inverse of the variance-covariance matrix for the least-squares estimates of the linear parameters of the model. These criteria can be divided into two classes according to the number of parameters used; the first class uses all parameters of the model and the second uses a sub –system of the parameters. In the first class, possible criteria to consider are, D–, A– and E– optimality criteria. Statistical models with several parameters have their mean of the parameters estimator as a vector making the variance of the parameters estimate a matrix whose inverse is called the ''information matrix''. The optimality properties of designs are determined by their moment matrices, Pukelsheim, (2006).

2.1.1 D-Optimality

For information-based criteria**,** the most prominent of such criteria is the D-optimality criterion that maximizes the determinant of the moment matrix. This amounts to the minimization of the size of the confidence region on the vector β in model Pazman, (1992). Determinant criterion is the most prominent design criterion in the life applications, which was introduced by Wald, (1943) it puts emphasis on the quality of the parameter estimates. It was called later, D–optimality by Kiefer and Wolfowitz, (1959). D–optimality is the most well studied problem which is seen in the literature by Kiefer, (1959), Atkinson and Donev, (1992) and Pukelsheim and Rosenberger, (1993) where they considered the construction of D–optimal designs in a variety of

examples. Its popularity is due its simple computation, and the many available algorithms. D-optimality is parameter estimation criterion which aims at seeking designs which maximize the determinant of the moment matrix.

2.1.2 E-Optimality

E–optimality was introduced by Ehrenfeld, (1955), but the Computations of E–optimal designs for the full mean parameter vector and for many subsets in univariate polynomial regression models were determined by Rissanen, (1983). A method for computing E–optimal designs for a broad class of two parameter models was presented by Dette and Haines, (1994).The evaluation of the smallest Eigen value of the moment matrix $X'X$ of a design is the same as minimizing the largest Eigen value of the dispersion matrix $(X'X)^{-1}$ according to Pukelsheim, (2006).

2.1.3 A-Optimality

A–optimality criterion was introduced by Smith, (1918) as reported in Chernoff, (1953) and it involves the use of Fisher's information matrix. An algebraic approach for constructing A–optimal design under generalized linear models was presented by Yang, (2008).

2.1.4 T-Optimality

T-optimality is neither information-based criterion, distance-based criterion nor compound design criterion but falls under other criteria. T–optimal design is a plan where the optimality is obtained by discriminating between two or more models, one of which is true. Atkinson and Fedorov, (1975) introduced experimental designs for discriminating between two models and also between several models.

2.2 Efficiency of a Design

If the experimental region (R) is either spherical or cuboidal, a standard response surface design such as central composite design (CCD), Box-Behnken designs or their variations such as face-centred cube designs are applied since they are quite general and flexible Montgomery, (2005).

But occasionally during experimentation these designs are not the obvious choice and Optimal designs are the alternative when the experimental region is irregular due to factor levels constraints or when the experimenter has prior knowledge about the process being studied which may suggest a non-standard model where some higher order terms or some interaction terms between factors may not be included in the model or even when the process factors are categorical or an unusual sample size may be of importance due to cost or time considerations. In such cases, designs of fewer trials are carried out with the aim of obtaining an efficient design for fitting a reduced quadratic or higher order model. The efficiency of an experiment is influenced by the adoption of an appropriate experimental design capable of representing the response surface design. Selecting an appropriate experimental design, is based on finding the best optimality criterion in which larger efficiency values imply a better design.

2.3 Process Variables for Ethanol Production from Pineapple Peels

Ethanol production has received considerable attention over the years as an octane booster, fuel extender, or a neat liquid fuel. Renewable feed stocks used in its production provide a domestic endless supply of raw materials that are immune to disruption by foreign suppliers as well as improving international balance of payments. Wyman and Hinman, (1990) and Onuki et al., (2008) reviewed published literature on current ethanol production, separation methods, and chemical and sensory analysis techniques. According to their work, ethanol produced by

fermentation, called bioethanol, accounts for approximately 95% of the ethanol produced in the world which had risen to nearly 13.5 billion gallons in 2006 although it had been part of alcoholic beverages for a long time, its application expanded tremendously during the 20th Century as an additive to gasoline. In their study, Corn in the Unites States and sugarcane in Brazil are noted to be the widely used raw materials for bioethanol production and that Cellulosic materials are expected to be the ultimate major source of ethanol since they represent a value-adding technology for agricultural co-products.

Periyasamy et al., (2009) used *Saccharomyces cerevisiae to* produce bio-ethanol from sugar molasses. The influencing parameters that affected the production were optimized. The optimal values of the parameters such as temperature, pH, substrate concentration, enzyme concentration and fermentation period were found to be 35°C, 4.0, 300 mg/l, 2 mg/l and 72 h respectively. Under these optimum operating conditions, a maximum of 53% bio-ethanol yield was achieved. However the optimization method was that one of one-parameter-at-a time where identification of bio-ethanol was done by taking about 5 to 10 ml of the fermented sample and adding a pinch of potassium dichromate and a few drops of H_2SO_4. The colour of the sample turned from pink to green indicating presence of bioethanol. Determination of sugar concentration was by rapid test method where 5 ml of fermented sample was taken and dissolved in 100 ml of distilled water and mixed with 5 ml of conc. Hydrochloric acid (HCL) and was heated at 70^0C for a period of 10 minutes. The obtained sample was neutralized by adding $NaOH$ and then prepared to 1000 ml and taken into a burette. The 5 ml of Fehling A and 5 ml of Fehling B were taken and mixed with 10 to 15 ml of distilled water in a conical flask and Methylene blue indicator was added where the conical flask solution was titrated with burette solution in boiling

conditions until disappearance of blue colour. The sugar concentration was calculated by using the following formula: Sugar Concentration (mg/l) = [(Dilution factor x Fehling factor) / Titrate value] x 100.While ethanol concentrations were determined by gas chromatography. PH was optimized by fermenting the sample to different pH values between 1.0 to 8.0 and to obtain maximum yield of bioethanol, lime or sulphuric acid were added and the samples were kept in anaerobic conditions for four days. The fermented solution was analysed for every 12 hour intervals indicating that bioethanol increased with increase in pH and reached a maximum when pH was 4 and reduction was noted as pH increased due to lesser activity of the yeast.

There have been several reviews of literature on bioethanol production from sugar molasses using yeast cells (*Saccharomyces cerevisiae*) from various authors and the optimization of the process variables for bioethanol yield. Fakruddin et al., (2013) used the one-factor-at-a-time approach in a study in which only one factor is varied at a time while all others are kept constant. Stress tolerant yeast strains were isolated from agro industry and optimized a process for ethanol production by considering all the factors at a time. Several fermentation batches were carried out by three stress tolerant strains by varying temperature, pH, sugar concentration, aeration, metal ions and immobilization. The fermentation was carried out at varying temperature, pH, reducing sugar concentration, agitation and immobilized condition at a time. Alcohol percentage in the fermentation broth was measured by redox back titration (micro-diffusion) method. They found out that the rate of ethanol production by yeast cells is highly affected by the pH of the fermentation medium. More acidic and basic conditions both retard the yeast metabolic pathways and hence the growth of cells also. Optimum pH is required for growth and ethanol yield by the yeast strains. To determine the optimum pH for ethanol yield, several experiment were conducted at

pH 4.5, 5.0 and 6. Increases or decreases of the initial pH from 5.5 of the fermentation media decreased the ethanol yield and it was also noted that all the strains produced maximum ethanol at pH 5.5 after 60 hours except T strain. This approach is time-consuming and expensive and possible interaction of effects between variables cannot be evaluated and misleading conclusions may be drawn. RSM overcomes these difficulties, since it allows accounting for possible interaction effects between variables. If adequately used, this powerful tool can provide the optimal conditions that will improve ethanol production. Hamouda et al., (2015) in their study, aimed at optimizing the production of ethanol by Egyptian yeast strain *Pichia Veronae* on the batch flasks scale. RSM based on CCFD of experiments was used to overcome the limitation of one-at-a-time-parameter optimization. Zentou et al., (2017),evaluated the potential of molasses as bioethanol feedstock by studying the effect of different operating conditions on fermentation yield including initial sugar concentration (25-150 g/L), pH (4.5-9.5), and temperature (30-50°C). Molasses composition analyses indicated its richness with sucrose and fermentable sugars which qualified it as a promising feedstock for bioethanol production. The maximum ethanol yield was noted for: initial sugar concentration of 50 g/L, pH of 4.5 and 30 °C of temperature which represented the optimum conditions for the fermentation. The kinetics study of fermentation experiment carried out under optimal conditions revealed that the fermentation reaction occurred in 3 phases: lag phase, acceleration phase and final phase. Microorganism and culture media was prepared using 2-3 loops of active dry *Saccharomyces cerevisiae (*yeast) which was dissolved in 50 mL of distilled water which was then added directly into 200 mL of culture media containing diluted molasses, ammonium sulphate (0.7g/L), ammonium phosphate (0.4 g/L) and incubated at 35°C and shaking with 250 rpm for 6 hours. Anaerobic fermentation was

carried out in batch mode using bioreactor where, one litre volume containing 250 mL of culture media with different initial sugar concentrations, different temperatures and different pH values was fermented, the fermentation process was carried out in three duplicates during 72 hours , pH value was adjusted to 4.5, 7 and 9.5 using sulphuric acid (0.1M) and sodium hydroxide (0.1M); Inoculum for fermentation assays was prepared with different dilution rates for (50 ,100 and 150 g/L) of initial sugar. The fermentation temperature was kept at 20, 30 and 50 °C Optimization was through plotting the yield against the three parameters that were being optimized. While El-Gendy et al., (2013) used CCDF in an attempt to optimize variables: incubation period, initial PH, incubation temperature and molasses concentration which affect Bioethanol production by fermentation using Saccharomyces cerevisiae. Hayder et al., (2018) used, the response surface methodology (RSM) based on central composite design (CCD), to estimate the number of runs and optimum conditions for four independent variables that affected fermentation process of lignocellulosic materials which usually exist in the organic fraction of Iraqi municipal solid waste, and the independent variables were initial concentration, pH, and inoculum size and fermentation time. In fermentation process, the *Saccharomyces cerevisiae* was used as an inoculum. The aim of their study was bioethanol production from the cellulosic biomass under controlled optimum conditions and a maximum bioethanol yield of 332.9 mg/L, was practically achieved following thirty different experimental runs, as specified by 2^4–full factorial CCD. The optimum values for the four parameters, corresponding to the maximum yield were; initial sugar weight = 75 g/L, pH = 6, fermentation time = 39 hrs. (Aerobic fermentation = 24 hrs. and anaerobic fermentation = 15 hrs.), and finally yeast inoculum = 2 ml/l. The obtained data was utilized to develop a semi-empirical model, based on a second degree polynomial,

which helped to predict bioethanol yield. The model adequacy was tested using anova and the R^2= 0.9771, made the model acceptable. The developed model was used to generate contour plots and yield response surface. Maximum bioethanol production was observed in a Lab scale bioreactor reaching up to 492.9 mg/L within optimum conditions. Adnan et al., (2014) studied ethanol fermentation processes using glycerol as carbon source using local isolate, ethanol genic bacterium Escherichia coli SS1 in a closed system. Factors affecting bioethanol production from pure glycerol were optimized via response surface methodology (RSM) with CCFD. Four significant variables were found to influence bioethanol yield; initial pH of fermentation medium, substrate concentration, salt content and organic nitrogen concentration with statistically significant effects ($p < 0.05$) on bioethanol production. The significant factors were then analysed using CCFD. The optimum conditions for bioethanol production were substrate concentration at 34.5 g/L, pH 7.61, and organic nitrogen concentration at 6.42 g/L which gave an ethanol yield of approximately 1.00 mol/mol. In addition, batch ethanol fermentation in a two-litre bioreactor was performed at the glycerol concentration of 20 g/L, 35 g/L and 45 g/L, respectively. The ethanol yields obtained from all tested glycerol concentrations were approaching theoretical yield when the batch fermentation was performed at optimized conditions. Yusuf et al., (2012) study, revealed that pineapple peels can be used to produce ethanol using baker's yeast (Saccharomyces cerevisae) in an aerobic degradation of sugar. The various parameters of fermentation evaluated compared favourably with the standard values. Tropea et al., (2014) observed that Pineapple waste, which is the by-product of the pineapple processing, is rich in cellulose, hemicelluloses, sugar and other carbohydrates. These wastes consist of residual pulp, peels and skin which can be dried mechanically making it easier to store throughout the year. They contain high

sugar and lignocelluloses amounts making them a potential source of valuable fermentation and non-fermentation products. Biomasses are renewable non-fossil carbon, such as energy crops and lignocelluloses residues from plants, grasses, fruit wastes, cereals algae etc. which can be converted to ethanol through microbial fermentations. Almarsdottir et al., (2012) argued that utilization of biomass for ethanol production would ensure a continual energy supply. Wyman and Hinman, (1990) records that Lignocellulosic material is the most abundant biopolymer on earth and its annual production is estimated at approximately 50 billion tons. Choonut et al., (2014) and Sun and Cheng, (2002) found out that Lignocellulose comprise of two main classes of structural polysaccharides, cellulose and hemicellulose which when hydrolyzed, provide sources of fermentable sugars (glucose and xylose, respectively). To produce ethanol at a cheap cost, the supply of cheap raw material is vital. The economics of biofuel production by fermentation are influenced by the cost of the raw materials used, which accounts for more than half of the costs of production Choonut *et al*., (2014). Further they observed that pineapple peels are novel and potential raw material for ethanol production. To enhance the bio-digestibility of the wastes, lignocelluloses pretreatment either physically; chemically or biologically to increase accessibility of the enzyme to the materials is necessary. Pre-treatment results in enrichment of the difficult biodegradable materials, and improves the yield of reducing sugar and ethanol from the wastes. An increase of 1.71 fold of cellulose was observed after pre-treatment of pineapple peel with water and heat at 100^{o}C for 240 min to 34.03±1.30 g/L when the substrate concentration was 20g/L was observed. This method of pre-treatment was reported to be the most economically feasible. The results of fermentable sugars increase and various methods of pre-treatments are shown in Table 2.1.

Table 2.1: Composition of Treated and Un-Treated Pineapple Peel (Choonut et al., 2014)

Physical pretreatment	Chemical Pretreatment	Cellulose (%)	Hemicellulose (%)	Lignin (%)	Ash (%)
Water bath	H_2O	37.68±6.97	51.13±6.77	10.24±0.63	0.96±0.81
(100°C, 240 min)	H_3PO_4	41.86±1.28	42.83±2.93	14.63±2.70	0.67±6.65
Microwave (3 min)	H_3PO_4	36.96±1.94	53.54±1.20	9.58±1.40	0.24±0.28
	$(NH_4)_2SO_4$	32.01±0.35	59.87±0.71	7.76±1.03	0.37±0.04
Ultrasonic (60 min)	CH_3COOH	35.99±1.85	52.77±1.00	11.02±2.50	0.24±0.34
	H_3PO_4	30.55±1.84	57.84±1.65	11.20±0.14	0.41±0.04
Stream explosion	CH_3COOH	25.54±3.41	69.10±3.42	5.06±0.21	0.31±0.21
(121°C, 60 min)	H_3PO_4	27.84±6.89	42.91±10.61	25.15±1.10	4.10±3.39
Un-treated peel		21.98±2.34	74.96±2.55	2.68±1.54	0.38±0.25

Table 2.2: Effects of Pretreatment on Reducing Sugar Yield from Pineapple Peel(Choonut et al., 2014)

Physical Pretreatment.	Chemical Pretreatment.	Cellulose (%)	Reducing Sugar (g/L)
Water bath	H_2O	37.68±6.97	34.03±1.30
(100°C, 240 min)	H_3PO_4	41.86±1.28	40.10±3.98
Microwave (3 min)	H_3PO_4	36.96±1.94	31.22±1.54
	$(NH_4)_2SO_4$	32.01±0.35	25.74±1.11
Ultrasonic (60 min)	CH_3COOH	35.99±1.85	30.14±2.47
	H_3PO_4	30.55±1.84	20.25±2.78
Stream explosion	CH_3COOH	25.54±3.41	20.25±1.56
(121°C, 60 min)	H_3PO_4	27.84±6.89	25.47±2.58

Enzymatic hydrolysis is regarded as the most promising approach to liberating fermentable sugars in an energy-efficient way from the carbohydrates found in lignocelluloses in order to produce ethanol Tropea *et al*., (2014) Sugars released by enzymes are then fermented to ethanol by yeasts. There are several fermentation approaches that can be employed as follows: Sequential enzymatic hydrolysis and fermentation referred to as separate hydrolysis and fermentation (SHF), while when the two steps are simultaneously done, the process is called simultaneous Saccharification and fermentation (SSF). SSF has the advantage of preventing the buildup of hydrolysis products such as cellobiose and glucose, which reduce the rate of further substrate hydrolysis. However, it has to be carried out at temperatures that suit the fermenting organism. In the case of yeast, the temperature is generally below

40 ºC, which is below the optimum temperature for enzymatic hydrolysis (50 ºC) according to Tengborg et al., (2001). Fermentation also seems to decrease the inhibition of the enzymes probably by converting some of the toxic compounds present in the hydrolysate. These mechanisms increase the overall productivity, the concentration and also the final ethanol yield Barta et al., (2010). Tables for various composition of pineapple peels after different forms of pretreatment are shown in the appendix in Table D and Table E. From the foregoing literature, several researchers used CCD and its counterpart CCFD which is not a rotatable design to optimize fermentation and other processes but none has been done in literature on fitting similar studies using rotatable designs constructed using balanced incomplete block designs yet second order rotatable designs with any number of factors with reasonably small number of runs can easily be obtained as shown by Das and Narasimham, (1962) which are also fully rotatable unlike the CCDF. The optimality criteria of the general design with four factors and its efficiency to optimal designs have not been investigated also. This study aimed at optimization of bioethanol yield (as the response variable) using pineapple waste as the feedstock and yeast as the fermenter and the determination of factor settings namely: incubation time, pH initial of the medium, incubation temperature during fermentation, and substrate concentration of the pineapple peels which all are of continuous nature associated with this optimal yield hence RSM was used. Other than that, there were constrains in the design data from the variables used and therefore the experimental design had to meet requirements of the constraints. From the three dimensions plots and contours plots of the response surface some results were displayed which helped in understanding how the response changes with changes in design variables and since the study had four

independent process variables, general mathematical methods of locating the point of maxima would not have been employed.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter discusses $D-, E-, A-and\ T-$ optimality criteria of the general design and its relative efficiencies to the $D-, E-, A-and\ T-$ optimal designs. Effects of process variables fermentation time, initial pH of fermentation broth, incubation temperature and concentration of substrate used during fermentation using Pineapple peels as substrate for ethanol production, modeling, diagnostic checks of the fitted model using various tools and optimization using second order model using graphical method and analytical method.

3.1 Optimality Criterion

RSM is concerned with the selection and construction of an appropriate design that can provide adequate and reliable information concerning a certain response variable, denoted by Y and determination of a suitable model that best fits the data that can be generated from using the design chosen. Such a model gives an approximate functional relationship between the response variable Y and a set of control variables believed by the experimenter to have an effect on the response , the input variables are denoted by $x_1, x_2, \dots, x_k$ .Then finally the determination of optimal settings on the control variables that produce maximum (or minimum) response values within a certain region of interest R.

$$y_u = f(x_{iu}) + e_u \tag{3.1}$$

Where, $u = 1,2, \dots, N$ are the N observations and x_{iu} is the level of the i^{th} factor at the u^{th} run. The function $f(x_{iu})$ describes the form in which the response and input variables are related and e_u is the experimental error at the u^{th} run with mean zero

and variance σ^2.The objective is to establish a functional relationship between the response and the control variables which gives a summary of an experiment and enables prediction of response y_u for values of x_{iu} that are not included in the experiment. When f is known, values of x_i for $i = 1,2, \dots, k$ which optimize the response can be determined using calculus methods or response surface plots. The function f is approximated within an experimental region (R) by a polynomial of suitable degree in variables x_i and analysis of variance method is used to test the adequacy of the fitted polynomial. If a polynomial adequately represents the response relationship, then it is called a response surface and a response surface design is a design that allows fitting of a response surface and provides a measure for testing their adequacy. The function f is of degree one in $x_{iu's}$ if

$$y_u = \beta_0 + \sum_{i=1}^{k} \beta_i x_{iu} + e_u. \qquad (3.2)$$

This model is used in the initial stages of experimentation to identify the important factors of the process and their importance in the process. Data analysis, determination of the significance of models parameters, estimation of the mean response and determination of optimum operating conditions on the control variables that result in maximum or minimum response over the region of interest (R) is carried out using a second order model given in equation (3.3).

$$y_u = \beta_0 + \sum_{i=1}^{k} \beta_i x_{iu} + \sum_{i=1}^{k} \beta_{ii} x_{iu}^2 + \sum \sum_{i<j}^{k} \beta_{ij} x_{iu} x_{ju} + e_u \quad (3.3)$$

To estimate the unknown parameters in equation (3.3), a series of experiments (runs) $N > p = k + 1$ are performed in which response y is measured for a specified settings of control variables which in total constitute the response surface design or just a design which is denoted by a design matrix $D_{N \times k}$. Given k variates each at s levels, a design formed with N of the s^k variates treatment combinations, can be

written as the following $N \times K$ matrix, which we call the design matrix and denote it with$D_{N\times K}$.

$$D_{N\times K} = \begin{bmatrix} x_{11} & x_{21} & x_{31} & \cdots & x_{k1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{1N} & x_{2N} & x_{3N} & \cdots & x_{kN} \end{bmatrix} \quad (3.4)$$

The treatment combinations are called points of the design. Since the investigator is mostly interested in the response surface near the Centre of the design, it is important to know if a particular design is rotatable or not. According to Box and Hunter, (1957), a design of the form described will be rotatable design of order d if a response polynomial surface

$$y_u = \beta_0 + \sum_i \beta_i x_{iu} + \sum_{i\leq j} \beta_{ij} x_{iu} x_{ju} + \sum_{i\leq j\leq k} \beta_{ijk} x_{iu} x_{ju} x_{ku} + \cdots \quad (3.5)$$

is obtained from the treatments, on the variables $x_i, i = 1,2, \ldots, k$ and with some suitable origin and scale, can be fitted so that the variance of the estimated response $var[\hat{y}(x)] = \sigma^2 x_s'(X'X)^{-1}x_s$ from any treatment is a function of the sum of squares of the levels of the factors in that treatment combination, in other words, the variance of the estimated response at a point is a function of the square of the distance of a point from a suitable origin, so that the variances of all estimated responses at points equidistant from the origin are the same. The variance of the estimate y_u is only a function of the distance $\delta^2 = \sum_{i=1}^{k} x_{iu}^2$ of the points $x_1, x_2, \ldots, x_k$ from the Centre of the design. Spherical variance of the estimation of the response surface is achieved if the design points satisfy the following conditions as recorded in Box and Hunter, (1957) and Das and Narasimham, (1962)

i) $\sum_{u=1}^{N} x_{iu} = 0$

ii) $\sum_{u=1}^{N} x_{iu} x_{ju} = 0$

iii) $\sum_{u=1}^{N} x_{iu}^2 = constant$

iv) $\sum_{u=1}^{N} x_{iu}^2 x_{ju}^2 = constant$

v) $\sum_{u=1}^{N} x_{iu}^4 = 3\sum_{u=1}^{N} x_{iu}^2 x_{ju}^2$ For all $i \neq j$ (3.6).

