

**MODELING THE YIELD OF SOYBEAN (*Glycine max* (L.) Merrill) USING A
FOUR COMPONENT MIXTURE EXPERIMENT IN THE PRESENCE OF
VARIATION CREATED BY TWO PROCESS VARIABLES WITHIN SPLIT-
PLOT DESIGN**

BY

WANYONYI WANGILA SAMSON

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
MATHEMATICS (BIOSTATISTICS) IN THE SCHOOL OF SCIENCE
UNIVERSITY OF ELDORET, KENYA**

JUNE, 2021

DECLARATION

Declaration by the Candidate

This thesis is my original work and has not been submitted for any academic awards in any institution; and shall not be reproduced in part or full, or in any format without prior written permission from the author and / or University of Eldoret

Wanyonyi Wangila Samson

Signature:  Date: 09/06/2021

SSCI/MAT/P/003/18

Declaration by Supervisors

This proposal has been submitted with our approval as university supervisors.

Signature:  Date: 09/06/2021

Dr. Ayubu Anapapa

Department of Mathematics and Actuarial Science

Murang'a University of Technology, Kenya

Signature:  Date: 09/06/2021

Dr. Julius Koech

Department of Mathematics and Computer Science

University of Eldoret, Kenya

DEDICATION

I dedicate my dissertation work to my family, church members, friends and thesis advisors who supported me throughout this PhD academic journey. In addition, I will always appreciate all they have done especially Nancy Ashibambo my wife for her endless support to help me pay school fees on time.

ABSTRACT

Use of mixture design has become well-known in statistical modeling due to its utility in predicting any mixture's response and thus serving as the foundation for optimizing the predicted response on blends of different components. In most split-plot designs utilizing mixture-process variable settings, restricted randomization always exist. This study's primary goal was to find the best split-plot design (SPD) for performing the *Glycine max* experiment with the settings mixture-process variables. The SPD was made up with use of a Simplex Centroid Design (SCD) of four mixture blends and a 2^2 factorial design with a Central Composite Design (CCD) of the process variable and this was compared with six different designs of split-plot structure arrangement. The JMP software version 15 was used to create D-optimal split-plot designs. The study compared the constructed designs' relative efficiency using A-, D-, I-, and G- optimality criteria respectively. Furthermore, use of graphical technique (fraction of design space plot) was used to display, explain, and evaluate experimental designs' performance in terms of precision of the six designs' variance prediction properties. Results showed that arranging subplots with more SCD points as compared to the pure mixture design points within SPD with two high process variables provided more precise parameter estimates. Also use of the restricted maximum likelihood method was employed to estimate parameter values within the SPD. The current study thoroughly investigated estimation of parameters from MPV settings in conjunction with CCD within SPD using the Scheffe polynomial and Cox mixture models to predict optimal responses of *Glycine max* yield. The predicted maximum optimum yield for the total number of seeds per plant stem of *Glycine max* was 102.06 equivalent to 15.7832g using the screening methodology. Furthermore, the expected response from the simulated-based technique for {4, 2} SCD and the actual results obtained from field experiments using MPV settings within SPD were compared. The variety of *Glycine max* Blyvoor was found to have higher production in terms of yield as compared to R 184. Thus the study recommends farmers to use Blyvoor as an alternative variety. In addition, the study also recommends farmers to embrace SPDs in the context of mixture settings formulations in order to measure the interaction effects of both the mixture components and the processing conditions like soil pH and the seeding rate.

TABLE OF CONTENTS

DECLARATION	ii
DEDICATION	iii
ABSTRACT	iv
LIST OF TABLES	ix
LIST OF FIGURES	xi
LIST OF PLATES	xiii
LIST OF ABBREVIATION	xiv
ACKNOWLEDGEMENT	xvi
CHAPTER ONE	1
INTRODUCTION.....	1
1.1 Background of the Study	1
1.2 Theory behind MPV within Split-plot Designs	3
1.3 Statement of the Problem.....	5
1.4 Objective.....	7
1.4.1 General Objective	7
1.4.2 Specific Objectives	7
1.5 Rationale of the Study.....	7
1.6 Significance of the Study	9
1.7 Scope of the Study	10
1.8 Assumption of the Study.....	10
1.9 Thesis Outline	10
CHAPTER TWO	12
LITERATURE REVIEW	12
2.1 Introduction.....	12
2.2 Mixture Design and Statistical Mixture Models	14
2.2.1 Simplex Lattice Design (SLD).....	16
2.2.2 Simplex Centroid Design (SCD).....	20
2.2.3 Mixture Design with Presence of Process Variable	23
2.2.4 The Combined MPV Model.....	24
2.3 Split-Plot Layout Structure	33
2.3.1 Hasse Diagrams and Expected Means Squares within SPD	37
2.3.2 Analysis of Split-Plot Designs	39