Das and Narasimham, (1962), demonstrated the construction of various designs of different factors one of which was for a four factor design with number of replications r of the BIBD being less than three the number of times (denoted by λ) pairs of treatments occurred together in the design that is ($r < 3\lambda$). The values for the coded levels of the design denoted by letters a and b were obtained as follows

$$\sum x_{iu}^4 = 24a^4 + 2b^4 \quad (3.7)$$

$$\sum x_{iu}^2 x_{ju}^2 = 16a^4 \quad (3.8)$$

Relating the two as in equation (3.6) part (v), gives

i). $24a^4 + 2b^4 = 3(16a^4) \rightarrow \frac{b^4}{a^4} = 12 \Longrightarrow \frac{b^2}{a^2} = 2\sqrt{3}$ and

ii). $\sum x_i^2 = 24a^2 + 2b^2 = N$ (3.9)

will completely determine the values of a and b. The number of points in this design is $N = 40$. They observed that no center points were necessary in these designs though they may be added if required. Solutions to equations (3.9) gave $a = \pm 1.137241371$ and $b = \pm 2.116644693$ as the coded levels of the factorial and the axial parts of the design respectively. For the purposes of analysis, both values were truncated at four significant figures in this work. To estimate the parameters of the developed model we proceed as follows:

Given $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$ $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix}_{N\times p}$ $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}_{N\times 1}$ $\epsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{N\times 1}$

Where y is an $(N \times 1)$ vector of observations at each run, X is an $(N \times p)$ matrix of levels of independent variables known as the model matrix with $p = k + 1$,

β is a $(N \times 1)$ vector of regression coefficients, and ϵ is an $(N \times 1)$ vector of random errors.

$$y = X\beta + \epsilon \tag{3.10}$$

We assume $\boldsymbol{\epsilon}$ is normally distributed with mean zero and Cov $(\epsilon) = \sigma^2 I$ according to Montgomery, (2005) and the estimates of the parameters is given by

$$\hat{\beta} = (X'X)^{-1}X'y \tag{3.11}$$

The variances of the parameters estimate $(\hat{\beta})$ is obtained as

$$\text{Var}\,(\hat{\beta}) = \sigma^2(X'X)^{-1} \tag{3.12}$$

$$\hat{y}_u = x_u\hat{\beta} \tag{3.13}$$

Rotatability requires that the model have a reasonably consistent and stable variance throughout the region of interest R. The variance of the predicted response at some point x is

$$V[\hat{y}(x)] = \sigma^2 x'(X'X)^{-1}x \tag{3.14}$$

Box and Hunter, (1957) suggested that a second order response surface should be rotatable meaning that the $V[y(x)]$ is the same at all points x that are the same distance from the design center i.e. the variance of predicted response is constant on spheres. Since the aim of RSM is optimization and the location of the optimum point

is unknown prior to running the experiment, it makes sense to use a design that provides equal precision of estimation in all the directions. The design problem therefore consists of selecting row vectors $X^{i\times p}$, $i = 1,2,\dots,N$ from the design space X such that the design defined by these N vectors is, in some sense, optimal. The moment matrix of the design is given as

$$M = \frac{X'X}{N} \tag{3.15}$$

We assume N is fixed. Solutions to this problem consist of developing some sensible criterion based on the above model and using it to obtain optimal designs, Montgomery, (2005).For the optimality criterion, the class of ϕ_p-criteria, that is T-, D-, A- and E- corresponding to parameter values 1, 0, -1 and -∞ respectively are summarized in equation (3.16) as given in Pukelsheim, (2006). The amount of information inherent to $C_k(M(\eta))$ is provided by ϕ_p-criteria with $C_k(M(\eta)) \in$ PD(m), defined by:

$$\phi_p(C) = \begin{cases} \lambda_{min}(C), if\ p = -\infty \\ \det(C)^{1/s}, if\ p = 0 \\ \left[\frac{1}{s} trace\ C^p\right]^p, if\ p \neq 0, \pm\infty \end{cases} \tag{3.16}$$

For all C in PD (m). By definition ϕ_p(C) is a scalar measure which is a function of the Eigen values, determinant, trace and average variance of C for all p $\in$[-∞; 1]

3.1.1 D-Optimality Criterion

Let C be a parameter subsystem moment matrix of S dimension, then D-optimality is given by

$$\phi_0(C) = (\det C)^{1/s} \tag{3.17}$$

Where C is the moment matrix and S is the number of parameters in the model. Maximization of the determinant of the moment matrix is the same as minimizing the determinant of the dispersion matrix that is $(\det C)^{-1} = \det(C)^{-1}$.The focus of D-optimality is on estimation of model parameters through good attributes of the moment matrix, which is defined as $C = \frac{X'X}{N}$, where $N$is the total number of runs in the design, which is used as a penalty for larger designs. D-optimality requires one to maximize the determinant of the moment matrix, i.e. Max $|X'X|$ =min $|(X'X)^{-1}|$.Under the standard normality assumptions, $|X'X|$ is inversely proportional to the square of the volume of the confidence region for the regression coefficients. Hence the larger the determinant of $X'X$ the better the estimation of the model parameters. Quite often for second order models, there is no finite D-optimal design; however, one can still compare the results for a particular design to the theoretical values.

3.1.2 E-Optimality

The procedure here builds on finding the design which maximizes the minimum eigenvalue of $X'X$ or equivalently, minimize the maximum eigenvalue of $(X'X)^{-1}$ .The aim of E-optimality is to minimize the maximum variance of all possible normalized linear combinations of parameter estimates.

$$\text{Max } \lambda_{min}(X'X) = \text{Min } \lambda_{max}(X'X)^{-1}. \tag{3.18}$$

It is the minimization of the largest Eigen value of the dispersion matrix. Which is given by

$$\frac{1}{\phi_{-\infty}(C_k(A))} = \lambda_{max}(C_k(A)^{-1}) \tag{3.19}$$

The Eigen value criterion $\phi_{-\infty}$ is one extreme member of the matrix means ϕ_p corresponding to the parameter $p = -\infty$. It is one of the four particular members of the one dimensional family of matrix means ϕ_p that submits itself to the principles that a reasonable criteria must meet as presented in Pukelsheim and Rosenberger, (1993) expressed in the form

$$\phi_{-\infty}(C) = \lambda_{min}(C) \tag{3.20}$$

3.1.3 A-Optimality

This criteria purpose at minimizing the sum of diagonal elements of the inverse of the moment matrix which is equivalent to minimizing the average variance of the parameter estimates given as Min $trace(X'X)^{-1}$ matrix. Invariance under reparameterization loses its appeal if the parameters of interest have a definite physical meaning. The average variance criterion saves the situation by providing a reasonable alternative. If the coefficients matrix is partitioned into its columns, $k = (C_1, C_2, \dots, C_k)$. Then the inverse $\frac{1}{\phi_{-1}}$ can be represented as

$$\frac{1}{\phi_{-1}(C_k(A))} = \frac{1}{s} trace\ C_k(A)^{-1} \tag{3.21}$$

This corresponds to the average of the standardized variances of the optimal estimates of the scalar parameter systems $c_1'\theta, \dots, c_s'\theta$ formed from the columns of matrix C, Pukelsheim and Rosenberger, (1993). Therefore, the average variance criterion is given by

$$\phi_{-1}(C) = \left[\frac{1}{S} traceC^{-1}\right]^{-1} \tag{3.22}$$

3.1.4 T-Optimality

There are two choices for defining T–optimality criterion according to the number of models under discrimination.To discriminate between competing models, Atkinson

and Fedorov, (1975) introduced T-optimality design criterion in the context of optimal design theory. The T-criterion is given by

$$\phi_1(C) = \frac{1}{s} traceC \tag{3.23}$$

3.2 Relative Efficiency and Optimal Designs

A procedure that works with a smaller sample is usually more efficient than one that requires a larger sample. In other words, an efficient procedure produces results that maximize use of materials, time and energy. An efficient experimental design produces the desired experimental results with the minimum amount of resources (e.g. time and money). The best experimental design in any given condition is the one which estimates the desired effects and contrasts with maximum precision or efficiency. Efficiency as discriminating criteria allows the comparison between any design and the best design. One of the main objectives of RSM is the determination of the optimum settings of the control variables that result in a maximum (or a minimum) response over a certain region of interest, R. This requires having a 'good' fitting model that provides an adequate representation of the mean response because such a model is to be utilized to determine the value of the optimum, Montgomery, (2005). A design can be made better by varying the proportion that a particular vector is run. When all the regression vectors are run an equal number of times, we have a uniform design. Pukelsheim, (2006) gives details of obtaining the optimal weights of a design for matrix means $\emptyset_p$with $p \in (-\infty, 1]$ the optimal weights satisfy

$$w_i = \frac{\sqrt{b_{ii}}}{\sum_{j \leq N} \sqrt{b_{jj}}} \forall i = 1, \dots, N \tag{3.24}$$

Where $b_{11}, \dots b_{NN}$ are the diagonal elements of matrix B given as equation (3.25)

$$B = UC^{p+1}U' \tag{3.25}$$

Where $U = (XX')^{-1}XK$ and C is the information matrix,K is the coefficient matrix and the N regression vectors $x_1, \dots, x_N$ form the rows of the design matrix X**.** Hence w_i will form the proportion that each regression vector will be run to obtain the $D-, A-, E-$ and $T-$ optimal designs. The optimal variance for the design is given by equation (3.26)

$$V(\emptyset_P) = \left(\frac{1}{S} trace\ C^P\right)^{1/p} = \left(\frac{1}{S}\left(\sum_{j \le N} \sqrt{b_{jj}}\right)^2\right)^{1/p} \text{ if } p \neq 0. \text{ For } p \in (-\infty, 1) \quad (3.26)$$

3.2.1 D-Optimal Efficiency

The optimal weights of D-optimal design were obtained using equation (3.24) and the appropriate matrix B after substituting the value of $p = 0$ in equation (3.25) where $p \in (-\infty, 1)$ and the corresponding optimal variance from equation (3.26). B_d Matrix is given as

$$B_d = UCU' \quad (3.27)$$

Factorial weight corresponding to $D-$optimal

$$D_{fw} = \frac{\sqrt{b_{11}}}{\sum_{j \le N} \sqrt{b_{jj}}} \quad (3.28)$$

while the weight corresponding to axial part is obtained as

$$D_{aw} = \frac{\sqrt{b_{NN}}}{\sum_{j \le N} \sqrt{b_{jj}}} \quad (3.29)$$

Where $b_{11}, \dots b_{NN}$ are the diagonal elements of matrix B_d. The $D-$optimal design was obtained by replicating the factorial and axial parts as per the weights obtained to give the design matrix X_d which was used to obtain the D-optimal moment matrix C_d

$$C_d = \frac{X_d' X_d}{N_d} \quad (3.30)$$

where N_d was the number of runs in the D −optimal design. The optimal variance corresponding to D −optimal was calculated as.

$$V(\emptyset_0) = (trace\ C_d)^{\frac{1}{S}} \tag{3.31}$$

and the relative efficiency of the general to D-optimal design is

$$\phi_{Deff(\xi)} = \frac{(\det C)^{\frac{1}{S}}}{(\det C_d)^{\frac{1}{S}}} \tag{3.32}.$$

3.2.2. E −Optimal Efficiency

The minimum Eigen value of the moment matrix of the general design was obtained as $\lambda_{min}(C)$ with a corresponding normalized eigenvector Z. Since the minimum Eigen value had multiplicity one, there exists a matrix E where

$$E = \frac{zz'}{\|z\|} \tag{3.33}$$

of trace one such that $x_i' E x_i \leq \lambda_{min}(C)$ for all $x_i \in \chi$ which was used to determine the E −optimal weights corresponding to the factorial and axial parts of the design. The optimal variance corresponding to the E −optimal design is the minimum Eigen value of the resulting moment matrix of the design i.e. $\lambda_{min}(C_e)$ where

$$C_e = E'E/N \tag{3.34}$$

and N is the number of rows of matrix E.The E- efficiency of the general to the E −optimal design was obtained as

$$\phi_{Eeff} = \frac{\lambda_{min}(C)}{\lambda_{min}(C_e)} \tag{3.35}$$

3.2.3.A −Optimal Efficiency

The B_a matrix for computing the A −optimal weight was obtained by putting $p = -1$ in equation (3.25) so that

$$B_a = UU' \tag{3.36}$$

The A −optimal factorial weight (A_{fw}) was obtained as

$$A_{fw} = \frac{\sqrt{b_{11}}}{\sum_{j \le N} \sqrt{b_{jj}}} \tag{3.37}$$

while the weight corresponding to axial part

$$A_{aw} = \frac{\sqrt{b_{NN}}}{\sum_{j \le N} \sqrt{b_{jj}}}. \tag{3.38}$$

Where $b_{11}, \ldots b_{NN}$ are the diagonal elements of matrix B_a. The A −optimal design was obtained by replicating the factorial and axial parts as per the weights obtained above to give the design matrix X_a which was used to obtain the moment matrix

$$C_a = \frac{X_a' X_a}{N_a} \tag{3.39}$$

where N_a was the number of runs in the A −optimal design. The optimal variance corresponding to the A −optimal design is given by equation (3.40)

$$V(\emptyset_{-1}) = \left(\frac{1}{S} trace\ C_a^{-1}\right)^{-1} = \left(\frac{1}{S}\left(\sum_{j \le N} \sqrt{b_{jj}}\right)^2\right)^{-1} \tag{3.40}$$

The efficiency of the general design to the A −optimal was obtained as

$$\varphi_{Aeff} = \frac{\left(\frac{1}{S} trace\ C^{-1}\right)^{-1}}{\left(\frac{1}{S} trace\ C_a^{-1}\right)^{-1}} \tag{3.41}$$

3.2.4. T −Optimal Efficiency

To compute the weights corresponding to the T −optimal design p is set at one in equation (3.25) and hence the B_t matrix is given as

$$B_t = UC^2U' \tag{3.42}$$

Factorial weight corresponding to T −optimal is

$$T_{fw} = \frac{\sqrt{b_{11}}}{\sum_{j\leq N}\sqrt{b_{jj}}} \tag{3.43}$$

while the weight corresponding to axial part is

$$T_{aw} = \frac{\sqrt{b_{NN}}}{\sum_{j\leq N}\sqrt{b_{jj}}}. \tag{3.44}$$

Where $b_{11}, \ldots b_{NN}$ are the diagonal elements of matrix B_t. The T −optimal design was obtained by replicating the factorial and axial parts as per the weights obtained as above to give the design matrix X_t which was used to obtain the moment matrix $C_t = \frac{X_t'X_t}{N_t}$ where N_t was the number of runs in the T −optimal design. The optimal variance corresponding to the T −optimal design was given by

$$V(\emptyset_1) = \left(\frac{1}{S} trace\ C_t^1\right)^1 = \left(\frac{1}{S}\left(\sum_{j\leq N}\sqrt{b_{jj}}\right)^2\right)^1 \tag{3.45}$$

The T-efficiency for T-optimal designs with the moment matrix C_t and the general design with moment matrix C is given as

$$\frac{\phi(C)}{\phi(C_t)} = \frac{\left\{\frac{1}{S}traceC\right\}}{\left\{\frac{1}{S}traceC_t\right\}} \tag{3.46}$$

Where s is the number of parameters in the model.

3.3 Substrate and Effects of Factors on Fermentation.

The challenges of municipal wastes disposal and the health risks these wastes pose to the urban dwellers and the need for alternative forms of energy are serious issues that require attention. Pineapple waste is one of the main municipal wastes in Thika town due to the large scale cultivation of pineapples by plantation firms and the surrounding farms of the local community. The high amount of reducing sugars in pineapple peels makes it an ideal raw material for bioethanol production. Zentou et al., (2017) observed that there are many factors that influence the process of fermentation such as substrate concentration, temperature, fermentation period, pH of the fermentation medium, yeast concentration e.t.c. The optimal sugar concentration depends primarily on the physiological properties of the yeast; a high sugar concentration can create an extracellular osmotic pressure greater than that of the intracellular environment which makes the water in the cell to diffuse through a membrane of a hypotonic solution to a hypertonic solution Klis, et al,. (2006). In the very dilute sugar solution; water passes from the external medium to the intracellular environment which creates internal pressure leading to swelling and bursting of the cells. Moreover, an increase in initial sugar concentration increases the concentration of product (ethanol) presence of which has been shown to have an inhibition effect on yeast growth and fermentation activity which totally stops at high ethanol concentration Zhang et al., (2015). There are many parameters which can affect the enzymatic activities. To achieve the most effective productivity, the appropriate parameters, which can maximize the enzymatic activities and minimize the cost, are required. In an ethanol industry, the most closely controlled parameters are temperature and pH. Basically, higher temperatures give higher productivity. However, above a certain temperature, the enzyme starts losing its activity. This is

because the protein form of the enzyme is broken by the heat. Too high temperature kills yeast, and low temperature slows down yeast activity. Thus, to keep a specific range of temperature is vital to ethanol production. Also, an enzyme has an optimal pH range. In the range, the enzyme shows high ethanol production. However, if the pH changes drastically from the range, the enzyme loses its activity again. This phenomenon is same as one with high temperature, that is to say, the extreme pH can break enzyme formation and it cannot be recovered. Yeast is a facultative anaerobe. In an aerobic environment, it converts sugars into carbon dioxide and water. In an anaerobic environment, it converts sugars into carbon dioxide and ethanol.

3.3.1 Materials and Substrate Preparations.

The study used pineapple peels for fermentation. The peels were obtained from a local market in Thika town. The peels were sun dried for three days and then oven dried at 35^0C for three hours and milled into fine particles using a Ramtons fruit blender. 40g of the mill was carefully weighed using an electronic balance (type AY220 number D440620174 of capacity 220g with a reliability index of 0.1mg (manufactured by Shimadzu Corporation). The mass was dissolved in 1000mL of distilled water and immediately pre-treated at 100 ºC for 240 minutes in an incubator under continuous mixing to inactivate endogenous enzymes and reduce microbial spoilage. The content was allowed to come to room temperature and the filtrate settled at the bottom of the glass jar. The pH was measured using a pH meter model number LMPH-10 Mark LABMAN serial number L9254 and was found to be 3.95. Batch fermentations was carried out in 250 ml conical flasks fitted with rubber stoppers and clearly marked with stickers indicating run number, required fermentation time, pH, temperature and concentration of substrate. Five batches (as per the design points) of 250 ml in conical flasks of the pre-treated substrate were prepared and concentrations varied by

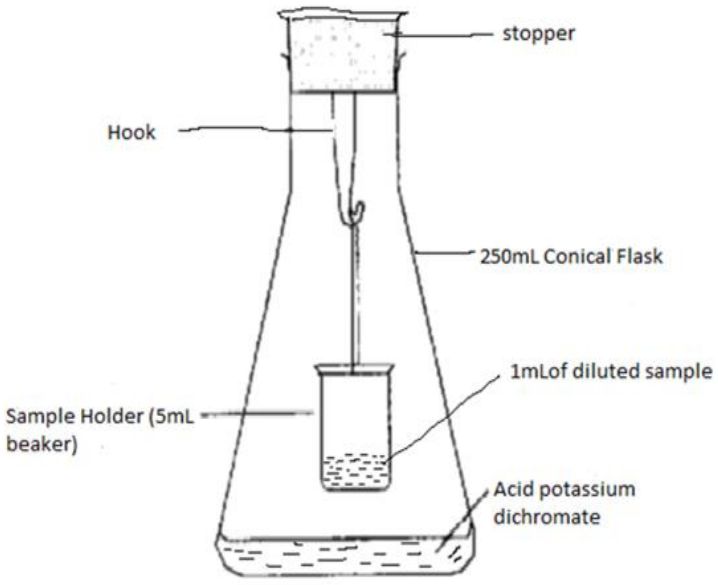
appropriately diluting using distilled water, further five other 100mL beakers were used where substrates with various concentration were placed in and pH adjusted by adding drops of 0.1M Sodium hydroxide to increase it from 3.95 which was found to be the pH of the pre-treated substrate. Then 10ml of these substrates with concentrations and pH's adjusted as per the experimental design points were now drawn using a Pipette and placed in the marked 250ml conical flasks of the 40 experimental runs. The fermentation started with addition of 10 ml of *Saccharomyces Cerevisiae* inoculum (10^7 cells per ml) which was prepared by dissolving 5g of Brewer's yeast in 1000mL of distilled water to the medium. The yeast contains an enzyme called invertase, which acts as a catalyst and helps to convert the sucrose sugars into glucose and fructose (both $C_6H_{12}O_6$). One mole of glucose is converted into two moles of ethanol and two moles of carbon dioxide in the first stage. In stage two, fructose and glucose react with another enzyme called zymase, which is also contained in the yeast to produce ethanol and carbon dioxide. Yeasts mainly metabolize glucose and fructose to form pyruvic acid, but the pyruvic acid generated is decarboxylated to acetaldehyde which then experiences dehydrogenation to ethanol. Yeasts often used in alcoholic fermentation are *Saccharomyces cerevisiae*, because it is highly tolerant to alcohol (12-18% v / v), resistant to high sugar levels and remains active in the fermentation at a temperature of 4-32℃. The fermentation process takes around three days to four days to be complete and is carried out at a temperature of between$25 - 40^0C$. The ethanol, which is produced from the fermentation process, still contains a significant quantity of water, which must be removed using the fractional distillation process. The process works by boiling the water and ethanol mixture. Since ethanol has a lower boiling point (78.3^0C) compared to that of water (100^0C), the ethanol turns into the vapour state before the water and can be

condensed and separated.For a fermentation process to be viable for generation of ethanol from lignocellulose, it must produce good yield of the product, at high rate, and generate concentrations that are economically recoverable. The utilization of glucose, xylose, mannose, galactose, arabinose, and rhamnose, in the presence of acetic, ferulic acid and a variety of degradation products from thermo-chemical pretreatment is a challange. *Saccharomyces cerevisiae* is a model eukaryotic organism, often used in research because it is easy to manipulate and culture. Being a eukaryotic cell it is similar in structure to human cells. Yeast is widely used in industries to manufacture enzymes and proteins for beer or wine making because it metabolizes glucose to ethanol. A strain of yeast called Baker's yeast is commonly used in leavening of bread and bakery products. This is because it is able to convert sugars present in dough to carbon dioxide and ethanol. The same species of yeast has a strain referred to as brewers yeast that was used in this fermentation process. The contents were placed in a rotation shaker at 30 rpm for two hours. Incubation was performed in four shaking incubators at 200rpm in Mount Kenya university analytical chemistry laboratories set at different temperatures as per the requirements and samples for analysis were taken after 24, 31.5,48, 60.9 and 72hours at different prescribed periods. Broth samples were drawn from the fermentation flasks using a 10 ml syringe: The drawn samples were immediately frozen at -10 ºC in a deep freezer until analysis time.

3.3.2 Experimental Set-Up for Ethanol Determination

Analysis of ethanol content in the sample was by the method of redox back titration as put forward by Krakwowiak et al., (1997) where for sharp end points detections during titrations, dilution of the fermented samples in the ratio of one to ten was necessary. Three samples of 1ml of the diluted sample were drawn using a micro

pippete and placed in a 5ml beaker (sample holder) and 10ml of acidified potassium dichromate (0.01M in 5.0M sulphuric acid) prepared by putting 1000 ml of distilled water in a 2000ml volumetric flask followed by slowly adding 543.5 ml of concentrated sulphuric acid with constant swirling and the flask being cooled under cold running tap water and 5.88g of potassium dichromate was carefully weighed and added into the cooled solution and the content was topped up to the mark with distilled water was placed in a 250ml conical flasks with matching rubber stoppers as shown in Figure 3.1 and the fermented samples were suspended over the dichromate for overnight.Three samples from each experimental run were prepared since the entire content of the conical flask was used in the titration. Then water and ethanol from the sample slowly evaporated from the sample holder and ethanol was oxidized to ethanoic acid by the acidified Potassium dichromate, the set up was left in a water bath at $25^0\text{C} - 30^0\text{C}$ degrees. After overnight, the flask was allowed to come to room temperature and the stopper loosened carefully where sample holder was removed and discarded.



:

Figure 3.1: Experimental set-up

The walls of the flasks were rinsed with distilled water then 100ml of distilled water was added turning the contents yellow in colour. Further 1ml of 1.2 molar Potassium iodide prepared by dissolving 49.8 g in a 250 volumetric flask and topping up to the mark using distilled water was added turning the contents dark brown and the flask swirled to mix. Ethanol is oxidized to acetaldehyde acid by reacting it with an excess of acidified Potassium dichromate solution as in equation (3.47)

$$2Cr_2O_7^{2-} + 16H^+ + 3C_2H_5OH \xrightarrow{yields} 4Cr^{3+} + 11H_2O + 3CH_3\text{COOH} \quad (3.47)$$

The excess potassium dichromate is oxidized by potassium iodide to produce iodine as in equation (3.48)

$$\text{Cr}_2\text{O}_7^{2-} + 14\text{H}^+ + 6\text{I}^- \xrightarrow{\text{yields}} 2\text{Cr}^{3+} + 7\text{H}_2\text{O} + 3\text{I}_2 \quad (3.48)$$

The iodine produced is then titrated with standard sodium thiosulfate ($Na_2So_3^{2-}OH$) solution (0.03molar) equation (3.49) prepared by placing 14.89g of sodium thiosulfate in a 2000L volumetric flask and adding distilled water while swirling to dissolve it and the content topped up to the mark.

$$\text{I}_2 + 2\text{S}_2\text{O}_3^{2-} \xrightarrow{\text{yields}} 2\text{I}^- + \text{S}_4\text{O}_6^{2-} \quad (3.49)$$

The brown color of the solution fades to pale-yellow and upon addition of Starch indicator solution (1.0 % starch solution) prepared by weighing 2g of soluble starch and dissolving it in 200ml of distilled water, the solution takes a blue-black color as a result of formation of starch-iodine complex and as more thiosulfate is added near the end-point, the blue-black color of iodine complex fades and the end-point of titration is reached just when the enough thiosulfate is added to react with all the iodine present and the solution becomes colorless. Three concordant results were

obtained (titers agreeing to within 0.1ml). One blank titration was carried out to inform us of how much acid dichromate was present at the start as no alcohol had been added it meant all the amount of dichromate was still present. To determine the amount of ethanol in the sample, the average volume of sodium thiosulfate used in titration of the sample is subtracted from the average volume of sodium thiosulfate used in titrating the blank sample and the corresponding moles are converted into grams per liter of substrate.

3.3.3 Modelling Ethanol Yield

Laboratory scale experimental data were fit to the design to study the effects of fermentation time, initial pH of the fermentation broth, incubation temperature and initial substrate concentration using a second order model equation (3.3) within the region of interest as shown in table 3.1

Table 3.1: Factors Settings in Coded and Natural Levels

Coded Levels	Time(hrs.)	pH	Temp(^{0}C)	Sub-Con(g/L)
2.116	72	7	40	40
1.137	60.9	6.3	36.5	35.4
0	48	5.5	32.5	30
-1.137	35.1	4.7	28.5	24.6
-2.116	24	4	25	20

The purpose of coding was to make mathematical progression easy. The factors were rescaled and therefore zero is in the middle of the center of the design while ± 1.137 and ± 2.116 are the distances from the center with directions which correspond to the factorial and axial parts of the design respectively. The natural variables x_i for $i = 1,2,3,4$ factors were converted into coded variables τ_j for $j = 1,2,3,4\ and\ 5$ using the relationship in equation (3.50)

$$\tau_j = \frac{\xi_j - [\max x_i + \min x_i]/2}{[\max x_i - \min x_i]/(2.116*2)} \quad (3.50)$$

The maximum and minimum values of $x_i$cover the range of dissimilarity in the input variables where ξ_j represents the natural variable corresponding to each coded level. The output response corresponding to each combination of input parameters was the mean Bioethanol produced in g/L per experimental run.

3.3.4 The Adequacy of the Fitted Model

The tests of validity of the fitted model provide an important examination to determine whether it offers an adequate approximation of the true response surface. Analysis of variance (anova), Tests for Significance of Regression coefficients, normality test and coefficient of determination were used to examine the fitted second order model.

i). Normality Test:

This test requires the error term $e_i's$ to be normally distributed with a mean of zero and a variance of δ^2 .This was achieved by plotting normal probability plot of residuals. If the residuals plots, approximately lie along a straight line, then the normality assumption is satisfied

$$e_i = Y_i - \hat{Y}_i \quad (3.51)$$

If $e_i \sim N(0, \delta^2)$, then the observations $Y_i's$ are also normally and independently distributed therefore test for the significance of the regression can be applied to determine if the relationship between the dependent and independent variables exists.

ii). Analysis of Variance for Significance of Regression

Table 3.2:Anova table

Variation Source	Degrees of Freedom	Sum of Squares	MSS	F
Regression	$p-1$	$SSR=\sum_{u}^{N}(\hat{y}_u-\bar{y})^2$	$SSR/P-1$	$MSSR/MSSE$
Error	$N-p$	$SSE=\sum_{u}^{N}(\bar{y}_u-\hat{y})^2$	$SSE/N-p$	
Total	$N-1$	$SST=\sum_{u=1}^{N}(y_u-\bar{y})^2$		

The error sum of squares (SSE) is a measure of the amount of variation explained by the regression model, the smaller the SSE the better the regression model.

$$SSE = SST - SSR \tag{3.52}$$

Where

$$SST = Y'Y - \frac{(\sum Y_i)^2}{n} \tag{3.53}$$

$$SSR = \hat{\beta}X'Y - \frac{(\sum Y_i)^2}{n} \tag{3.54}$$

Hence

$$SSE = Y'Y - \hat{\beta}X'Y \tag{3.55}$$

iii). Tests for Significance of Regression

A good estimated regression model will explain the variation of the dependent variable in a sample. Test of hypotheses about model parameters helps an experimenter to measure the effectiveness of the model. In general, the $F-$test is used for more than one coefficient or joint hypotheses and it is applied for testing the significance of the parameters of either the main effects or of two way interactions parameters or the parameters of the quadratic effects.

The hypotheses to be tested are

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 \text{ Verses } H_1: \beta_i \neq 0 \text{ for at least one } i. \quad (3.56)$$

Then comparison of the F value to the critical $F_{\alpha,p,N-p-1}$ is made and if $F_{cal} > F_{\alpha,p,N-p-1}$, H_0 is rejected or still reject the null hypotheses when the $p-value$ for F_{cal} is less than the level of significance α. When the hypotheses test is particular to one coefficient at a time , $t-$test is employed. For example to test the significance of contribution of variable i in a model the appropriate hypotheses is $H_0: \beta_i = 0$ verses $H_1: \beta_i \neq 0$ and the test for this hypotheses is called the $t-$statistic expressed as

$$t_0 = \frac{\widehat{\beta_i}}{\sqrt{\widehat{\sigma^2} w_{jj}}} \quad (3.57)$$

Where w_{jj} is the diagonal element of matrix $(X'X)^{-1}$ appendix (1.C) corresponding to $\widehat{\beta_i}$ the denominator of equation (3.57) being standard error of coefficient $\widehat{\beta_i}$. The statistic t_0 is compared with critical $t-$values and null hypothesis is rejected whenever $\|t_0\| > t(_{\frac{\alpha}{2},N-q-1)}$ with the implication that the variable contributes significantly to the model.

iv). Coefficient of Determination

In order to determine how well the estimated model fits the data, R-squared value can be used. The R-squared lies in the interval [0, 1], when it is closer to one, the better the regression equation fits the sample data. R-squared Measures the percentage of the variation of Y around $\bar{Y}$ that is explained by the regression equation, however adding a variable to the model will always increase R-squared regardless of whether or not the variable is statistically significant. Thus adjusted R-squared is preferred since when variables are added to the model, adjusted R-Squared will not necessarily increase in fact it decreases if unnecessary variables are added.

$$R^2 = \frac{SS_R}{SS_T} = \frac{SS_T - SS_E}{SS_T} = 1 - \frac{SS_E}{SS_T} \quad (3.58)$$

$$\bar{R}^2 = 1 - \frac{SS_E/_{N-p-1}}{SS_T/_{N-1}} \quad (3.59)$$

Where equation (3.58) represents the formula for finding R-squared and equation (3.59) is the formula for adjusted R-squared. When both are close to one it means estimated regression equation fits the data well.

3.4 Optimization of Bioethanol Yield

Optimization is the synonym of the word maximization or minimization that means choosing the best option Montgomery, (2005).It is a process of getting the optimal setting of an experiment.Optimization plays a key role in any response surface investigation. One of the main objectives of modelling the response is to use the fitted model in determining optimum conditions on the model's control variables that result in a maximum (or minimum) response over a certain region of interest, *R.* This, of course, assumes that the model has been screened to determine its suitability for providing an adequate representation of the mean response over the region *R.* Quite often, a second-degree model is employed since it incorporates curvature after a series of experiments have been sequentially carried out leading up to a region that is believed to contain the location of the optimum response, Khuri, (2017). RSM is equipped with statistical tools for determining the significance of a factor over a response. The evaluation of factors using the RSM uses experimental design in order to distribute the selected variables within the boundaries of the design. The basic strategy involves four steps as put forward by Box and Wilson, (1951) as follows:

i. Move into the optimum region.
ii. Study the behavior of the response in the optimum region.
iii. Estimate the optimum conditions.
iv. Verify the model.

Model-based Optimization uses model fits to identify the system optimum. Determination of the optimum type, is by plotting contour and image plots in case of two factors, and response surface plots or the analytical determination using calculus method.

3.4.1 Response Surface Plots

The location of the stationary point which may be characterized as maxima, minima or a saddle point is a point in the levels of $x_i's \;\; i = 1,2,\dots,k$, that optimize response. This point if it exists is a point in levels of $x_i's$ for which the partial derivatives are

$$\frac{\delta\hat{y}}{\delta x_1} = \frac{\delta\hat{y}}{\delta x_2} = \cdots = \frac{\delta\hat{y}}{\delta x_k} = 0 \qquad (3.60)$$

This point say $x_{1s}, x_{2s}, \dots, x_{ks}$ is called the stationary point. By drawing contour plots using computer soft wares, experimenter can usually characterize the shape of the response surface and locate the optimum with reasonable precision Montgomery, (2005).Single factor experiments were performed to determine the appropriate range of conditions for ethanol production using pineapple waste where temperature, incubation time, substrate concentration and pH levels were considered. Each input variable was varied over a range of values for each level keeping others constant.

The function $f(x_1, x_2)$ that relates the control variables to the response Y can be plotted as shown in Figure 3.2

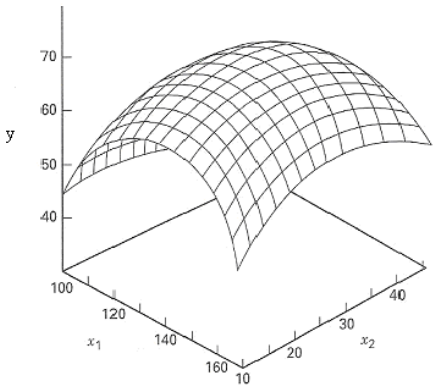


Figure 3.2: Response Surface Plot (Montgomery, 2005)

In this graph, each value of x_1 and x_2 generates a y-value. This three-dimensional graph shows the response surface from the side and it is called a response surface plot. Sometimes, it is less complicated to view the response surface in two-dimensional graphs. The contour plots show contour lines of x_1 and x_2 pairs that have the same response value y. An example of contour plot is shown in Figure 3.3.

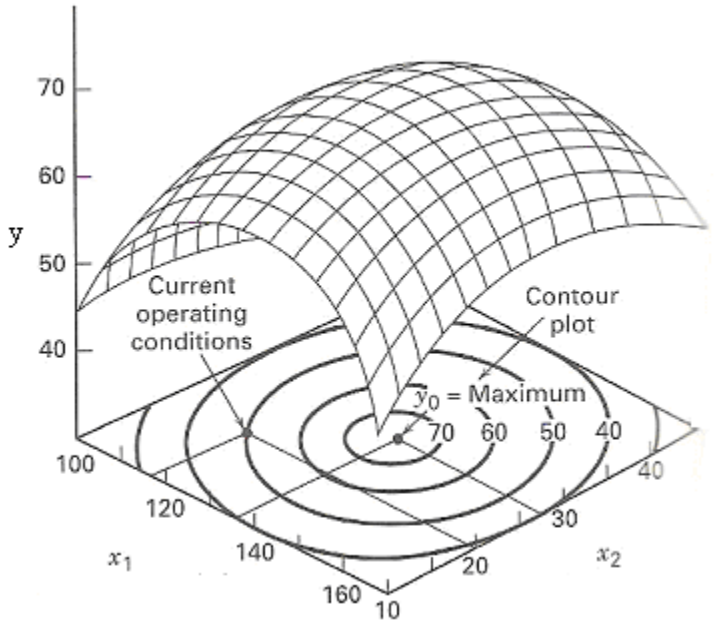


Figure 3.3: Contour Plot (Montgomery, 2005)

In order to understand the surface of a response, graphs are helpful tools, but when there are more than two independent variables, graphs are difficult or almost impossible to use to illustrate the response surface, since it is beyond three-dimensions. For this reason, response surface models are essential for analyzing the unknown function$f$.The exploration of an experimental region using response surface methods revolves around the assumption that the expected response, E(y), is a function of controllable variables $x_1, x_2, \dots, x_k$ where the $x_i's$ are suitably scaled and centered. Linear transformations of the independent variables with aim of studying RSM is accomplished by

i. Understanding the topography of the response surface (local maximum, local minimum, ridge lines), and

ii. Finding the region where the optimal response occurs and the goal is to move rapidly and efficiently along a path steepest to get to a maximum or a minimum response so that the response is optimized Montgomery, (2005).

3.4.2 Optimization Analytically

The optimum point may be determined analytically by using the equations (3.61)

$$\hat{Y} = \hat{\beta}_0 + X'b + X'BX \tag{3.61}$$

Which can be written in matrix form as

Where $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$, $b = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$ and $B = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \dots & \hat{\beta}_{1k}/2 \\ & \hat{\beta}_{22} & \dots & \hat{\beta}_{2k}/2 \\ & & \ddots & \vdots \\ sym & & & \hat{\beta}_{kk} \end{bmatrix}$

Where b is a $k \times 1$ vector of first order regression coefficients and B is a $k \times k$ symmetric matrix whose main diagonal elements are the pure quadratic coefficients β_{ii} and whose off diagonal elements are one-half the mixed quadratic coefficients β_{ij} $(i \neq j)$. The derivative of $\hat{Y}$ with respect to vector X equated to zero is

$$\frac{\delta \hat{Y}}{\delta X} = b + 2BX = 0 \tag{3.62}$$

The stationary point is

$$X_s = -\frac{1}{2}B^{-1}b \tag{3.63}$$

The predicted response is

$$\hat{Y} = \hat{\beta}_0 + \frac{1}{2}X_s'b \tag{3.64}$$

Once the stationary point is found we characterized the response surface in the immediate vicinity of this point i.e. determined whether the point is a point of

maximum, minimum or saddle point. Also to study the relative sensitivity of the response to the variables $x_1, x_2, \ldots, x_k$ contour plots of the fitted model whose construction and interpretation is relatively easy when two or three response variables are involved were examined. A formal analysis called canonical analysis was used which involved transforming model into a new co-ordinate system with the origin at the stationary point and then rotating the axes of this system until they are parallel to the principal axis of the fitted response surface which results into fitted model of the form

$$\hat{Y} = \hat{y}_s + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \cdots + \lambda_k w_k^2 \qquad (3.65)$$

Where w_i's in equation (3.65) are the transformed independent variables and λ_i's the Eigen values or characteristic roots of matrix B ($BX = \lambda X$ **),** equation (3.65) is known as the "canonical form "of the model. The nature of the response surface was determined from the stationary points and the signs and magnitude of the $\lambda_i's$. If all the eigenvalues are negative, then $x_s$is a point of maximum and it's a minimum point if all are positive and x_s is a saddle point if λ's are of mixed signs.

3.4.3 Modelling and Optimization using E-Optimal Design

The most efficient design relative to the general design which was the E-optimal design was employed in modelling the effects of the process variables on the ethanol production as well as the optimizing ethanol yield.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.0 Introduction

In this chapter, $D-, A-, E-$ and $T-$optimal values of the general design, the weights corresponding to $D-, A-, E-$ and $T-$optimal designs as well as the optimal variances of these optimal designs and relative efficiencies to the general design were computed and presented. The experimental, predicted as well residues values were presented. Fitted second order model developed was presented and results of its adequacy check displayed in form of plot of residuals versus fitted values, normal probability plot of the residuals, standardized residues against run numbers as well as plot of predicted versus actual values. The images, contours and response surface plots for the yield were also presented. Steepest ascent and analytical determination of the stationary point and the maximum yield results are presented.

4.1 The $D-, A-, E-$ and $T-$Optimal Values of the General Design.

The general design matrix X was obtained by substituting in the values of $a = \pm 1.137$ and $b = \pm 2.116$ as the factorial and axial points of the design as per equation (3.9) and fitting a full second order model. The model matrix X is displayed in Appendix table. A. The moment matrix for the general design with $N = 40$ is M_G given in Table 4.1.