2.4	Robust Parameter Design for MPV with Noise Variable	40
2.5	The Design Experimental Region (Design Space).....	41
2.6	Graphical Techniques for Evaluating Design Optimality	41
2.6.1	Fraction of Design Space Plots Construction	43
2.6.2	Algorithm Used to Generate FDS Plots.....	44
2.6.3	Variance Dispersion Graphs (VDGs)	45
2.6.4	Evaluation of FDS plot for MPV designs	47
2.6.5	The Average Relative Variance	48
2.7	General Optimal Design Criteria for Comparing and Examining Designs.....	51
2.7.1	D-optimality Criterion	53
2.7.2	G-optimality Criterion	55
2.7.3	A-optimality Design Criterion	56
2.7.4	I -optimal Criterion	57
2.8	D-Optimal Designs for Construction of Mixture Designs	60
2.9	I-Optimal Criteria for Construction of Mixture Designs	61
2.10	Comparison of D- And I-optimal Designs.....	63
2.11	Central Composite Design	64
2.12	The Ecology and Significant Production of Soybean Globally	69
2.12.1	Effect of organic and inorganic fertilizer on <i>Glycine max</i> Crop.....	71
2.12.2	Effect of Row Spacing and Seeding Rate on <i>Glycine max</i> Variety of Seeds.....	74
2.12.3	Effect of Soil pH on <i>Glycine max</i> (L.) Merrill Yield.....	76
2.13	Screening Methodology in Mixture design with application to <i>Glycine max</i>	77
CHAPTER THREE.....		86
RESEARCH METHODOLOGY		86
3.1	Introduction.....	86
3.2	Developing a Parsimonious Model for Use of Split-Plot Design and MPV	86
3.2.1	New Design for Split-Plot Layout to Support Fitting MPV	87
3.2.3	Construction of SPD for Combined MPV with CCD Formulated.....	97
3.2.4	Evaluation of MPV Design with Split Plot Structure	111
3.2.4.1	Prediction Variance for MPVD with a Split Plot Structure	111
3.2.4.2	Fraction Design Space Plots for MPV Design within SPD	113
3.2.4.3	Sliced Fraction Design Space Plots for MPVD within SPD.....	113
3.2.6	Evaluation of a Desirable Design for MPV within SPD.....	114

3.3	Modified MPV Model in Predicting the Yield of <i>Glycine Max</i>	117
3.3.1	Data Source.....	117
3.3.2	Description of Experimental Sites	118
3.3.3	Composting Farmyard Manure through the Framework of Mixture Design.....	120
3.3.4	Treatment combinations Manure through of MPV within SPD Context.....	124
3.3.5	Soil Sample Analysis	126
3.3.6	Exploration and Estimation of Parameters for MPV within Split Plot Layout.....	128
3.3.6.1	Maximum Likelihood and Ordinary Least Square Method	128
3.3.6.2	Estimation of θ and δ Using REML Method	131
3.3.6.3	Estimation Whole Plot Random Effect Using Bayesian Perspective Approach.....	134
3.3.6.4	Estimation of Parameters Using EMA Method	135
3.3.7	SPD Analysis Using a Full REML Method.....	139
3.4	Screening Methodology in the Framework of a Cox MPV Model.....	139
3.5	Estimating the Optimal Yield of <i>Glycine Max</i> in the Framework of SPD.....	148
3.5.1	Plants per Acre.....	150
3.5.2	Pods per Plant	150
3.5.3	Seeds per Pod.....	150
3.5.4	Seeds per Pound (Seed Size).....	150
	CHAPTER FOUR.....	152
	RESULTS AND DISCUSSION	152
4.1	Introduction.....	152
4.2	The Modified MPV Model in Predicting the Yield of <i>Glycine Max</i>	159
4.2.1	Data Source.....	159
4.2.2	Exploration and Estimation of Parameters for MPV within Split Plot Layout.....	166
4.2.3	ANOVA Tables Based on Analyzing of Mixture Process Data	169
4.3	The Screening Methodology for Estimating the Predicted Yield.	187
4.3.1	Formulated of Cox MPV Model	188
4.3.2	Screening in a MPV Settings	209
4.3.3	The Joint Factor Tests of MPV Using the Cox Reference Mixture Model.....	211
4.4	Estimated Optimal Yield of <i>Glycine max</i> in the Framework of SPD	212
4.4.1	Response optimization in SPD in the presence of MPV.....	212
4.4.2	Optimal Yield of <i>Glycine max</i> within SPD.....	217
	CHAPTER FIVE	219
	CONCLUSIONS AND RECOMMENDATIONS.....	219