Table 4.1: Moment Matrix of the General Second Order Design

$$M_G = \begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.999 & 0.000 & 0.000 & 0.000 & 0.999 & 0.000 & 0.000 & 0.999 & 0.000 & 0.999 \\
0.000 & 0.999 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.999 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.999 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.00 & 0.000 & 0.999 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.999 & 0.000 & 0.00 & 0.000 & 0.000 & 2.005 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.669 & 0.000 & 0.669 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.999 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.000 & 2.005 & 0.000 & 0.000 & 0.669 & 0.000 & 0.669 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.000 \\
0.999 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 2.005 & 0.000 & 0.669 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 \\
0.999 & 0.000 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.000 & 0.669 & 0.000 & 0.000 & 0.669 & 0.000 & 2.005
\end{bmatrix}$$

4.1.1 D −Optimal Value for the General Design

Determinant criterion was obtained using equation (3.17) with C being the moment matrix of the design M_G and $s = 15$ being the number of parameters in the model.

$$\phi_0(M_G) = (\det\ M_G)^{\frac{1}{15}} = 0.6796529 \qquad (4.1)$$

4.1.2 E −Optimal Value for the General Design

The smallest Eigen-value criterion $\phi_{-\infty}(C) = \lambda_{min}(C)$ and $C = M_G$ therefore using equation (3.20) the E −optimal value becomes

$$\phi_{-\infty}(C) = \lambda_{min}(C) = \lambda_{min}(M_G) = 0.002856958 \qquad (4.2)$$

4.1.3 A −Optimal Value for the General Design

The average variance criterion ($A - optimal\ value$) is as per equation (3.21) where C as above is the moment matrix M_G and s is the number of parameters in the model to be estimated.

$$\phi_{-1}(C) = \left(\frac{1}{15} trace\ {M_G}^{-1}\right)^{-1} = 0.04104631 \qquad (4.3)$$

4.1.4 $T-$Optimal Value for the General Design

The trace criterion is useless if the regression vectors $x \in \chi$ have a constant squared length α say, then the moment matrix $M(\xi)$ of any design $\xi \in \Xi$ satisfies

Trace $M(\xi) = trace \int_{\chi} \chi\chi' d\xi = trace \int_{\chi} \chi'\chi \, d\xi = c$ whence ϕ_1 constant providing no distinction whatsoever, according to Pukelsheim, (2006) but in this case, the sum of the squares of the regression vectors $x \in \chi$ have different values as follows

$$\sum x_0^2 = \sum x_1^2 = \sum x_2^2 = \sum x_3^2 = \sum x_4^2 = 40,$$

$$\sum (x_i^2)^2 = 80.28822 \text{ for } = 1,2,3, \&4 \text{ ,}$$

$$\sum (x_i x_j)^2 = 26.76274 \text{ .} \qquad (4.4)$$

Hence the trace will be of significance as an optimality criterion. The trace criterion with $C = M_G$ and s being the number of parameters to be estimated, the $T-$optimal value became

$$\phi_1(C) = \frac{1}{s} trace\,(C) = \frac{1}{15} trace\,(M_G) = 1.135448 \qquad (4.5)$$

Table 4.2: The Four Optimality Criteria for the General Design

$D-$	$A-$	$E-$	$T-$
0.6796529	0.04104631	0.002856958	1.135448

4.2 Relative Efficiencies of the Optimal Designs

To obtain an optimal design, equation (3.24) was used to determine the number of times each regression vector would be run (i.e. the weight) and the corresponding B matrices obtained as per equation (3.25) and optimal values were obtained using equation (3.26)

4.2.1. D −Optimal Design and its Relative Efficiency

The weight corresponding to factorial part is 0.02387036 and that one corresponding to axial part is 0.02951858 as per equation (3.24),Hence a D −optimal design was a design with factorial part replicated approximately two times i.e. ($n_f = 2$) with the axial part being replicated three times (i.e. $n_a = 3$) giving a design with $32 \times 2 + 8 \times 3 = 88$ points. The D −optimal moment matrix is $M_d = \frac{1}{88} X'X$ is:

Table 4.3: Moment Matrix for the D-Optimal Design.

$$M_d = \begin{bmatrix}
1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0109 & 0.0000 & 0.0000 & 0.0000 & 1.0109 & 0.0000 & 0.0000 & 1.0109 & 0.0000 & 1.0109 \\
0.0000 & 1.0109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 1.0109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 1.0109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
1.0109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 2.2809 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.6082 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
1.0109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.0000 & 2.2809 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.6082 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.0000 \\
1.0109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 2.2809 & 0.0000 & 0.6082 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 \\
1.0109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 0.0000 & 0.6082 & 0.0000 & 2.2809
\end{bmatrix}$$

The D-optimal value is $V(\emptyset_0) = (\det M_d)^{1/15}$ .Hence the optimal variance is $V(\emptyset_0) = 0.6965612$, while the optimal value for the general design was 0.6796529 .Therefore the relative efficiency of the general design to the D −optimal design was obtained as per equation (3.28)

$$\phi_{eff(\xi)} = \frac{0.6796529}{0.6965612} = 97.714587031\% \cong 98\%. \qquad (4.6)$$

4.2.2. E − Optimal Design and its Efficiency

The minimum Eigen value of the moment matrix of the general design is $\lambda_{min}(MG) = 0.002893284$ with a corresponding normalized Eigenvector:

$$Z' = [0.895\ 0.00\ 0.00\ 0.00\ 0.00\ -0.233\ 0.00\ 0.00\ 0.00\ -0.233\ 0.00\ 0.00\ -0.233\ 0.00\ -0.233]$$

According to equation (3.33), the E matrix for all x_i for $i = 1, \dots, 32$. $x \in \chi$, was obtained. Giving $x'Ex = 0.0009351244$ while for all x_i and $i = 33, \dots, 40$, we had $x'Ex = 0.01089277$ which is greater than the minimum eigenvalue therefore the E −optimal design assigns a weight of one to the factorial part and zero to the axial part of the design. The moment matrix of the E −optimal design is $M_e = \frac{1}{32}X'X$

Table 4.4: Moment Matrix for the E-Optimal Design

$$M_e = \begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.970 & 0.000 & 0.000 & 0.000 & 0.970 & 0.000 & 0.000 & 0.970 & 0.000 & 0.970 \\
0.000 & 0.970 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.970 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.970 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.970 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.970 & 0.000 & 0.000 & 0.000 & 0.000 & 1.255 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.836 & 0.000 & 0.836 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.970 & 0.000 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.000 & 1.255 & 0.000 & 0.000 & 0.836 & 0.000 & 0.836 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.000 \\
0.970 & 0.000 & 0.000 & 0.000 & 0.000. & 0.836 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 1.255 & 0.000 & 0.836 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 \\
0.970 & 0.000 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.000 & 0.836 & 0.000 & 0.000 & 0.836 & 0.000 & 1.255
\end{bmatrix}$$

With an optimal variance of 0.4182000.The relative efficiency of the general to the $\phi_{-\infty}$ −optimal design was obtained using equation (3.35) which was found to be 1%.

4.2.3 A −Optimal Design and its Efficiency

For $A-$ optimal design, $p = -1$, in equation (3.25) the $A-$ optimal weights corresponding to factorial and axial parts are 0.01704711 and 0.05681391 respectively. Therefore the $A-$ optimal design was obtained by replicating the factorial part two times (i.e.$n_f = 2$) and the axial part six times (i.e. $n_a = 6$) giving a

design with $32 \times 2 + 8 \times 6 = 112$ and $A-$ optimal moment matrix was $M_A = \frac{1}{112} X'X$

Table 4.5: Moment Matrix for A-Optimal Design.

$$M_A = \begin{bmatrix}
1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0343 & 0.0000 & 0.0000 & 0.0000 & 1.0343 & 0.0000 & 0.0000 & 1.0343 & 0.0000 & 1.0343 \\
0.0000 & 1.0343 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 1.0343 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 1.0343 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0343 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
1.0343 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 2.8674 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.4779 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
1.0343 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.0000 & 2.8674 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.4779 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.0000 \\
1.0343 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 2.8674 & 0.0000 & 0.4779 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 \\
1.0343 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 0.0000 & 0.4779 & 0.0000 & 2.8674
\end{bmatrix}$$

The A-optimal variance was found to be 0.05798174 while the optimal value for the general design was 0.04154701. The relative efficiency of the general compared to the optimal was $0.71655334938 \cong 71.7\%$.

4.2.4 *T*-Optimal Design and its Efficiency

The $T-$optimal weights corresponding to factorial and axial parts were 0.01691205 and 0.05735179 respectively as per equations (3.22) when $p = 1$, the $T-$optimal design was approximated by replicating the factorial part of the general design two times and the axial part six times giving a design with a total of $32 \times 2 + 8 \times 6 = 112$ runs .The T-optimal design moment matrix was found to be the same as the one in table 4.5. The $T-$optimal variance was 1.29828. Relative efficiency of the general design to the optimal design was $\frac{1.136227}{1.29828} \cong 87.5\%$.As per equation (3.46).

4.3 Process Variables of Bioethanol Yield

4.3.1 Ethanol Content Determination

The volume of sodium thiosulfate used in titrating the blank sample was found to be 9.7mL. From the reaction equation (3.49) six moles of $Na_2SO_3^{2-}$ is equivalent to one mole of $Cr_2O_7^{2-}$ and from equation (3.47) two moles of $Cr_2O_7^{2-}$ is equivalent to three moles of C_2H_5OH therefore one mole of $Na_2O_3^{2-}$ is equivalent to 0.25moles of C_2H_5OH. Let xmL be the average titer volume of three concordant titrations for sample i, then the mass of ethanol in the sample is equal to $(9.7 - x) \times 0.03 \times \frac{1}{4} \times 10 \times 46$ where 0.03 is the moles of sodium thiosulfate while $\frac{1}{4}$ is the equivalent moles of ethanol in the sample and 10 is the dilution factor and 46 is the mass in grams of one mole of ethanol $(2 \times C = 24 + 5 \times H = 5 + 1 \times O = 16 + 1 \times H = 46)$.The procedure was applied for the forty experimental runs replicated thrice and the average per run is the observed yield of ethanol in g/L in table 4.6.

Table 4.6: Experimental, Estimated and Residuals Values Ethanol

Run	X1	X2	X3	X4	Time	pH	Temp	Conc	Y Observed	Y Estimated	Errors
1	-1.137	0	-1.137	-1.137	35.1	5.5	28.5	24.6	7.1	6.3	0.8
2	1.137	0	-1.137	-1.137	60.9	5.5	28.5	24.6	5.1	5.2	-0.1
3	-1.137	0	1.137	-1.137	35.1	5.5	36.5	24.6	6.8	7.2	-0.4
4	1.137	0	1.137	-1.137	60.9	5.5	36.5	24.6	9.8	10.6	-0.8
5	-1.137	0	-1.137	1.137	35.1	5.5	28.5	35.4	7	7.5	-0.5
6	1.137	0	-1.137	1.137	60.9	5.5	28.5	35.4	6.6	6.4	0.2
7	-1.137	0	1.137	1.137	35.1	5.5	36.5	35.4	6.3	6.5	-0.2
8	1.137	0	1.137	1.137	60.9	5.5	36.5	35.4	10.5	9.9	0.6
9	-1.137	-1.137	0	-1.137	35.1	4.7	32.5	24.6	8.4	8.2	0.2
10	1.137	-1.137	0	-1.137	60.9	4.7	32.5	24.6	10.9	11.3	-0.4
11	-1.137	1.137	0	-1.137	35.1	6.3	32.5	24.6	6.7	8.4	-1.7
12	1.137	1.137	0	-1.137	60.9	6.3	32.5	24.6	5.6	7.6	-2
13	-1.137	-1.137	0	1.137	35.1	4.7	32.5	35.4	5.1	7.3	-2.2
14	1.137	-1.137	0	1.137	60.9	4.7	32.5	35.4	9.1	10.4	-1.3
15	-1.137	1.137	0	1.137	35.1	6.3	32.5	35.4	9.9	9.8	0.1
16	1.137	1.137	0	1.137	60.9	6.3	32.5	35.4	8.9	9	-0.1
17	-1.137	-1.137	-1.137	0	35.1	4.7	28.5	30	6.9	7.7	-0.8
18	1.137	-1.137	-1.137	0	60.9	4.7	28.5	30	7.4	8.5	-1.1
19	-1.137	1.137	-1.137	0	35.1	6.3	28.5	30	7	9	-2
20	1.137	1.137	-1.137	0	60.9	6.3	28.5	30	3.6	6	-2.4
21	-1.137	-1.137	1.137	0	35.1	4.7	36.5	30	7.4	7.7	-0.3
22	1.137	-1.137	1.137	0	60.9	4.7	36.5	30	11.7	13	-1.3
23	-1.137	1.137	1.137	0	35.1	6.3	36.5	30	9.3	9	0.3
24	1.137	1.137	1.137	0	60.9	6.3	36.5	30	10.4	10.5	-0.1
25	0	-1.137	-1.137	-1.137	48	4.7	28.5	24.6	9.2	7.9	1.3
26	0	1.137	-1.137	-1.137	48	6.3	28.5	24.6	4	6.2	-2.2
27	0	-1.137	1.137	-1.137	48	4.7	36.5	24.6	10.8	11.1	-0.3
28	0	1.137	1.137	-1.137	48	6.3	36.5	24.6	9.3	9.4	-0.1
29	0	-1.137	-1.137	1.137	48	4.7	28.5	35.4	7.3	8	-0.7
30	0	1.137	-1.137	1.137	48	6.3	28.5	35.4	9.1	8.5	0.6
31	0	-1.137	1.137	1.137	48	4.7	36.5	35.4	6.9	9.3	-2.4
32	0	1.137	1.137	1.137	48	6.3	36.5	35.4	9.9	9.8	0.1
33	2.116	0	0	0	72	5.5	32.5	30	9.3	8.3	1
34	-2.116	0	0	0	24	5.5	32.5	30	5.9	6.1	-0.2
35	0	2.116	0	0	48	7.0	32.5	30	8.4	11.2	-2.8
36	0	-2.116	0	0	48	4.0	32.5	30	9.5	12.4	-2.9
37	0	0	2.116	0	48	5.5	40	30	7.9	8.5	-0.6
38	0	0	-2.116	0	48	5.5	25	30	4.6	4.3	0.3
39	0	0	0	2.116	48	5.5	32.5	40	6.8	6.9	-0.1
40	0	0	0	-2.116	48	5.5	32.5	20	6.3	6.5	-0.2

4.3.2 Modelling Ethanol Yield

Based on the general Second order rotatable design constructed using balanced incomplete block design and observed yields, the statistical combinations of the variables in coded form along with the experimental and predicted data are presented in table 4.6. The fitted full second order model for the general design was obtained using R programming as equation (4.7).

$$Y = 11.821 + 0.5127X_1 - 0.2687X_2 + 0.9937X_3 + 0.1089X_4 - 0.7590X_1X_2 + 0.8654X_1X_3 + 0.2127X_1X_4 + 0.4448X_2X_3 + 1.1168X_2X_4 - 0.3723X_3X_4 - 1.03X_1^2 - 0.6274X_2^2 - 1.2102X_3^2 - 1.1382X_4^2 \quad (4.7)$$

The regression equation characterizes the influence of the different variables on bioethanol yield. Positive signs in front of the terms indicate synergetic effect while negative sign indicate antagonistic effects. The model shows that within the studied range of the variables, pH(X_2), had an antagonistic effect on ethanol yield of 0.2687 while all the other variables had synergetic effect. Incubation temperature had the highest effect of 0.9937 followed by time while substrate concentration of the pineapple peel had the least effect. A comparison of the observed results between experimental and predicted readings showed a good matching in between and over the defined range.

4.3.3 Adequacy of the Model

i). Plot of Residuals versus Fitted Values

If the fitted model is correct, and if the assumptions of normality for the errors are satisfied, then the residuals should be structure less, in particular they should be unrelated to any other variables including the response and the predicted values. A plot of the residuals versus fitted values should not reveal any obvious pattern. The normal probability plot, of the residuals approximate a straight line hence the model explains variations of the dependent variables in the sample meaning that the responses Y_i are also normally and independently distributed.

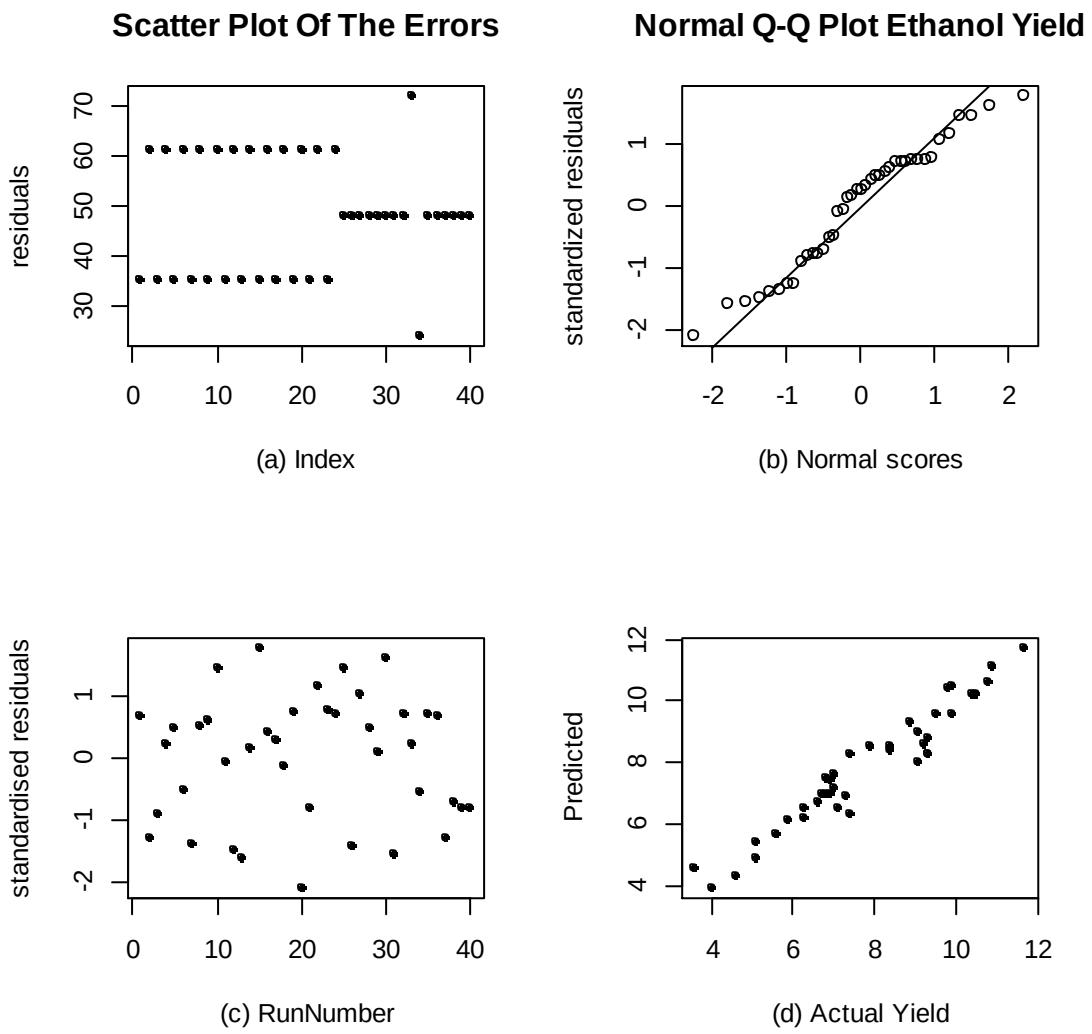


Figure 4.1: Plots of Residuals

In Figure 4.1(b), the normal probability plot of standardized residuals clearly indicates a straight line and in figure 4.1(c), the standardized residuals are plotted against the run numbers and again the plot indicates that the points are randomly scattered within the constant range of residuals across the graph thus the model is adequate and finally the plot of predicted values against the experimental values figure 4.1(d) indicates a strong positive correlation and therefore there is no reason to suspect violation of independence or constant variance assumption.

ii). Regression Analysis and Anova

Analysis of the fitted model to test whether the model adequately approximates bioethanol yield well was carried out using the anova and regression analysis whose R-out puts are displayed in tables 4.7 and 4.8

Table 4.7: Call: Rsm (Formula = Y ~ So(X1, X2, X3, X4), Data = D3)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.82132	1.71799	6.8809	3.266e-07 ***
X1	0.51267	0.10282	4.9861	3.864e-05 ***
X2	-0.26866	0.10282	-2.6129	0.014978 *
X3	0.99367	0.1028	9.6641	6.360e-10 ***
X4	0.10893	0.10282	1.0594	0.299524
X1:X2	-0.75903	0.12573	-6.0371	2.628e-06 ***
X1:X3	0.86539	0.12573	6.8831	3.249e-07 ***
X1:X4	0.21272	0.12573	1.6919	0.103088
X2:X3	0.44478	0.12573	3.5377	0.001606 **
X2:X4	1.11679	0.12573	8.8827	3.311e-09 ***
X3:X4	-0.37226	0.12573	-2.9609	0.006632 **
X1^2	-1.02999	0.43578	-2.3635	0.026185 *
X2^2	-0.62737	0.43578	-1.4396	0.162377
X3^2	-1.21016	0.43578	-2.7770	0.010248 *
X4^2	-1.13815	0.43578	-2.6117	0.015017 *

--Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.Multiple R-squared: 0.9323, Adjusted R-squared: 0.8944 F-statistic: 24.59 on 14 and 25 DF, p-value: 3.692e-11.Call: Analysis of variance showed that magnitude of the F-value 24.59 and the low probability value (< 0.0001), as proof of the significant model fit. Using the $t-$test statistics and 5% level of significance, the regression analysis indicates that

out of the four control variables, incubation time (X_1) and incubation temperature (X_3)were very positively significant to the yield of bioethanol with their $P-$values of their $t-$statistics being far much less than 0.001 the level of significance while initial PH (X_2) had a significant negative effect at $P-$value of 0.01 while the concentration of the substrate was not significant at $\alpha = 5\%$. Therefore increase in incubation time and temperature resulted in increased bioethanol yield while increase in initial PH resulted in decrease of bioethanol production as indicated by the signs of their parameters in the model. The interactive effect of incubation temperature and time, initial PH and substrate concentration, had significant synergetic effect on bioethanol yield while time and pH had an antagonistic effect significant at 0.001 interactive effect of temperature and pH was also significant at 0.01implying that increase in both incubation temperature and initial PH resulted in increase in ethanol yield. The intercept term of the model was significant at P-value of 0.001 while the quadratic effects of time, temperature and substrate concentration were all significant at a P-value of 0.05 affecting ethanol yield antagonistically. The $t-$statitics values could also have been determined analytically using equation (3.57) as follows

$$t_{X_1} = \frac{0.51267}{\sqrt{0.4227 \times 0.025}} = 4.987142 \qquad (4.8)$$

The reduced model after removing the non-significant terms become

$$\hat{Y} = 11.82 + 0.51X_1 - 0.27X_2 + 0.99X_3 + 0.11X_4 - 0.76X_1X_2 + 0.87X_1X_3 + 0.44X_2X_3 + 1.12X_2X_4 - 0.37X_3X_4 - 1.03X_1^2 - 1.21X_3^2 - 1.14X_4^2 \qquad (4.9)$$

Equation (4.9) was used to obtain the estimated value of ethanol yield in table 4.6

To test the adequacy of the fitted model, analysis of variance and F-test were employed. The Anova for full second order model is given in Table 4.8

Table 4.8: Analysis of Variance (Anova)

Response:Y	D.o.f	Sum Sq	Mean Sq	F value	Pr(>F)
FO(X1, X2, X3, X4)	4	53.346	13.3364	31.5516	1.940e-09
TWI(X1, X2, X3, X4)	6	78.988	13.1646	31.1451	1.944e-10
PQ(X1, X2, X3, X4)	4	13.157	3.2894	7.7821	0.0003228
Residuals	25	10.567	0.4227		
Lack of fit	25	10.567	0.4227		
Pure error	0	0.000			

The results indicate that the F-ratio for main effects of the incubation time, initial PH, incubation temperature and substrate concentration was 31.5516 which is more than the F-table value at 5% percent level of significance with 4 and 35 degrees of freedom i.e table $F(5\%, 4,35) = 2.69$, therefore we reject the the null hypothesis ($H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \; vs \; H_1: \beta_j \neq 0$ for at least one of the regression parameters) of equality of the regression parameters at 5% level of significance likewise for the two terms interaction and quadratic regression parameters the null hypothesis ($H_0 = \beta_{12} = \beta_{13} = \beta_{14} = \beta_{23} = \beta_{24} = \beta_{34} = 0, H_0: \beta_{11} = \beta_{22} = \beta_{33} = \beta_{44} = 0$) respectively of the equality of the parameters is rejected since the two F-ratio values i.e. 31.1451 and 7.7821 have their P-values 1.944e^{-10} and 0.0003228 which were both less than $\alpha = 5\%$ and the model is therefore considered adequate and since the second order terms of two-way interactions (TWI) and polynomial quadratics (PQ) terms contributed significantly to the model, canonical analysis was necessary in order to determine how well the estimated model fits the data, R-squared value was used. The value of the Multiple R-squared was 0.9323 and Adjusted R-squared was 0.8944 which measure model fitting reliability and for the fitted model,

R-squared of 0.9323 indicates aptness of the model, Bhunia and Dey, (2012). Also most values of (Probability F) are less than 0.05, which confirms that the model terms are significant. Therefore only 6.8% of the total variation could not be explained by the model which ensures good adjustment of the model to experimental data. Model adequacy was also confirmed by the good agreement between the experimental data and predicted data as shown in Table 4.6.

4.4 Optimization of Ethanol Yield

4.4.1 Graphical Analysis

For more understanding of ethanol yield under optimum conditions and interaction between the various independent variables within the given ranges, the second order model was used to build images, contours and response surface plots by RSM (response surface method). The three dimensional plots were built by fixing two of independent variables in their midpoint i.e. their stationary values and changing the other two variables over their experimental range. The resulting graphics show clear effects of initial incubation time, pH value, fermentation temperature and concentration of substrate on ethanol production. The effect of fermentation time and pH-value on the bioethanol yield, as a contour plot, image and response surface are presented by Fig. 4.2, 4.3 and 4.4 respectively.

i). Contour Plots

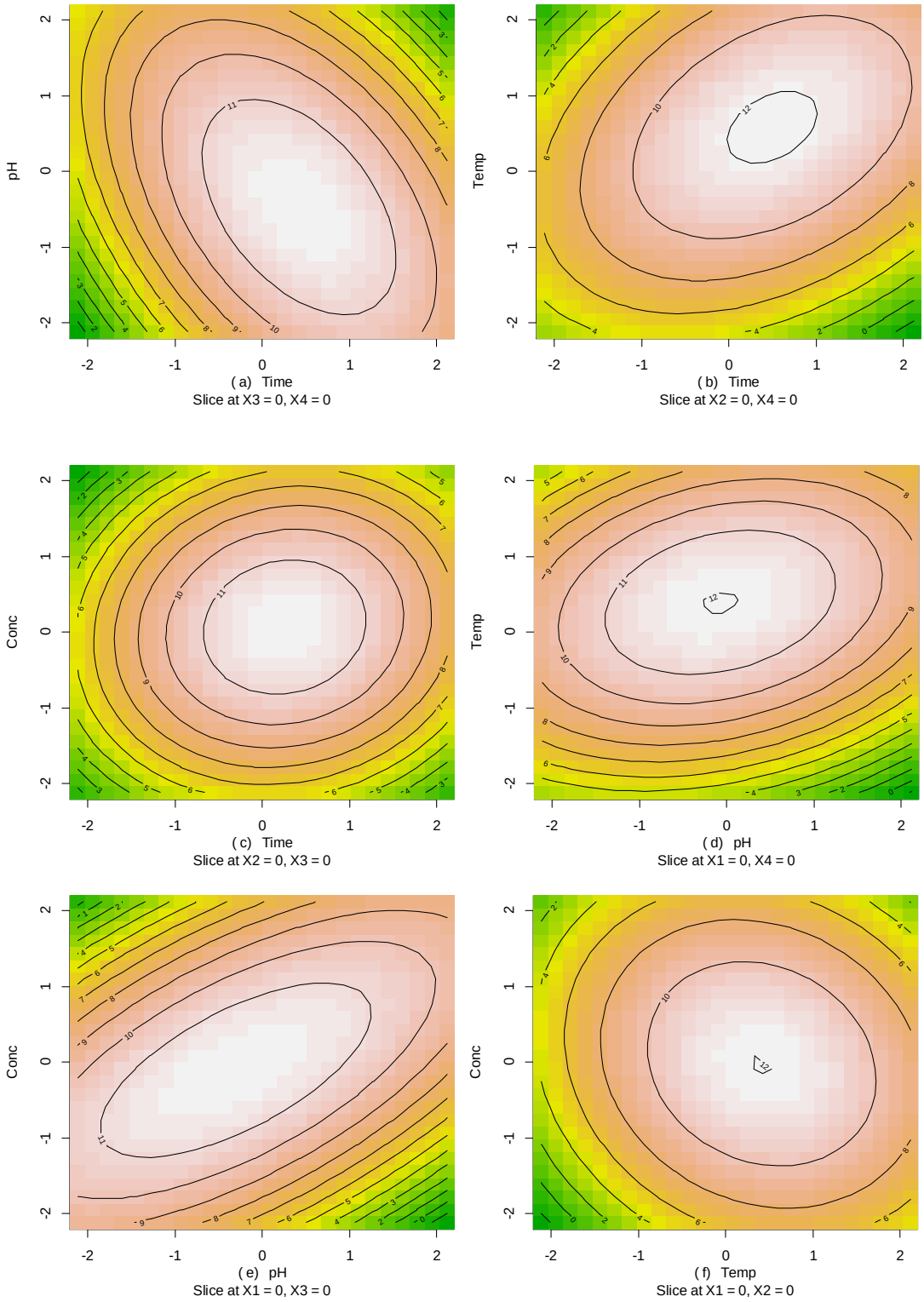


Figure 4.2: The Contour Plots for Bioethanol Yield

The response surface model was used to predict the ethanol yield by contour plots. A Contour plot is the projection of the response surface as a two dimensional plane Box

and Hunter, (1957). The shapes of contour plots indicate the nature and extent of the interaction between different factors Prakash et al., (2017),where less prominent or negligible interactions were shown by the circular nature of the contour plots, while comparatively prominent interactions were otherwise shown by the elliptical nature of the contour plots. The contour plots developed using the fitted quadratic polynomial equation obtained from regression analysis are in Figure 4.2(a-f). Each figure presents the effect of two variables on the production of bioethanol, while the other two variables are held at zero level Liu et al., (2013). Figure 4.2 (f) shows the effects of varying temperature X_3 and substrate concentration X_4 while fermentation time and pH are held constant at the stationary point. The enlarged contour Figure 4.2 (f) with colors is depicted in Appendix Figure C.

ii). Images of Yield

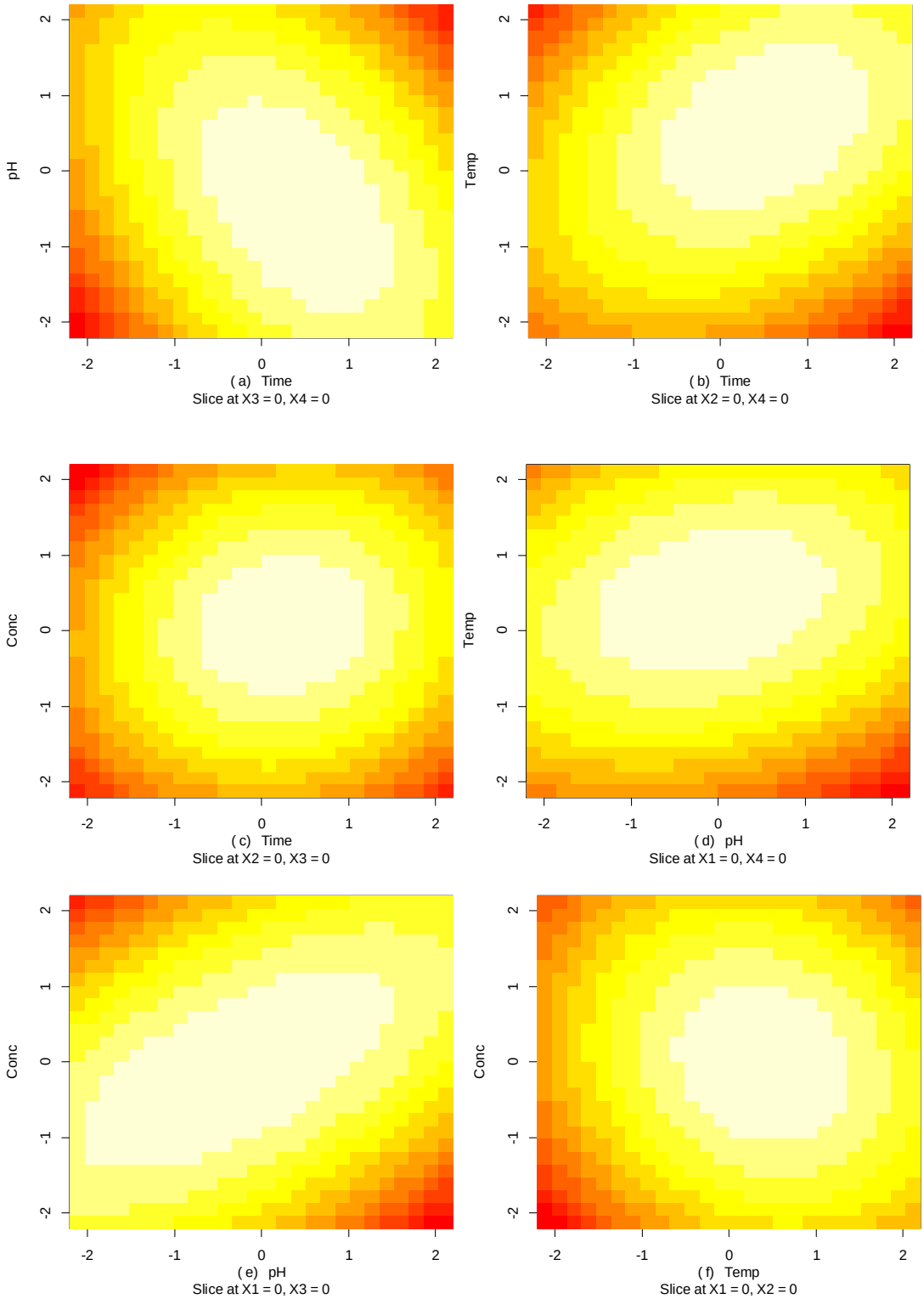


Figure 4.3: Images of Bioethanol Yield

iii). Response Surface Plots

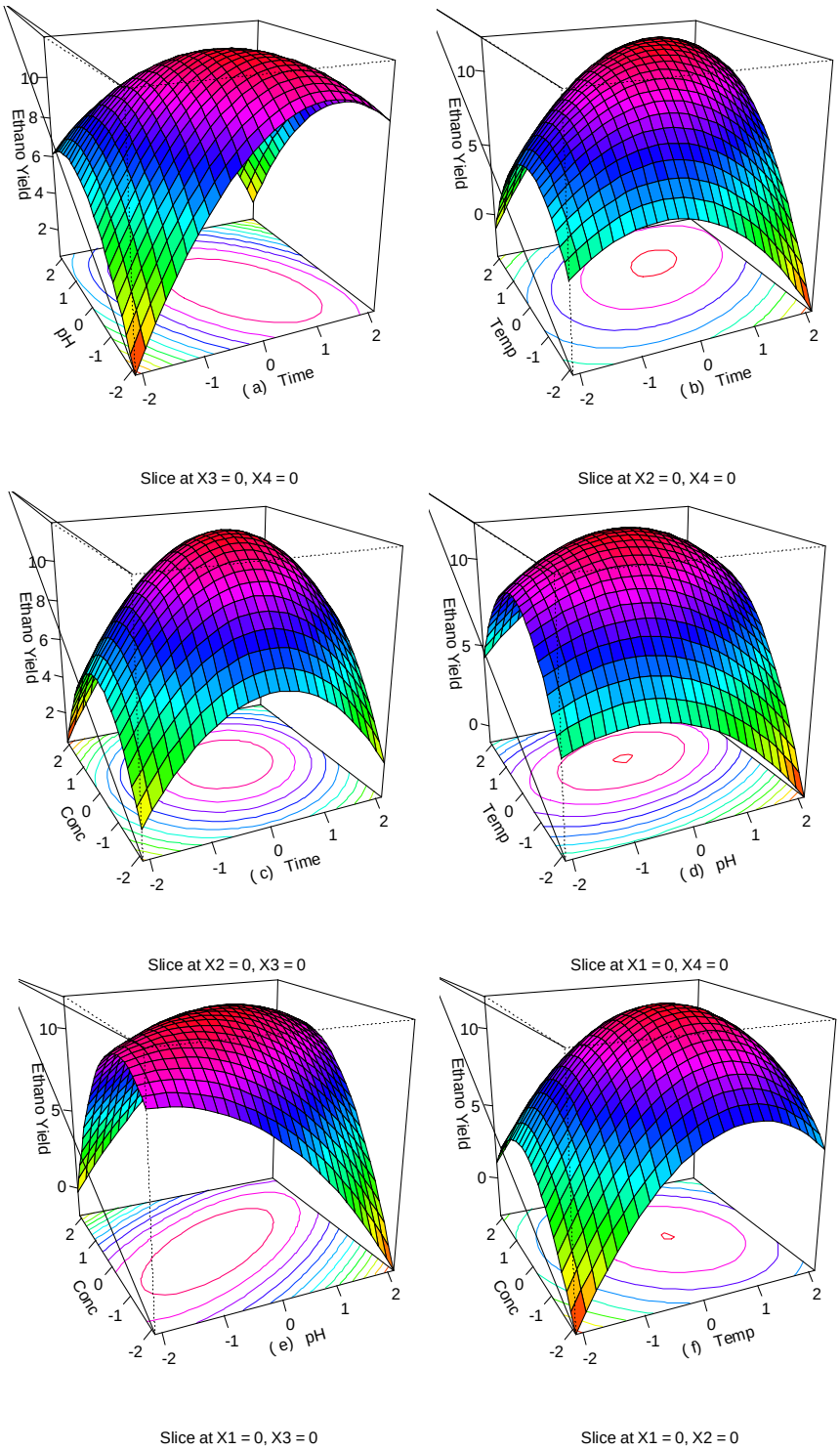


Figure 4.4: Response Surface Plots General Design

4.4.2 Analytical Determination of the Optimum Settings

Characterization of the Response Surface was achieved by determining the stationary point of the response surface using computer software R programming.

Table 4.9: The Stationary Point of Response Surface:

X_1	X_2	X_3	X_4
0.7455032	-0.7762284	0.5897951	-0.3597595

The conditions of the stationary point were determined analytically as follows

$$X_s = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \qquad b = \begin{bmatrix} 0.51267 \\ -0.26866 \\ 0.99367 \\ 0.10893 \end{bmatrix}$$

$$\beta = \begin{bmatrix} -1.02997 & -0.75903/2 & 0.86539/2 & 0.21272/2 \\ -.75902/2 & -0.62737 & 0.44478/2 & 1.11679/2 \\ 0.86539/2 & 0.44478/2 & -1.21016 & -0.37226/2 \\ 0.21272/2 & 1.11679/2 & -0.37226/2 & -1.13815 \end{bmatrix}$$

The derivative of $\hat{Y}$ with respect to vector X equated to zero was obtained using equation (3.62) and therefore the stationary point is given as equation (3.63)

$$X_s = \begin{bmatrix} 0.745503 \\ -0.7762284 \\ 0.589751 \\ -0.3597595 \end{bmatrix} = -0.5 \times \begin{bmatrix} -1.029990 & -0.379515 & 0.432695 & 0.106360 \\ -0.379515 & -0.62737 & 0.222390 & 0.558395 \\ 0.432695 & 0.22239 & -1.21016 & -0.186130 \\ 0.106360 & 0.558395 & -0.186130 & -1.138150 \end{bmatrix}^{-1} \times \begin{bmatrix} 0.51267 \\ -0.26866 \\ 0.99367 \\ 0.10893 \end{bmatrix}$$

$$B^{-1} = \begin{bmatrix} -1.5612644 & 1.29560157 & -0.4056709 & 0.5560856 \\ 1.2956016 & -4.02630437 & 0.0087569 & -1.8557290 \\ -0.4056709 & 0.00875697 & -0.9894949 & 0.1282058 \\ 0.5560856 & -1.85572902 & 0.1282057 & -1.7580701 \end{bmatrix}$$

And the predicted response at this stationary point is given by equation (3.64) as

$$12.3901 = 11.82132 + 0.5 \times \begin{bmatrix} 0.745503 & -0.776228 & 0.589751 & -0.3597595 \end{bmatrix} \times \begin{bmatrix} 0.51267 \\ -0.26866 \\ 0.99367 \\ 0.10893 \end{bmatrix}$$

Which gives a yield of 12.3901g/L of bioethanol. The stationary point gives the coded levels of the control variables i.e. $X_1 = 0.745503$ is the coded level of incubation time at the stationary point. While -0.776228 , $0.589751\ and\ -0.3597595$ are the coded levels of the initial pH, incubation temperature and the substrate concentration respectively. The natural levels of the factors were obtained as follows:

For the incubation time, the natural level at the stationary point was $0.745503 = \frac{\xi_t - 48}{48/_{2.117\times2}} \rightarrow \xi_t = 56.45162 \cong 56.5$ hours of incubation and for the initial PH, the natural level was $-0.776228 = \frac{\xi_h - 5.5}{\frac{3}{2.117\times2}} \rightarrow \xi_{ph} = 4.949995 \cong 4.95$ and similarly for incubation temperature, the natural level at the stationary point was $0.589751 = \frac{\xi_l - 32.5}{\frac{15}{2.117\times2}} \rightarrow \xi_{tp} = 34.5903332.8^oC \cong 34.6^0C$ and finally the substrate concentration in natural variable was $-0.3597595 = \frac{\xi_m - 30}{\frac{20}{2.117\times2}} \rightarrow \xi_c = 28.30062 \cong 28.3$g/L and with a production of 12.39g/L of alcohol, it is worthwhile to note that $\frac{12.39\text{g/L}}{28.3\text{g/L}}$ this is an yield of 0.4377996g of ethanol per gram of the substrate which compares well with most other findings in literature since it translates to 85.6% of the theoretical yield, where 0.511g of ethanol per gram of substrate is the theoretical yield as proposed by Tropea et al., (2014) at the stationary point. The response surface was characterized by expressing the fitted model in canonical form as equation (4.8)

$$\hat{Y} = \hat{Y}_s + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \lambda_3 w_3^2 + \lambda_4 w_4^2 \qquad (4.8)$$