5.1	Conclusions.....	219
5.2	Recommendation	221
	REFERENCE.....	223
	APPENDICES.....	235
	APPENDIX I: R Code	235
	APPENDIX II: Research Licenses	237
	APPENDIX III: Similarity Report.....	238

LIST OF TABLES

Table 2.1: The EMS for SPD.....	36
Table 2.2: I-Optimal weights for second order mixture model.....	62
Table 3.1: Proposed design A1.....	98
Table 3.2: Proposed design A2.....	100
Table 3.3: Proposed design A3.....	102
Table 3.4: Proposed design A4.....	105
Table 3.5: Proposed design A5.....	107
Table 3.6: Proposed design A6.....	109
Table 3.7: Scaled seeding rate and soil pH according to 2 ² factorial design.....	125
Table 3.8: Soil analysis from the two sites.....	128
Table 4.1: Optimality criterion efficiency of design A4 Relative to design 1, 2, 3, 5 and A6.....	153
Table 4.2: Design optimality efficiency of design A6 Relative to design 1, 2, 3, 4 and A5.....	154
Table 4.3: Optimality criterion efficiency of design A6 Relative to design 1, 2, 3, 4 and A5.....	155
Table 4.4: Optimality criterion efficiency of design A3 Relative to design 1, 2, 4, 5 and A6.....	156
Table 4.5: Optimality criterion efficiency of design A2 Relative to design 1, 3, 4, 5 and A6.....	157
Table 4.6: Optimality criterion efficiency of design A1 Relative to design 2, 3, 4, 5 and A6.....	158
Table 4.7: D-optimality criterion efficiency of design design A1, A2, A3, A4 , A5and A6.....	158
Table 4.8: The simulated response of <i>Glycine max</i>	160
Table 4.9: The experimental results of Glycine max for R 184.....	162
Table 4.10: The experimental results of Glycine max for Blyvoor.....	164
Table 4.11: The REML variance component estimates obtained from the model.....	166
Table 4.12: The summary fit of the eight responses obtained using MPV setting.....	167
Table 4.13: The t student test for the fitted MPV model for the No. of branches.....	169
Table 4.14: The t student test for the fitted MPV model for the No.of Pods on branches.....	170
Table 4.15: The t student test for the fitted MPV model for the No. of Pods on Main stem.....	173
Table 4.16: The least square means for the No. of pods on main stem.....	176
Table 4.17: The t student test for the fitted MPV model for the No. of Pods per branch.....	177
Table 4.18: The t student test for the fitted MPV model for the total No. of Pods per plant.....	180
Table 4.19: The t student test for the fitted MPV model for the No. of seeds per pod.....	181
Table 4.20: The t student test for the fitted MPV model for the total No. of seeds per plant.....	184
Table 4.21: The t student test for the fitted MPV model for yield of seeds in grams per plant.....	185
Table 4.22: The t student test for the fitted Cox MPV model for the No. of branches.....	188

Table 4.23: The t student test for the fitted Cox MPV model for the No. of Pods on branches.....	190
Table 4.24: The t student test for the fitted Cox MPV model for the No. of Pods on Main stem....	192
Table 4.25: The t student test for the fitted Cox MPV model for the No. of Pods per branch.....	194
Table 4.26: The t student test for the fitted Cox MPV model for the total No. of Pods.....	196
Table 4.27: The fitted Normal distribution for the No. of pods on branches.....	200
Table 4.28: The t student test for the fitted Cox MPV model for the No. of seeds per pod.....	204
Table 4.29: The t student test for the fitted Cox MPV model for the total No. of seeds.....	206
Table 4.30: The t student test for the fitted Cox MPV model for yield of seeds in grams.....	208
Table 4.31: ANOVA for the joint factor tests of MPV.....	211
Table 4.32: The MPV setting that can lead to optimum yield of Glycine max.....	215
Table 4.33: Estimation of yield of Blyvoor and R 184 in Bushel per acre.....	217