Where λ_i for $i = 1,2,3,4.$ are simply the roots of $\| \beta - \lambda I \|$ equated to zero with

$\lambda = \begin{bmatrix} -0.1795987 \\ -0.7201155 \\ -1.1548720 \\ -1.9510952 \end{bmatrix}$ being the vector of Eigen values. The canonical form of the fitted model is given in equation (4.9)

$$\hat{Y} = 12.390 - 0.17959w_1^2 - 0.72012w_2^2 - 1.15487w_3^2 - 1.95109w_4^2 \qquad (4.9)$$

Since all the Eigen values are negative the stationary point is a maximum point within the region of exploration. The relationship between the canonical variables $\{w_i\}$ and the design variables $\{x_i\}$ was determined .This is necessary when it is difficulty to run the experiment at the stationary point. The process variables are related to canonical variables by $W' = M'(X - X_s)$; Where the columns of M are the normalized Eigen vectors associated with λ_i ie if m_i is the i^{th} column of matrix M, then m_i is the solution to $(\beta - \lambda I)m_i = 0$ for which $\sum_{i=1}^{k} m_i^2 = 1$.

$$M = \begin{bmatrix} 0.3342068 & 0.7031772 & 0.3793281 & -0.4999579 \\ -0.82565661 & 0.2194700 & -0.2670610 & -0.4458727 \\ 0.04385817 & 0.6513284 & -0.5200101 & 0.5508515 \\ -0.45240843 & 0.1820609 & 0.7172016 & 0.4977975 \end{bmatrix}$$

So the relationship between the W and X variables is

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \begin{bmatrix} 0.3342068 & 0.7031772 & 0.3793281 & -0.4999579 \\ -0.82565661 & 0.2194700 & -0.2670610 & -0.4458727 \\ 0.04385817 & 0.6513284 & -0.5200101 & 0.5508515 \\ -0.45240843 & 0.1820609 & 0.7172016 & 0.4977975 \end{bmatrix}' \times \begin{bmatrix} x_1 - 0.7455032 \\ x_2 + 0.7762284 \\ x_3 - 0.589751 \\ x_4 + 0.3597595 \end{bmatrix}$$

Implying that if w_1 is required, the equation (4.10) is solved

$$w_1 = 0.33(x_1 - 0.75) + 0.70(x_2 + 0.78) + 0.38(x_3 - 0.59) - 0.50(x_4 + 0.36) \qquad (4.10)$$

To explore the response surface in the vicinity of the stationary point, the appropriate points at which to take observations in the $w_i's$ space are obtained as such and then equation (4.8) was used to obtain $x_i's$ so that the runs may be made. As a confirmation test, one experiment of fermentation using the optimized conditions was conducted to

validate the model. Three samples of the diluted ferment were set up for oxidation experiment and an ethanol yield of $11.6 \pm .15$ g/L was obtained. This actual value closely agreed with the predicted value, with a difference of only 0.8%. Hence, it confirmed that the model developed from the response surface methodology reliably predicted ethanol production. According to Irshad and Asgher, (2011) differences between experimental and predicted values of less than 10% confirm the validity of a model. This actual value was in close agreement with the predicted value with difference of less than 10%. Hence, the model developed from the response surface methodology was confirmed and therefore the model reliably predicts ethanol yields. The "steepest" function in R which computes the curved path of the steepest ascent based on the ridge analysis was employed to further study ethanol yield at the point of maximum yield table 4.10 is the graph of steepest Ascent.

Table 4.10: Path of Steepest Ascent:

Run	Dist.	X1	X2	X3	X4 | yhat
1.	0.0	0.000	0.000	0.000	0.000 | 11.821
2.	0.5	0.307	-0.176	0.350	-0.047 | 12.220
3.	1.0	0.607	-0.540	0.534	-0.234 | 12.371
4.	1.5	0.846	-0.966	0.623	-0.462 | 12.379
5.	2.0	1.053	-1.395	0.680	-0.695 | 12.279
6.	2.5	1.246	-1.821	0.722	-0.927 | 12.082
7.	3.0	1.431	-2.244	0.758	-1.158 | 11.791
8.	3.5	1.612	-2.665	0.791	-1.388 | 11.407
9.	4.0	1.788	-3.082	0.820	-1.616 | 10.935
10.	4.5	1.964	-3.503	0.848	-1.846 | 10.366
11.	5.0	2.138	-3.919	0.875	-2.074 | 9.711

After the midpoint range the bioethanol yield decreased despite the fact that all variables were increasing. This was due to growth-inhabiting effect of high sugar concentrations, high pH levels, high temperature, as well as ethanol-formation which with time raised its levels thereby distorting microorganisms (yeast) causing what is referred to as "metabolism-poisonous effect", Thatipamala et al., (1992). The point of maximum ethanol yield from ridge analysis lies between the fourth and fifth points

which depicts the greatest yield of 12.379 g/L at the fourth point from table (4.10) followed by a decrease at the fifth point to12.279g /L. The variables setting in coded form at the fourth point are 0.846, -0.966, 0.623 and -0.462 for the time, initial PH, incubation temperature and Substrate concentration, from the design center respectively which corresponds to 57.6hours,pH of 4.8,34.7^{0}C and a concentration of 27.82g/L of the substrate in natural levels .The path of steepest ascent is shown in Figure 4.5.

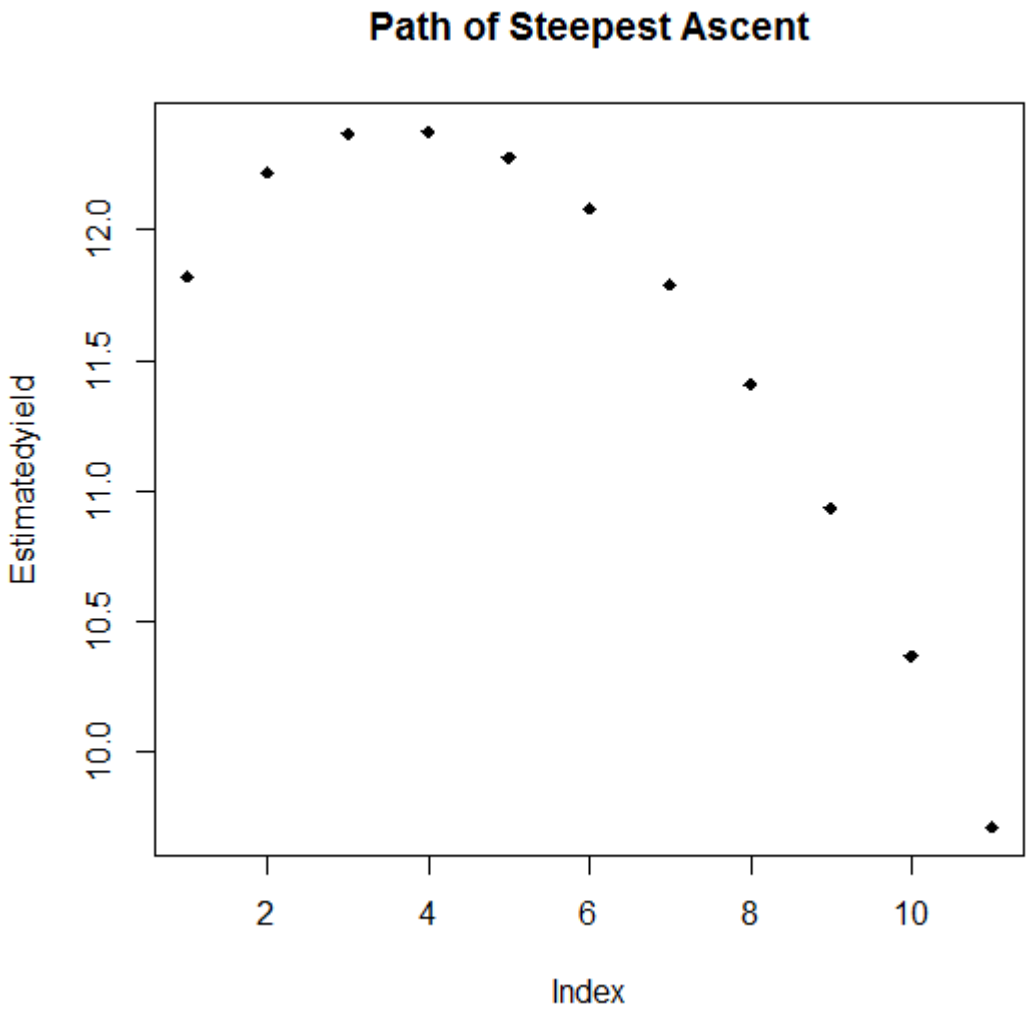


Figure 4.5: Path of Steepest Ascent Plot

4.4.3. Optimization using the E −Optimal Design

The most efficient design was the E −optimal with a relative efficiency the general to the E − optimal design being equal to 1%. E − $o$ptimal design required only 32 runs i.e only factorial part of the design was necessary but this design was found to be inadequate in estimating parameters of the full second order model since some coefficients were aliased and it was not possible to use function 'rsm' instead it returned a linear model. But augmenting the design with one center point enabled fitting of the full second order model without factors aliasing. Table 4.11 is the regression analysis for the full second order model using the E-optimal design

Table 4.11: Regression Analysis for the E −Optimal Design

Estimate		Std. Error	t value	Pr(>|t|)
(Intercept)	11.82100	0.64028	18.4623	3.815e-13 ***
X1	0.42876	0.11495	3.7300	0.0015326 **
X2	-0.27118	0.11495	-2.3592	0.0298177 *
X3	1.05541	0.11495	9.1816	3.268e-08 ***
X4	0.10627	0.11495	0.9245	0.3674353
X1:X2	-0.75903	0.12382	-6.1302	8.642e-06 ***
X1:X3	0.86539	0.12382	6.9892	1.586e-06 ***
X1:X4	0.21272	0.12382	1.7180	0.1029487
X2:X3	0.44478	0.12382	3.5922	0.0020830 **
X2:X4	1.11679	0.12382	9.0195	264e-08 ***
X3:X4	-0.37226	0.12382	-3.0065	0.0075773 **
X1^2	-1.29141	0.22606	-5.7127	2.041e-05 ***
X2^2	-0.58556	0.22606	-2.5903	0.0184723 *
X3^2	-1.10770	0.22606	-4.9000	0.0001153 ***
X4^2	-1.02068	0.22606	-4.5151	0.0002680 ***

---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1,Multiple R-squared: 0.95, Adjusted R-squared: 0.911 , F-statistic: 24.41 on 14 and 18 DF, p-value: 7.464e-09

Stationary point of response surface: The F −value of 24.41 with a P −value of 7.464e-09 which is almost negligible indicates that the model is very significant. From the Table 4.11, the intercept term and incubation temperatures, the interactive effects of initial pH and incubation time, time and temperature, pH and concentration as well as the quadratic effects of time, temperature and concentration were all very

significant at P-values 0.001 While the incubation time and the interactions effects of pH and time as well as time and concentration were significant at P-values of 0.01 and with a multiple R-squared of 0.95,only 0.05 variations of Y around the mean $\bar{Y}$ is not explained by the model hence as expected the model by the $E-$optimal design achieves better parameters estimations with the least number of runs. The final model for the E-optimal design was

$$\hat{Y} = 11.82 + 0.43X_1 - 0.27X_2 + 1.05X_3 + 0.11X_4 - 0.76X_{12} + 0.87X_{13} + 0.44X_{23} + 1.12X_{24} - 0.37X_{34} - 1.29X_{11} - 0.59X_{22} - 1.11X_{33} - 1.02X_{44} \quad (4.11).$$

Table 4.12: Analysis of Variance Table

Response: Y1:	D.f	Sum Sq	Mean Sq	F value	Pr(>F)
FO(X1, X2, X3, X4)	4	42.896	10.7240	26.159	2.744e-07
TWI(X1, X2, X3, X4)	6	78.988	13.1646	32.112	1.137e-08
PQ(X1, X2, X3, X4)	4	18.217	4.5543	11.109	0.0001019
Residuals	18	7.379	0.4100		
Lack of fit	18	7.379	0.4100		
Pure error	0	0.000			

The first order terms, the two-way interaction terms and the quadratic terms of the model were all significant at $\alpha = 5\%$ with the $P-$values of their $F-$ratios being less the0.05. Stationary point of response surface in coded levels of the variables were X_1= 0.5600981, X_2= -0.7591368, X_3= 0.6127829 and X_4= -0.4166317.The corresponding levels in actual variables were $X_1 = 54.35$ hours of incubation time, $X_2 = 4.96$ initial pH, $X_3 = 34.67$ hrs. and $X_4 = 28.03 g/L$ substrate concentration for an yield of 12.35g /L of ethanol which is slightly less than the estimated maximum value by the general design by $0.04 g/L$.But the yield achieved by the E-optimal design was slightly higher at 86.22% of the theoretical yield as

compared to 85.6% of the general design by 1.00724%. Hence this design would be preferred to general design since it achieves higher percentages of ethanol yield at a lower number of experimental runs.

Eigen analysis: From the R out- put, the Eigen values were

[-0.1440843, -0.7681515, -1.1212032, -1.9719169] all of them negative hence once more a point of maximum yield is suggested and confirmed by the ridge analysis for the steepest ascent. The R-output for the ridge analysis for the E −optimal design is given in Table 4.13

Table 4.13: Ridge Analysis E-Optimal

	Distance	X1	X2	X3	X4 | yhat
1.	0.0	0.000	0.000	0.000	0.000 | 11.821
2.	0.5	0.253	-0.172	0.391	-0.060 | 12.215
3.	1.0	0.488	-0.576	0.580	-0.305 | 12.337
4.	1.5	0.656	-1.032	0.645	-0.583 | 12.328
5.	2.0	0.798	-1.476	0.677	-0.852 | 12.232
6.	2.5	0.930	-1.911	0.697	-1.117 | 12.058
7.	3.0	1.058	-2.342	0.712	-1.378 | 11.808
8.	3.5	1.181	-2.767	0.724	-1.635 | 11.486
9.	4.0	1.303	-3.191	0.734	-1.892 | 11.090
10.	4.5	1.423	-3.612	0.743	-2.147 | 10.623
11.	5.0	1.545	-4.038	0.752	-2.405 | 10.075

The optimal bioethanol yield was at the third point of 12.337g/L.

i). Images for the E-Optimal Design

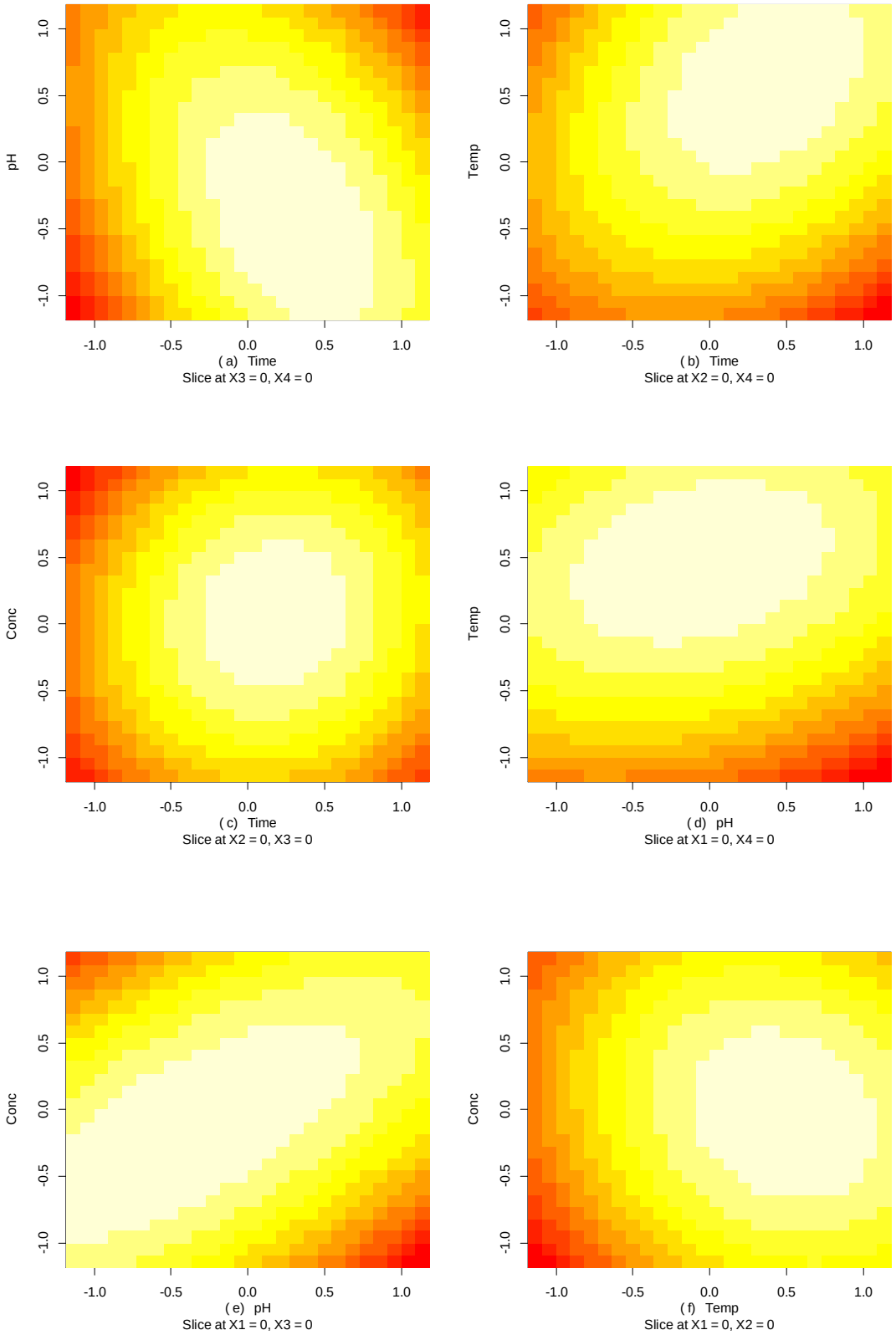


Figure 4.6: Images of Yield for E-Optimal Design

ii). The Contour Plots E-Optimal

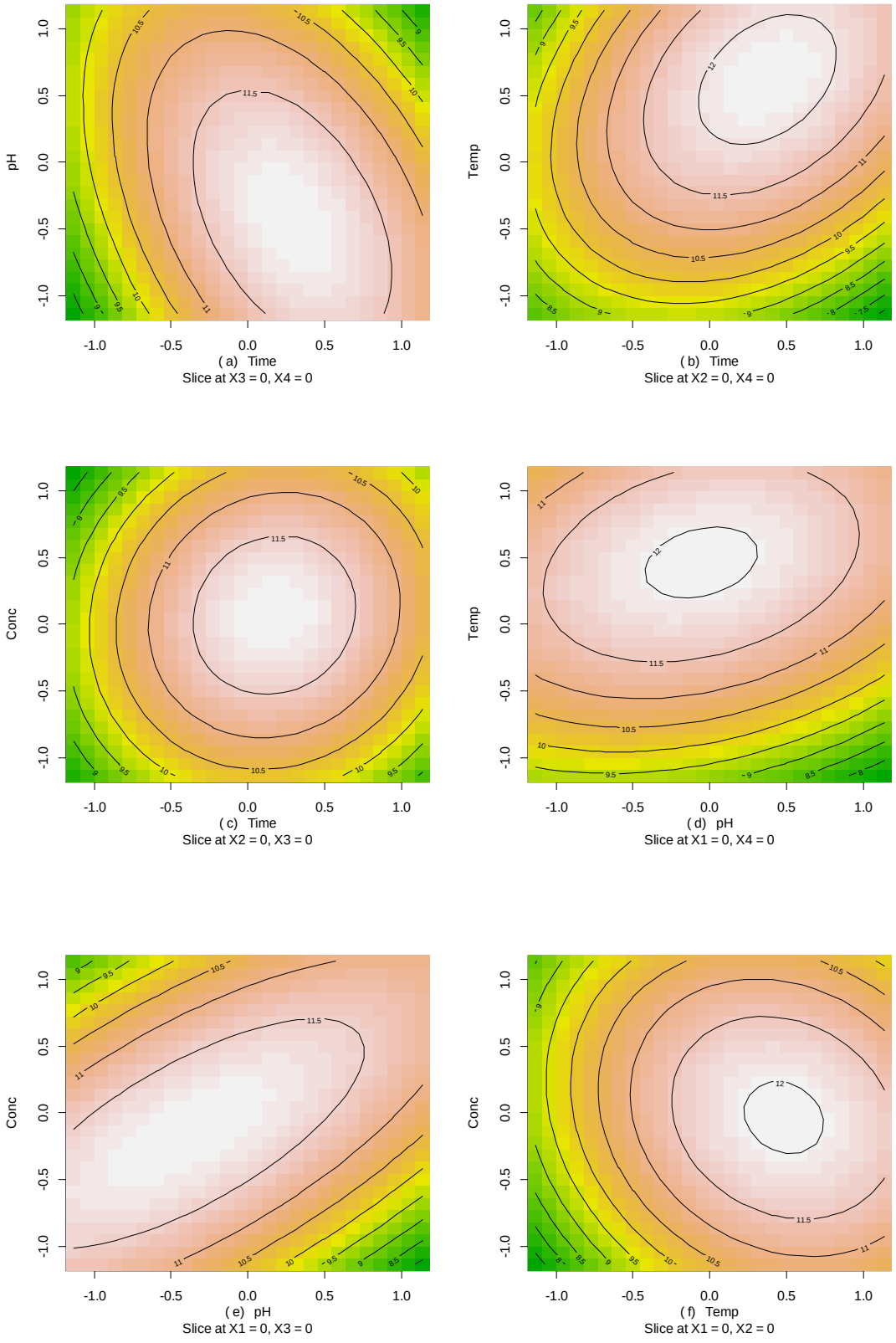


Figure 4. 7: Contour Plots from the E-Optimal

In the contour plots of figure 4.7, when variables other than those on the coordinate axes are involved, the display is a slice of the response surface, holding the other variables fixed at certain values. By default, the averages of the numeric predictors and their first levels are used. The contour plots in figure 4.7(a) show a slice of the response surface when substrate concentration (X_4) and incubation time (X_1) are held constant at -0.42 and 0.56 the interest is in the behavior of the response in the neighborhood of the stationary point.

iii). Surface Plots E-Optimal

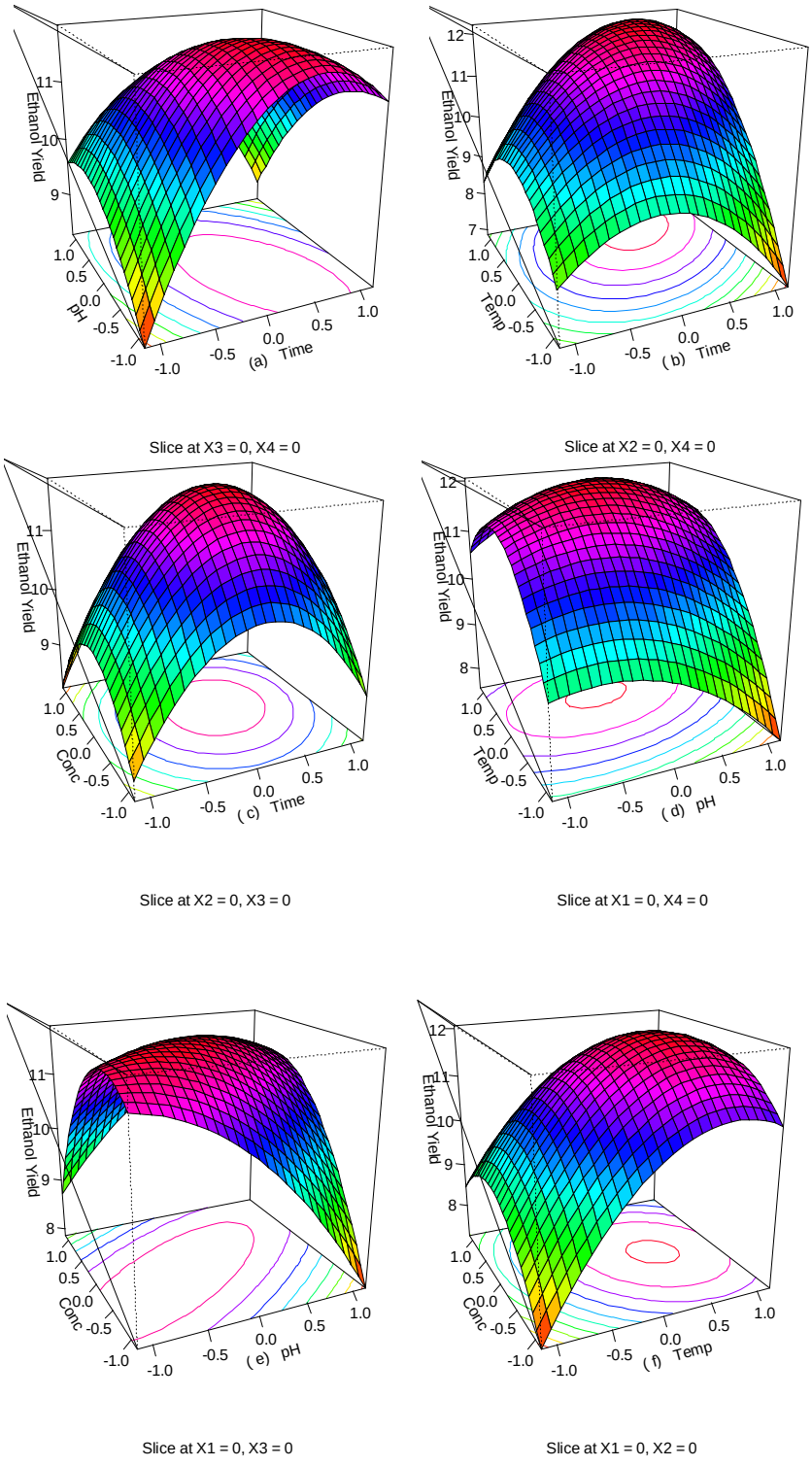


Figure 4.8: Response Surface Plot E-Optimal Design

The response surface plots in Figure 4.8 depicts the changes in response when two regression variables are varied while two more others are held constant at the design

center. For example, in figure 4.8(a), yield increases with increase in both incubation time and initial pH up to a point where incubation time is -0.76 and pH is 0.56 while incubation temperature and substrate concentrations are equal to zero for a yield of 11.55g/L. Further increase in both time and pH results in decreased yield for example when time is increased to -0.90 from -0.76 and pH coded level decreases to 1 from 0.56 while temperature and concentration are still zero in coded levels, the yield decreases to 11.45597g/L.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.0 Introduction

This chapter gives summary, conclusion and recommendations of the findings of this research and also suggests areas for further research which have emerged during the course of this study.

5.1 Conclusion

The $D-, A-, E-and\ T-$optimal values of the general design were obtained as well as the relative efficiencies of the general second order design to the optimal designs. The $E-$ optimal was a design of 32 runs and it was the most the efficient design over the general design with a relative efficiency of one Wald, (1943). An application to a four factor experiment was fit to the design to determine the optimal conditions for ethanol yield where incubation time, initial pH, incubation temperature and substrate concentrations, were the variables under investigation. The ethanol yield increased with increase in the process variables up to a certain point before ethanol produced started decreasing as indicated by the plot of the path of steepest ascent and the response surface plots. This was due to "metabolic poison effect" and an optimal yield of 12.35g/L of ethanol was realized at factor settings of 0.5600981,-0.7591368,0.6127829,-0.4166317 for time, pH, temperature and substrate concentration respectively in coded levels translating to 54.35hours, 4.96 level of pH, 34.67^0C temperature and 28.03g/L of substrate concentration of factors in natural levels for the $E-$ optimal design while the an optimal yield of 12.390g/l of ethanol was realized using the general design at factor settings in coded and natural levels as 0.75,-0.78,0.59,-0.36 and 56.45hours, 4.95 pH level, 34.59^0C level of temperature and 28.30g/l of substrate concentrations respectively. The yield by the E-optimal design

was noted to be slightly lower than the yield obtained using the general design by 0.04488981g/L. Response surface methodology and the rotatable design constructed using balanced incomplete block design for four factors with $r < 3\lambda$ was found reliable in modeling, optimizing and studying the effects of the factors and their interaction to the process of fermentation of pineapples peels waste as substrate for ethanol production using Saccharomyces cerevisiae. The high values of coefficient of determination R^2 of 0.95 and adjusted R-squared of 0.911 for the E-optimal design indicated that the model fitted the data well. The yields of 12.35g/L of ethanol for a substrate concentration of 28.03g/L translates into 0.441g of ethanol per gram of substrate which compares well with many other findings in literature from similar studies. This was roughly 86% of the theoretical yield. Theoretical yield in this study was calculated as the maximum ethanol yield in relation to dry matter, (0.511 g alcohol per 1.0 g dry matter of substrate as proposed by Tropea *et al*., (2014).The design was found reliable in modeling, optimizing and studying effects of the factors to the processes of fermentation of pineapples peels for ethanol production. The study established the factor settings that yield maximum ethanol from pineapple peels. These wastes if not properly disposed can be a major source of pollution. A cheaper fuel than fossil fuel is provided while managing wastes.

5.2 Recommendations

The design can be applied to any four factor experiment provided the appropriate coded values and actual value transformations are made. Further investigation of the $E-$optimal design is recommended to determine the number of center points that will give it the optimal value of ethanol yield during fermentation of pineapple peels. A comparison of the results obtained using this design and the result of rotatable design with four factors constructed using balanced incomplete block design when

replications (r) are more than three the number of times (λ) pairs of treatments occur together applied in fermentation of pineapple peels is also suggested. Investigation of optimal settings of incubation time and temperature as well as initial pH levels at constant substrate concentrations is also recommended since the full second order model fitted indicated that the concentrations of pineapple peels were not significant at all as a main factor as well as quadratic term and the corresponding optimal yield of bioethanol at these optimal settings should be investigated. Further suggestion is made on increasing ethanol yield using pineapple peels by employing enzymatic hydrolysis to increase percentages of fermentable sugars after pre-treatment as well as determination of the amount of ethanol yielded from fermentation samples through distillation, since this would increase the precision with which ethanol is determined.

REFERENCES

Abdullah. (2007). Solid and Liquid Pineapple Waste Utilization for Lactic Acid Fermentation. *Reaktor*, *11*(1), 50–52.

Adnan, N. A. A., Suhaimi, S. N., Abd-Aziz, S., Hassan, M. A., & Phang, L. Y. (2014). Optimization of bioethanol production from glycerol by Escheric h ia coli SS1. *Renewable Energy*, *66*, 625–633. https://doi.org/10.1016/j.renene.2013.12.032

Almarsdottir, A. R., Sigurbjornsdottir, M. A., & Orlygsson, J. (2012). Effect of various factors on ethanol yields from lignocellulosic biomass by Thermoanaerobacterium AK 17. *Biotechnology and Bioengineering*, *109*(3), 686–694. https://doi.org/10.1002/bit.24346

Atkinson, A. and Donev, A. (1992). *Optimum experimental designs.* Oxford University Press.

Atkinson, A. C., & Fedorov, V. V. (1975). The design of experiments for discriminating between two rival models. *Biometrika*, *62*(1), 57–70. https://doi.org/10.1093/biomet/62.1.57

Ban-Koffi, L., & Han, Y. W. (1990). Alcohol production from pineapple waste. *World Journal of Microbiology & Biotechnology*, *6*(3), 281–284. https://doi.org/10.1007/BF01201297

Barta, Z., Reczey, K., & Zacchi, G. (2010). Techno-economic evaluation of stillage treatment with anaerobic digestion in a softwood-to-ethanol process. *Biotechnology for Biofuels*, *3*, 1–11. https://doi.org/10.1186/1754-6834-3-21

Bhunia, B., & Dey, A. (2012). Statistical Approach for Optimization of Physiochemical Requirements on Alkaline Protease Production from Bacillus licheniformis NCIM 2042 . *Enzyme Research*, *2012*, 1–13. https://doi.org/10.1155/2012/905804

Bose, R. C., & Draper, N. R. (1959). Second Order Rotatable Designs in Three Dimensions. *The Annals of Mathematical Statistics*, *30*(1097–1112), 1097–1112. https://doi.org/10.1214/aoms/1177706093

Box, G. E. ., & Wilson, K. . (1951). On The Experimental Attainment of Optimum Conditions. *Journal Royal Statistics, Soc.*, *13*(1), 1-45.

Box, G. E. P., & Behnken, D. W. (1960). Some New Three Level Designs for the Study of Quantitative Variables. *Technometrics*, *2*(4), 455–475. https://doi.org/10.1080/00401706.1960.10489912

Box, G., & Hunter, J. (1957). Multi-Factor Experimental Designs for Exploring Response Surfaces Author (s): G . E . P . Box and J . S . Hunter MUJLTI-FACTOR RESPONSE SURFACES ’. *The Annals of Mathematical Statistics*, *28*(1), 195–241.

Chandel, A. K., Es, C., Rudravaram, R., Narasu, M. L., Rao, V., & Ravindra, P. (2007). *Economics and environmental im*. *2*(February), 14–32.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Statistics*, *24*, 586–602.

Choonut, A., Saejong, M., & Sangkharak, K. (2014). The Production of Ethanol and Hydrogen from Pineapple Peel by Saccharomyces cerevisiae and Enterobacter aerogenes. *Energy Procedia*, *52*, 242–249. https://doi.org/10.1016/j.egypro.2014.07.075

Das, M., & Narasimham, V. (1962). Construction of rotatable desingns through balanced incomplete block desings. *Annals of Mathematical Statistics*, *33*(4).

Dette, H., & Haines, L. M. (1994). E-optimal designs for linear and nonlinear models with two parameters. *Biometrika*, *81*(4), 739–754. https://doi.org/10.1093/biomet/81.4.739

Draper, R. (1985). *Response Transformation : 50*, 523–525.

Ehrenfeld, E. (1955). "On the efficiency of experimental design." *Annals of Mathematical Statistics*, *26*, 247–255.

El-Gendy, N. S., Madian, H. R., & Amr, S. S. A. (2013). Design and optimization of a process for sugarcane molasses fermentation by Saccharomyces cerevisiae using response surface methodology (International Journal of Microbiology). In *International Journal of Microbiology* (Vol. 2013, Issue 0). https://doi.org/10.1155/2013/815631

Ergun, M., & Ferda Mutlu, S. (2000). Application of a statistical technique to the production of ethanol from sugar beet molasses by Saccharomyces cerevisiae. *Bioresource Technology*, *73*(3), 251–255. https://doi.org/10.1016/S0960-8524(99)00140-6

Fakruddin, M., Islam, A., Ahmed, M. M., & Chowdhury, N. (2013). Process optimization of bioethanol production by stress tolerant yeasts isolated from agro-industrial waste. *International Journal of Renewable and Sustainable Energy*, *2*(4), 133–139. https://doi.org/10.11648/j.ijrse.20130204.11

Gardiner, D., Grandage, A., & Hader, R. (1985). Third order rotatable desings for exploring responce surfaces. *Annals of Statistics*, *13*(3), 845–874.

Goettemoeller, J., & Goettemoeller, A. (2007). *Sustainable Ethanol: Biofuels, Biorefineries, Cellulosic Biomass, Flex-fuel Vehicles, and Sustainable Farming for Energy Independence*. Prairie Oak Pub.

Hamouda, H. I., Nassar, H. N., Madian, H. R., Amr, S. S. A., & El-gendy, N. S. (2015). Response Surface Optimization of Bioethanol Production from Sugarcane Molasses by Pichia veronae Strain HSC-22. *Biotechnology Research International*, *2015*, 1–10.

Hansen, A. C., Zhang, Q., & Lyne, P. W. L. (2005). Ethanol-diesel fuel blends - A review. *Bioresource Technology*, *96*(3), 277–285. https://doi.org/10.1016/j.biortech.2004.04.007

Hayder, N. H., Flayeh, H. M., & Ahmed, A. W. (2018). Optimization of Bioethanol Production from Biodegradable Municipal Solid Waste using Response Surface Methodology (RSM). *Journal of Engineering and Sustainable Development*, *22*(10), 47–64.

Hill, W. J., & Hunter, W. G. (1966). A Review of Response Surface Methodology: a literature Survey. *Technometrics*, *8*(4), 571–590.

Irshad, M., & Asgher, M. (2011). *Production and optimization of ligninolytic enzymes by white rot fungus Schizophyllum commune IBL-06 in solid state medium banana stalks*. *10*(79), 18234–18242. https://doi.org/10.5897/AJB11.2242

Isaias, M., Núcleo, M., Planejamento, I. De, & Copersucar, C. C. T. C. (2004). *Assessment of greenhouse gas emissions in the production and use of fuel ethanol in Brazil* (Issue March). Government of the State of São Paulo.

Jones, R. P., Pamment, N., & Greenfield, P. F. (1981). *Alcohol fermentation by yeasts - the effect of environmental and other variables*.

Khuri, A. I. (2017). A General Overview of Response Surface Methodology. *Biometrics & Biostatistics International Journal*, *5*(3), 1–8. https://doi.org/10.15406/bbij.2017.05.00133

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society*, *B21*, 272–319. https://doi.org/10.1007/s00184-011-0348-6

Kiefer, J., & Wolfowitz, J. (1959). Optimum designs in regression problems. *Annals of Mathematical Statistics*, *30*, 271–294.

Klis, F. M., Boorsma, A., & De Groot, P. W. J. (2006). Cell wall construction in Saccharomyces cerevisiae. *Yeast*, *23*(3), 185–202. https://doi.org/10.1002/yea.1349

Koske, J. K., Kosgei, M. K., & Mutiso, J. M. (2011). *a New Third Order Rotatable Design in Five Dimensions*. *13*(1), 157–163.

Krakwowiak, T., Polytechnic, F., & State, B. (1997). *Production of Ethyl Alcohol From Molasses Using Continuous Process*. *8*(I), 35–39.

Liu, Y., Wei, S., & Liao, M. (2013). Optimization of ultrasonic extraction of phenolic compounds from Euryale ferox seed shells using response surface methodology. *Industrial Crops and Products*, *49*, 837–843. https://doi.org/10.1016/j.indcrop.2013.07.023

Mead, R., & Pike, D. (1975). A Review of Response Surface Methodology from a Biometric Viewpoint. *Biometrics*, *31*, 803–851.

Montgomery, D. C. (2005). *Design and Analysis of Experiments* (Fifth). John Wiley $ sons.

Myers, R. H., Khuri, A., & Carter, W. H. (1989). Response surface methodology: 1966-l988. *Technometrics*, *31*(2), 137–157. https://doi.org/10.1080/00401706.1989.10488509

Onuki, S., Leeuwen, J. Van, Jenks, W. S., & Grewell, D. A. (2008). *Ethanol production , purification , and analysis techniques : a review*.

Osunkoya, O., & Okwudinka, N. (2011). Utilization of sugar refinery waste (molasses) for ethanol production- using Saccharomyces Cervicae. *American Journal of Scientific and Industrial Research*, *2*(4), 694–706. https://doi.org/10.5251/ajsir.2011.2.4.694.706

Pazman, A. (1992). A Classification of Non Linear Regression Models. *Kybernetika*, *28*(6), 444–453.

Periyasamy, S., Venkatachalam, S., Ramasamy, S., & Srinivasan, V. (2009). Production of Bio-ethanol from Sugar Molasses Using Saccharomyces Cerevisiae. *Modern Applied Science*, *3*(8). https://doi.org/10.5539/mas.v3n8p32

Prakash Maran, J., Manikandan, S., Vigna Nivetha, C., & Dinesh, R. (2017). Ultrasound assisted extraction of bioactive compounds from Nephelium lappaceum L. fruit peel using central composite face centered response surface design. *Arabian Journal of Chemistry*, *10*, S1145–S1157. https://doi.org/10.1016/j.arabjc.2013.02.007

Pukelsheim, F. (2006). On Linear Regression Designs which Maximize Information. *Statistical Planning and Inference*, *4*, 339–364.

Pukelsheim, Friedrich. (2006). *Optimal Design of Experiments.* Society for industrial and applied mathematics.

Pukelsheim, Friedrich, & Rosenberger, J. L. (1993). Experimental designs for model discrimination. *Journal of the American Statistical Association*, *88*(422), 642–649. https://doi.org/10.1080/01621459.1993.10476317

Rattanapan, A., Limtong, S., & Phisalaphong, M. (2011). Ethanol production by repeated batch and continuous fermentations of blackstrap molasses using immobilized yeast cells on thin-shell silk cocoons. *Applied Energy*, *88*(12), 4400–4404. https://doi.org/10.1016/j.apenergy.2011.05.020

Razmovski, R., & Vucurovic, V. (2011). Enzyme and Microbial Technology Ethanol production from sugar beet molasses by S . cerevisiae entrapped in an alginate – maize stem ground tissue matrix. *Enzyme and Microbioal Techololgy*, *48*, 378–385. https://doi.org/10.1016/j.enzmictec.2010.12.015

Rissanen, J. (1983). Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to The Annals of Statistics. ® www.jstor.org. *The Annals of Statistics*, *11*(2), 416–431.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constant and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, *12*, 1–85.

Sun, Y., & Cheng, J. (2002). <*Sun and Cheng.Pdf*>. *83*, 1–11. https://doi.org/10.1016/S0960-8524(01)00212-7

Tengborg, C., Galbe, M., & Zacchi, G. (2001). Influence of enzyme loading and physical parameters on the enzymatic hydrolysis of steam-pretreated softwood. *Biotechnology Progress*, *17*(1), 110–117. https://doi.org/10.1021/bp000145+

Thatipamala, R., Rohani, S., & Hill, G. A. (1992). Effects of high product and substrate inhibitions on the kinetics and biomass and product yields during ethanol batch fermentation. *Biotechnology and Bioengineering*, *40*(2), 289–297. https://doi.org/10.1002/bit.260400213

Tropea, A., Wilson, D., Torre, L. G. La, Curto, R. B. Lo, Saugman, P., Troy-davies, P., Dugo, G., & Waldron, K. W. (2014). Bioethanol Production From Pineapple Wastes. *Journal of Food Research*, *3*(4), 60–70. https://doi.org/10.5539/jfr.v3n4p60

Tsuyoshi, N., Fudou, R., Yamanaka, S., & Kozaki, M. (2005). Identification of yeast strains isolated from marcha in Sikkim , a microbial starter for amylolytic fermentation. *International Journal of Food Microbiology*, *99*, 135–146. https://doi.org/10.1016/j.ijfoodmicro.2004.08.011

Wald, A. (1943). On the Efficient Design of Statistical Investigations. *The Annals of Mathematical Statistics*, *14*(2), 134–140. http://www.jstor.org/stable/2235815

Wang, Z., Zhuge, J., Fang, H., & Prior, B. A. (2001). Glycerol production by microbial fermentation: A review. *Biotechnology Advances*, *19*(3), 201–223. https://doi.org/10.1016/S0734-9750(01)00060-X

Wyman, C. E., & Hinman, N. D. (1990). Ethanol:Fundamentals of Production from Renewable Feedstocks and Use as a Transportation Fuel. *Applied Biochemistry and Biot Echnology*, *24/25*, 735–736.