LIST OF FIGURES

Figure 2.1: SLD with 3 mixture components.....	17
Figure 2.2: The comparison of factor space within simplex lattice design.....	18
Figure 2.3: Mixtures of mixtures within SLD for three mixture blends.....	19
Figure 2.4: A four-component mixture of SLD.....	20
Figure 2.5: A second order model, a SCD of three mixture components.....	21
Figure 2.6: The comparison between SCD and SLD for q mixture blends.....	22
Figure 2.7: The expected response nomenclature for four component mixture in SLD.....	22
Figure 2.8: Proposed mixture designs for the 3-2 Case.....	31
Figure 2.9: Illustration of L- Pseudo components.....	33
Figure 2.10: Two Hasse Diagrams for SPD.....	38
Figure 2.11: The total variation resulting from both and within Whole plot and sub plot	40
Figure 2.12: The contours plots for objective function space for D-, G- and I-optimality.	58
Figure 2.13: Comparison of the SPV created by the two optimal designs.....	62
Figure 2.14: Diagram of CCD generation for two factor components.....	66
Figure 3.1: A 2^2 factorial design with CCD of process variable.....	90
Figure 3.2: A Proposed Design for Split-Plot layout to fit MPV model with CCD.....	90
Figure 3.3: A proposed design for split-plot layout for combined 2nd order MPV with CCD.....	91
Figure 3.4: The map of Lugari Sub County in Kakamega County for field experiment.....	119
Figure 3.5: Only Sheep and Chicken manure considered for site 1 and 2, respectively.....	122
Figure 3.6: Only Cow manure and Goat manure considered for site 3 and 4, respectively.....	122
Figure 3.7: Both cow and Goat, Sheep and Goat manure were considered.....	123
Figure 3.8: Both chicken and Goat manure considered in this case.....	123
Figure 3.9: Both chicken and Cow, Sheep and Cow manure were considered.....	124
Figure 3.10: All the four-manure component considered	124
Figure 3.11: The soil pH chart indicating the ideal PH range for plant growth.....	126
Figure 4.1: Sliced FDS plot of design A4 Relative to design A1, A2, A3, A5 and A6.....	152
Figure 4.2: Sliced FDS plot of design A6 Relative to design A1, A2, A3, A4 and A5.....	153
Figure 4.3: Sliced FDS plot of design A5 Relative to design A1, A2, A3, A4 and A6.....	154
Figure 4.4: Sliced FDS plot of design A3 Relative to design A1, A2, A4, A5 and A6.....	155
Figure 4.5: Sliced FDS plot of design A2 Relative to design A1, A3, A4, A5 and A6.....	156
Figure 4.6: Sliced FDS plot of design A1 Relative to design A2, A3, A4, A5 and A6.....	157

Figure 4.7: The QQ plot for the number of pods on branches verses the fitted quantile.....	172
Figure 4.8: The effect summary of MPV settings on Pod development on main stem.....	175
Figure 4.9: The LSM plot for the response Y4 on main stem within the whole plots.....	176
Figure 4.10: The graph of Studentized Residuals for No. of Pods per branch.....	179
Figure 4.11: The No. of seeds per plant observed in each subplot within SPD.....	183
Figure 4.12: The QQ plot and PP plot for the yield of seeds in grams per plants.....	187
Figure 4.13: Histogram, Box plot and Normal Quantile plot fitted for pods on branches.....	198
Figure 4.14: The distribution profiler for the normal distribution for pods on branches.....	200
Figure 4.15: The response prediction for the No. of pods on a branches.....	201
Figure 4.16: The interaction plot resulting from MPV setting in the framework of SPD.....	202
Figure 4.17: The overlaid survival and failure plots for each group.....	203
Figure 4.18: The response prediction profiler plus interaction effect.....	214
Figure 4.19: The response contours of mixture experiment models on a ternary plot.....	214
Figure 4.20: The MPV setting with optimum predicted response of Glycine max.	216

LIST OF PLATES

Plate 1.1: Typical production of Glycine max on scale farm.....	9
Plate 3.1: Soybean plants at reproductive stage.....	149

LIST OF ABBREVIATION

AIC:	Akaike Information Criterion
ANOVA:	Analysis of Variance
BIC	Bayesian Information Criterion
CAN:	Calcium Ammonium Nitrate
CCD:	Central Composite Design
CIAT:	International Center for Tropical Agriculture
DAP:	Diammonium Phosphate
ECF	Easy to Change Factor
EMS:	Expected Means Square
FAO:	Food and Agriculture Organization
FDS:	Fraction Dispersion Space
FSD	Factorial Screen Design
FURP:	Fertilizer Use Recommendation Project
FYM:	Farmyard Manure
HCF	Hard to change Factor
IITA:	International Institute of Tropical Agriculture
INM:	Integrated Nutrient Management
ML:	Machine Learning
MOA:	Ministry of Agriculture
MPV:	Mixture Process Variable

MPVD:	Mixture Process Variable Design
OLS	Ordinary Least Square
PV:	Process variable
REML:	Restricted Maximum Likelihood
RVP:	Relative Variance Prediction
SCD:	Simplex Centroid Design
SDA	Seventh Day Adventist
SLD:	Simplex Lattice Design
SPD:	Split Plot Design
SPE:	Split Plot Error
SPV:	Scaled Prediction Variance
VDG:	Variance Dispersion Graph
WP:	Whole Plot
WPE:	Whole plot Error

ACKNOWLEDGEMENT

I'm very grateful to my supervisors Dr. Ayubu A. Okango, Dr. Julius K. Koech and Dr. Betty C. Korir for their valuable piece of advice, support and guidance throughout the writing this thesis. Their consistent dedication and encouragement ensured success for this thesis. In addition, I wish to extend my deepest acknowledgement to Head of department of Mathematics and Computer science Dr. Argwings Otieno for ensuring successful completion of this thesis within the stipulated time. I also take this opportunity to thank Ms. Sevu Mueni from Kenya Agriculture and Livestock Organization, Food crop Research institute KALRO Alupe, Busia, Kenya for her great effort contribution in soil sample analysis and interpretation. Last but not least, my wife, children, parents, brothers, sisters and friends for standing with me during my down moments. Further, I give my deepest gratefulness to almighty God for giving good health, his grace in each and everything I have done during this entire period of study and research.

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

A mixture design has become popular in statistical modeling in a Mixture Process Variable (MPV) experiment owing to its usefulness in modeling the blending surface that predicts the response of any mixture empirical (Anderson Cook *et al.*, 2004; Cho, 2010; Njoroge *et al.*, 2017; Sitinjak and Syafitri, 2019). In MPV, the response is a function of the mixture part proportion and the process variable. Thus the explanatory variable and the response variable in a mixture experiment are dependent only on the relative proportion of the mixture ingredient, not on the mixture's volume. (Cornell, 1988; Goldfarb *et al.*, 2004b). However, process variables are variables that do not make up a portion of the mixture in an experiment but influence the ingredients' blending properties when their levels are modified. (Goos and Donev, 2007).

In the agricultural sector, crop yield is determined by the relative proportions of mixture components and one or more of hidden factors such as soil pH, planting time, irrigation, and seed variety planted, all of which affect the blending properties of the mixture ingredients. The quality of the optimal yield obtained is determined by the relative proportions of the components used in blending, such as applying different proportions of mixture components of fertilizer, such as Diammonium Phosphate((18% N-48% P_2O_5), Single supers (18-20% P_2O_5), and triple supers (40-42 % P_2O_5) depending on the soil pH.

As Goldfarb *et al.* (2003) and Myers *et al.* (2009) discussed, these difficult-to-change factors are referred to as noise variables. As a result, the number of runs in MPV design

tends to increase as the number of hidden variables grows (Cho *et al.*, 2009). Therefore, due to functional and financial constraints, noise variables often prevent complete randomization of experimental runs. This situation can be solved when the split-plot design is applied subject to the restricted constraints (Goldfarb *et al.*, 2004b; Cho, 2010). The MPV should be carried out in a split-plot structure to counter the effect of noise variables that inhibit complete randomization in experimental trials (Goos and Donev, 2007). Therefore, complete randomization is always required for statistical purposes (Kowalski *et al.*, 2002; Sitinjak and Syafitri, 2019). The process variable determines the whole plot (primary treatment) in a split-plot design. The subplot treatments, on the other hand, were identified by the group of mixture ingredients. This method of setting the noise variable at a fixed level and then running all combinations of the other category at that level usually results in correlated observations (Cho, 2010). Therefore, the first main objective is to extend MPV by introducing a simplex centroid design using the four mixture components in the presence of two process variables. A simplex centroid of four mixture components and a 2^2 factorial design with Central Composite Design (CCD) of process variables were designed to make up a Split-Plot Design (SPD). The parameters of the two sources of errors in the split-plot design were estimated using the method of Restricted Maximum Likelihood (REML) as described by Kowalski *et al.* (2002) and Wanyonyi *et al.* (2018).