Yang, M. (2008). A-optimal designs for generalized linear models with two parameters. *Journal of Statistical Planning and Inference*, *138*(3), 624–641. https://doi.org/10.1016/j.jspi.2006.12.007

Yusuf.o.Raji, Jibril, M., Misau, I. M., & Danjuma, B. Y. (2012). Production of Vinegar From Pineapple Peel. *International Journal of Advanced Scientific Research and Technology*, *3*(2), 656–666. https://doi.org/10.1017/CBO9781107415324.004

Zentou, H., Abidin, Z. Z., Zouanti, M., & Greetham, D. (2017). Effect of operating conditions on molasses fermentation for bioethanol production. *International Journal of Applied Engineering Research*, *12*(15), 5202–5506.

Zhang, J., Ianora, A., Wu, C., Pellegrini, D., Esposito, F., & Buttino, I. (2015). How to increase productivity of the copepod Acartia tonsa (Dana): Effects of population density and food concentration. *Aquaculture Research*, *46*(12), 2982–2990. https://doi.org/10.1111/are.12456

APPENDICES

Appendix I: R-Code for Selected Runs

```
a= 1.137241371

a1= -a

b=2.116644693

b1= -b

r0=c (rep(1,40))

r1=c(rep(c(a1,a),12),rep(0,8),b,b1,rep(0,6))

r2=c(rep(0,8),rep(c(a1,a1,a,a),4),rep(c(a1,a),4),0,0,b,b1,rep(0,4))

r3=c(rep(c(a1,a1,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a),rep(c(a1,a1,a,a),2),rep(0,4),b,b
1,0,0)

r4=c(rep(c(a1,a1,a1,a1,a,a,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a),rep(0,6),b,b1)

Y=cbind(r0,r1,r2,r3,r4)

X=matrix(c(r0,r1,r2,r3,r4,r1*r1,r1*r2,r1*r3,r1*r4,r2*r2,r2*r3,r2*r4,r3*r3,r3*r4,r4*r4
),nrow=40)

MG=round((1/40)*(t(X)%*%X),4)

Dcri=(det(MG))^(1/15)

Acri=(sum(diag(solve(MG)))/15)^-1

Ecri=min(eigen(MG)$value)

Tcri=(sum(diag(MG)))/15

U1=X%*%t(X)#U1 is singular therefore

library(MASS)

U=(ginv(U1))%*%X

Bd=round(U%*%MG%*%t(U),7)# B opt matrix for D-opt Matrix

qd=(Bd[1,1])^.5 #bii for D-opt factorial

qqd=(Bd[40,40])^.5 # bii for D-opt for axial part

dwd=(sum((diag(Bd))^.5))# denominator of Bd for D-optimal weights

Dopwf=qd/dwd #D-opt weight corresponding to factorial part

DopwA=qqd/dwd #D-opt weight for axial part
```

```
Ba=round(U%*%t(U),7)

Bt=round(U%*%MG^2%*%t(U),7)

dr=sum((r1^2)^2)

qt=(Bt[1,1])^.5 #numerator for factorial part T-opt

qqt=(Bt[40,40])^.5 #numerator for axial part T-opt

dwt=(sum((diag(Bt))^.5))#denominator of Bt for T-optimal weights

Topwf=qt/dwt #for factorial part T-optimal weight

TopwA=qqt/dwt #for axial part T-optimal weight

dwa=(sum((diag(Ba))^.5))# denominator Ba for A-optimal weights

qa=(Ba[1,1])^.5 #bii for factorial

qqa=(Ba[40,40])^.5 #bii for axial

Aopwf=qa/dwa #A-optimal weight for factorial

AopwA=qqa/dwa #A-optimal weight for axial part
```

The D-optimal design

```
a1=-1.137241371

a=-a1

b=2.116644693

b1=-b

ra=c(rep(1,32))

r0=c(rep(ra,2),rep(1,24))

r1=c(rep(c(rep(c(a1,a),12),rep(0,8)),2),rep(c(b,b1,rep(0,6)),3))

r2=c(rep(c(rep(0,8),rep(c(a1,a1,a,a),4),rep(c(a1,a),4)),2),rep(c(0,0,b,b1,rep(0,4)),3))

r3=c(rep(c(rep(c(a1,a1,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a),rep(c(a1,a1,a,a),2)),2),rep
(c(rep(0,4),b,b1,0,0),3))

r4=c(rep(c(rep(c(a1,a1,a1,a1,a,a,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a)),2),rep(c(rep(0,
6),b,b1),3))

X=matrix(c(r0,r1,r2,r3,r4,r1*r1,r1*r2,r1*r3,r1*r4,r2*r2,r2*r3,r2*r4,r3*r3,r3*r4,r4*r4
),nrow=88)

MG=round((1/88)*(t(X)%*%X),4)

Doptcri=(det(MG))^(1/15)
```

Doptcri

The A=optimal run

a1=-1.137241371

a=-a1

b=2.116644693

b1=-b

ra=c(rep(1,32))

r0=c(rep(ra,2),rep(1,48))

r1=c(rep(c(rep(c(a1,a),12),rep(0,8)),2),rep(c(b,b1,rep(0,6)),6))

r2=c(rep(c(rep(0,8),rep(c(a1,a1,a,a),4),rep(c(a1,a),4)),2),rep(c(0,0,b,b1,rep(0,4)),6))

r3=c(rep(c(rep(c(a1,a1,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a),rep(c(a1,a1,a,a),2)),2),rep(c(rep(0,4),b,b1,0,0),6))

r4=c(rep(c(rep(c(a1,a1,a1,a1,a,a,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a)),2),rep(c(rep(0,6),b,b1),6))

X=matrix(c(r0,r1,r2,r3,r4,r1*r1,r1*r2,r1*r3,r1*r4,r2*r2,r2*r3,r2*r4,r3*r3,r3*r4,r4*r4),nrow=112)

MG=round((1/112)*(t(X)%*%X),4)

Aoptcri=((1/15)*(sum(diag(solve(MG)))))^-1

Aoptcri

The T-Optimal run

a1=-1.137241371

a=-a1

b=2.116644693

b1=-b

r0=c(rep(1,112))

r1=c(rep(c(rep(c(a1,a),12),rep(0,8)),2),rep(c(b,b1,rep(0,6)),6))

r2=c(rep(c(rep(0,8),rep(c(a1,a1,a,a),4),rep(c(a1,a),4)),2),rep(c(0,0,b,b1,rep(0,4)),6))

r3=c(rep(c(rep(c(a1,a1,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a),rep(c(a1,a1,a,a),2)),2),rep(c(rep(0,4),b,b1,0,0),6))

```
r4=c(rep(c(rep(c(a1,a1,a1,a1,a,a,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a)),2),rep(c(rep(0,
6),b,b1),6))

X=matrix(c(r0,r1,r2,r3,r4,r1*r1,r1*r2,r1*r3,r1*r4,r2*r2,r2*r3,r2*r4,r3*r3,r3*r4,r4*r4
),nrow=112)

MG=round((1/112)*(t(X)%*%X),4)

Toptcri=((1/15)*(sum(diag(MG))))
```

THE E-OPTIMAL run

```
a1=-1.137241371

b=2.116644693

r0=c(rep(1,32))

r1=c(rep(c(a1,a),12),rep(0,8))

r2=c(rep(0,8),rep(c(a1,a1,a,a),4),rep(c(a1,a),4))

r3=c(rep(c(a1,a1,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a),rep(c(a1,a1,a,a),2))

r4=c(rep(c(a1,a1,a1,a1,a,a,a,a),2),rep(0,8),c(a1,a1,a1,a1,a,a,a,a))

Y=cbind(r0,r1,r2,r3,r4)

X=matrix(c(r0,r1,r2,r3,r4,r1*r1,r1*r2,r1*r3,r1*r4,r2*r2,r2*r3,r2*r4,r3*r3,r3*r4,r4*r4
),nrow=32)

MG=round((1/32)*(t(X)%*%X),4)

Z=eigen(MG)$vectors

z=Z[1:15,15]

Z1=round(eigen(MG)$values,5)

#Vo=((1/40)*(sum((diag(B))^.5)^2))^-1

d=read.csv("D:\\timf.csv",header=T)

library(rsm)

ft=rsm(formula=Y~SO(X1,X2,X3,X4),data=d)

summary(ft)

sa1=steepest(ft)

library(MASS)

d.lm=lm(Y~X1+X2+X3+X4,data=d)

par(mfrow=c(3,2))
```

```
plot(d1,pch=20,xlab="(a) Index",ylab="residuals",main="Scatter Plot Of The Errors")

d.stdres=rstandard(d.lm)

qqnorm(d.stdres,

ylab="standardized residuals",

xlab="(b) Normal scores",

main="Normal Q-Q Plot Ethanol Yield")

qqline(d.stdres)

plot(d.stdres~Run,ylab="standardised residuals",xlab="(c)
RunNumber",pch=20,data=d)

plot(d.stdres~Yh,ylab="standardised residuals",xlab="(d) Predicted",pch=20,data=d)

plot(d.stdres~Y,ylab="standardised residuals",xlab="(e) Actual
Yield",pch=20,data=d)

plot(Yh~Y,ylab="Predicted",xlab="(f) Actual Yield",pch=20,data=d)

d=read.csv("D:\\timf.csv",header=T)# design matrix and experimental values

library(rsm)

ft=rsm(formula=Y~SO(X1,X2,X3,X4),data=d)#fit full second order model

summary(ft)#anova output

sa1=steepest(ft)# path of steepest acsent

d=read.csv("D:\\timf.csv",header=T)

d1=d[1:32,]# E-optimal design  matrix

d2=c(33,0,0,0,0,11.821)# argumenting with one centre point

d3=rbind(d1,d2)

library(rsm)

ft=rsm(formula=Y~SO(X1,X2,X3,X4),data=d3)

summary(ft)

sa1=steepest(ft)# pathof steepest ascent

ept1=dupe(sa1[2:9,])

we.lm=rsm(Y~SO(X1,X2,X3,X4),data=d)

par(mfrow=c(2,3))
```

```
image(we.lm,~X1+X2+X3+X4)
xs=canonical(we.lm)$xs# stationary point of full model
myhook=list()
myhook$post.plot=function(lab){idx=sapply(lab[3:4],grep,names(xs))
points(xs[idx[1]],xs[idx[2]],pch=2,col="red")}
par(mfrow=c(2,3))
contour(we.lm,~X1+X2+X3+X4,image=TRUE,# Contour plots
at=xs,hook=myhook)
```

Table A: The Design Matrix *X* for the Full Second Order Model

X_0	X_1	X_2	X_3	X_4	$X_1 X_2$	$X_1 X_3$	$X_1 X_4$	$X_2 X_3$	$X_2 X_4$	$X_3 X_4$	X_1^2	X_2^2	X_3^2	X_4^2
1	-1.137	0	-1.137	-1.137	0	1.2928	1.2928	0	0	1.2928	1.2928	0	1.2928	1.2928
1	1.137	0	-1.137	-1.137	0	-1.2928	-1.2928	0	0	1.2928	1.2928	0	1.2928	1.2928
1	-1.137	0	1.137	-1.137	0	-1.2928	1.2928	0	0	-1.2928	1.2928	0	1.2928	1.2928
1	1.137	0	1.137	-1.137	0	1.2928	-1.2928	0	0	-1.2928	1.2928	0	1.2928	1.2928
1	-1.137	0	-1.137	1.137	0	1.2928	-1.2928	0	0	-1.2928	1.2928	0	1.2928	1.2928
1	1.137	0	-1.137	1.137	0	-1.2928	1.2928	0	0	-1.2928	1.2928	0	1.2928	1.2928
1	-1.137	0	1.137	1.137	0	-1.2928	-1.2928	0	0	1.2928	1.2928	0	1.2928	1.2928
1	1.137	0	1.137	1.137	0	1.2928	1.2928	0	0	1.2928	1.2928	0	1.2928	1.2928
1	-1.137	-1.137	0	-1.137	1.2928	0	1.2928	0	1.2928	0	1.2928	1.2928	0	1.2928
1	1.137	-1.137	0	-1.137	-1.2928	0	-1.2928	0	1.2928	0	1.2928	1.2928	0	1.2928
1	-1.137	1.137	0	-1.137	-1.2928	0	1.2928	0	-1.2928	0	1.2928	1.2928	0	1.2928
1	1.137	1.137	0	-1.137	1.2928	0	-1.2928	0	-1.2928	0	1.2928	1.2928	0	1.2928
1	-1.137	-1.137	0	1.137	1.2928	0	-1.2928	0	-1.2928	0	1.2928	1.2928	0	1.2928
1	1.137	-1.137	0	1.137	-1.2928	0	1.2928	0	-1.2928	0	1.2928	1.2928	0	1.2928
1	-1.137	1.137	0	1.137	-1.2928	0	-1.2928	0	1.2928	0	1.2928	1.2928	0	1.2928
1	1.137	1.137	0	1.137	1.2928	0	1.2928	0	1.2928	0	1.2928	1.2928	0	1.2928
1	-1.137	-1.137	-1.137	0	1.2928	1.2928	0	1.2928	0	0	1.2928	1.2928	1.2928	0
1	1.137	-1.137	-1.137	0	-1.2928	-1.2928	0	1.2928	0	0	1.2928	1.2928	1.2928	0
1	-1.137	1.137	-1.137	0	-1.2928	1.2928	0	-1.2928	0	0	1.2928	1.2928	1.2928	0
1	1.137	1.137	-1.137	0	1.2928	-1.2928	0	-1.2928	0	0	1.2928	1.2928	1.2928	0
1	-1.137	-1.137	1.137	0	1.2928	-1.2928	0	-1.2928	0	0	1.2928	1.2928	1.2928	0
1	1.137	-1.137	1.137	0	-1.2928	1.2928	0	-1.2928	0	0	1.2928	1.2928	1.2928	0
1	-1.137	1.137	1.137	0	-1.2928	-1.2928	0	1.2928	0	0	1.2928	1.2928	1.2928	0
1	1.137	1.137	1.137	0	1.2928	1.2928	0	1.2928	0	0	1.2928	1.2928	1.2928	0
1	0	-1.137	-1.137	-1.137	0	0	0	1.2928	1.2928	1.2928	0	1.2928	1.2928	1.2928
1	0	1.137	-1.137	-1.137	0	0	0	-1.2928	-1.2928	1.2928	0	1.2928	1.2928	1.2928
1	0	-1.137	1.137	-1.137	0	0	0	-1.2928	1.2928	-1.2928	0	1.2928	1.2928	1.2928
1	0	1.137	1.137	-1.137	0	0	0	1.2928	-1.2928	-1.2928	0	1.2928	1.2928	1.2928
1	0	-1.137	-1.137	1.137	0	0	0	1.2928	-1.2928	-1.2928	0	1.2928	1.2928	1.2928
1	0	1.137	-1.137	1.137	0	0	0	-1.2928	1.2928	-1.2928	0	1.2928	1.2928	1.2928
1	0	-1.137	1.137	1.137	0	0	0	-1.2928	-1.2928	1.2928	0	1.2928	1.2928	1.2928
1	0	1.137	1.137	1.137	0	0	0	1.2928	1.2928	1.2928	0	1.2928	1.2928	1.2928
1	2.116	0	0	0	0	0	0	0	0	0	4.4775	0	0	0
1	-2.116	0	0	0	0	0	0	0	0	0	4.4775	0	0	0
1	0	2.116	0	0	0	0	0	0	0	0	0	4.4775	0	0
1	0	-2.116	0	0	0	0	0	0	0	0	0	4.4775	0	0
1	0	0	2.116	0	0	0	0	0	0	0	0	0	4.4775	0
1	0	0	-2.116	0	0	0	0	0	0	0	0	0	4.4775	0
1	0	0	0	2.116	0	0	0	0	0	0	0	0	0	4.4775
1	0	0	0	-2.116	0	0	0	0	0	0	0	0	0	4.4775

TABLE B: THE MATRIX $X'X$

40.00	0.00	0.00	0.00	0.00	39.99	0.00	0.00	0.00	39.99	0.00	0.00	39.99	0,00	39.99
0.00	39.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	39.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	39.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	39.99	0..00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
39.99	0.00	0.00	0.00	0.00	80.21	0.00	0.00	0.00	26.75	0.00	0.00	26.75	0.00	26.75
0.00	0.00	0.00	0.00	0.00	0.00	26.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	26.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	26.75	0.00	0.00	0.00	0.00	0.00	0.00
39.99	0.00	0.00	0.00	0.00	26.75	0.00	0.00	0.00	80.21	0.00	0.00	26.75	0.00	26.75
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	26.75	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	26.75	0.00	0.00	0.00
39.99	0.00	0.00	0.00	0.00	26.75	0.00	0.00	0.00	26.75	0.00	0.00	80.21	0.00	26.75
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	26.75	0.00
39.99	0.00	0.00	0.00	0.00	26.75	0.00	0.00	0.00	26.75	0.00	0.00	26.75	0.00	80.21

Table C: The Matrix $X'X$ Inverse

7.001	0.000	0.000	0.000	0.000	−1.745	0.000	0.000	0.000	−1.75	0.000	0.000	−1.745	0,000	−1.745
0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.025	0..000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
−1.745	0.000	0.000	0.000	0.000	0.450	0.000	0.000	0.000	0.432	0..000	0.000	0.432	0.000	0.432
0.000	0.000	0.000	0.000	0.000	0.000	0.432	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.432	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.432	0.000	0.000	0.000	0.000	0.000	0.000
−1.745	0.000	0.000	0.000	0.000	0.432	0.000	0.000	0.000	0.450	0.000	0.000	0.432	0.000	0.432
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0,432	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.432	0.000	0.000	0.000
−1.745	0.000	0.000	0.000	0.000	0.432	0.000	0.000	0.000	0.432	0.000	0.000	0.450	0.000	0.432
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.432	0.000
−1.745	0.000	0.000	0.000	0.000	0.432	0.000	0.000	0.000	0.432	0.000	0.000	0.432	0.000	0.450

Table D: Monosaccharide Compositions of Pineapple Waste Cell Walls.(Tropea et al., 2014)

		Residue total	Rhamnose		Fucose		Arabinose		Xylose		Mannose		Galactose		Glucose		Galacturonic acid	
Test	Hours	(%)	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
	0	3.9	657.1	30.6	3.0	0.3	2.5	0.2	57.0	2.1	224.8	9.8	22.3	1.1	34.8	4.2	251.7	25.1	60.9	3.7
DF	12	3.7	642.5	29.8	2.8	0.2	1.5	0.2	59.8	0.2	214.7	14.9	24.1	1.7	35.8	1.6	239.3	16.6	64.5	1.5
	48	1.7	528.6	17.8	2.7	0.3	0.9	0.1	31.4	4.2	242.3	15.8	21.7	3.0	18.2	2.7	173.2	10.6	38.1	4.9
	0	3.2	660.1	53.5	1.5	0.1	1.1	0.1	62.6	2.7	246.4	16.0	16.5	2.0	36.0	3.1	235.4	20.2	60.4	8.9
SHF	12	1.4	581.1	8.9	0.8	0.1	0.4	0.0	61.9	2.2	244.9	2.6	6.6	0.4	38.7	1.7	116.6	8.1	111.2	10.2
	30	0.7	565.6	4.4	1.5	0.1	1.0	0.1	54.6	3.3	236.0	15.9	14.5	0.5	36.7	2.7	123.0	3.4	98.4	4.7
	0	3.4	640.4	29.8	1.6	0.2	1.5	0.1	59.8	0.2	198.5	14.9	24.1	1.7	38.4	1.6	239.3	16.5	77.3	1.5
SSF	12	1.2	432.4	37.3	1.2	0.2	0.6	0.1	35.0	4.2	174.9	33.6	28.7	1.9	21.5	2.5	133.7	13.3	36.7	5.8
	30	0.8	375.4	12.3	1.2	0.1	0.7	0.1	32.4	2.3	97.3	6.3	29.7	2.5	22.6	1.9	85.2	1.6	106.3	12.5

Fiber Compositions; at Time 0, After 12 Hours and at The end of the process, for Direct Fermentation (DF), Separate Hydrolysis and Fermentation (SHF) and Simultaneous Saccharification and Fermentation (SSF) expressed as µg/mg anhydrous sugars in original sample. Results are shown as mean value (m) and standard deviation (SD); residue (%) = proportion of biomass recovered as alcohol insoluble residue (air).

Table E:Dry Matter (% Fwt), Fiber and Soluble Sugars (% Dry Matter).(Tropea et al., 2014)

	% of dry matter		% fibre in dry matter		% soluble sugars		Ethanol % (v/v) Amount		pH	
	initial	final	initial	final	initial	final	EY	TY	initial	final
DF	9.2	3.1	27.8	5.4	57.8	4.7	34 ± 0.2	86	4.5	3.4
SHF	8.5	2.6	25	5.3	48.6	6.2	3.7 ± 0.1	89	5	3.3
SSF	9	2.7	23.9	3.4	42.2	7.5	3.9 ± 0.1	96	4.5	3.3

Ethanol Yield (EY), Theoretical Yield (TY) and PH for Direct Fermentation (DF), Separate Hydrolysis and Fermentation (SHF) and Simultaneous Saccharification and Fermentation (SSF).

Figure A: Fresh Pineapple Peels Left: Right Sun Dried Peels after three Days At 25^0C

Figure B: Left; Milled Peels after Oven Drying: Right Dissolved and Pre-Treated Peels.

Figure C: Ccontours Plot Figure 4.2f Expanded