The second goal was to find the best SPD by using the chosen optimality criteria to test MPV by computing the design efficiencies of the relative designs as defined by Goos and Vanderbroek (2003), Vining *et al.* (2009), Goos and Donev (2007), and Wong (1994).

1.2 Theory behind MPV within Split-plot Designs

Cornell (2011) presents detailed information about mixture and MPV experiments. When the process variable is hard to change (Noise variable), Goldfarb *et al.* (2003), Goos and Donev (2007), and Cho (2010) examined and evaluated robust MPV designs taking into account the usual process means and the Variance. The research begins at this MPV design entailing a hard-to-change process variable (Soil pH) and adapts the MPV with split-plot designs structure.

However, the first Key articles focus on MPV design with hard change factor, commonly known as noise variable. In these papers, the researchers aimed to develop a model comprising mixture blends and controllable and uncontrollable process variables. The controllable process variables are easy to change factors, whereas uncontrollable factors are hard to change. They also took into account the models, which allow correlations between hard-to-change factors. They used a study technique involving a robust process in establishing variable levels, responsive to alterations in the uncontrollable process variable. In a situation involving rigorous analysis, the Delta method was used to evaluate the Variance and mean of a targeted variable. The researchers will use this technique to find the best combinations that generate the desired mean value while reducing Variance.

The graphical technique paper presented by Goldfarb *et al.* (2004b) is another essential paper relevant to this research. FDS plots for MPV designs are the focus of this approach. Furthermore, Giovannitti-Jensen and Myers *et al.* (2009) concentrate on prediction variances over design space using a variance dispersion graph (VDG), which allows the experimenter to see patterns of prediction variance in the design space. The FDS plots were initially introduced by Zahran *et al.* (2003), not as a substitute for VDGs but

complementary approach. The FDS plots provide sufficient information on the prediction variance distribution over the experimental area. This method was used by Goldfarb *et al.* (2004a) to create FDS plots for mixture designs. They showed that Piepel *et al.* (1993) random sampling technique and shrunken area approach yield equivalent results for fraction design space values and plots. They also provided the global FDS plot and sliced FDS plot over different process area shrinkage values for MPV designs.

FDS plots for split-plot designs are discussed in Liang *et al.* (2006). When the design is entirely well randomized, the scaled prediction variance (SPV) is usually based on the experimental design and presumed model. Due to the covariance of the response affecting the entire plot error variance and subplot error variance, SPV becomes more complicated when SPD is taken into account. To study the relationship between fundamental plot errors and split-plot (subplot) errors, several researchers used the paradigm of variable variance ratio as the basis for FDS plots.

$$d = \frac{\sigma_{\delta}^2}{\sigma_{\varepsilon}^2},$$

They have used sliced FDS plots at different whole plot levels to investigate prediction capability across the entire split-plot area in the design space. They also consider the influence of the variance ratio factor on design efficiency.

Weese's (2010) paper focuses on mixture design screening methods using the Cox polynomial mixture model context. If there are many mixture components or complicated constraints, and the experiment is thought to be a complex process, this method is used. Wanyama's (2013) paper focuses on the response of *Glycine max* (L.) Merrill is based on soil nutrient management. He talks about how soil pH, row spacing, seeding rate, organic

manure from livestock, and inorganic fertilizer affect. Glycine max is a measured response. Our research investigates the impact of soil pH and row spacing of variety seed on the response of soybean using the framework of MPV design within SPD. The soil pH and row spacing of the seed act as the two process variables where the soil pH is considered an uncontrollable process variable (noise variable). The row spacing is a controllable variable.

1.3 Statement of the Problem

Mixture experiments are a common problem in many scientific fields, including food, chemical, pharmaceutical, and process industries. Cornell (1988) defined the mixture experiment in terms of reaction (response), which is thought to be based on the relative proportions of the mixture blends rather than the mixture's different quantities. However, in many industrial processes and agricultural settings, the measurement response is influenced not only by the proportion of mixture components but also by one or more process variables that affect the mixture ingredients' blending properties or the crop's optimal yield. However, due to functional and economic considerations, some process variables (noise variables) are challenging to modify in some cases. According to Goldfarb *et al.* (2004b), these limitations prevent complete randomization of the experimental runs. Many researchers (Chad and Herbek, 2005; Lee *et al.*, 2008; Tittonell *et al.*, 2008; Wanyama, 2013) believe that other variables are also to blame if they are not well considered, such as seed row spacing, seeding rates, soil nutrient management strategies, soil pH. Some factors mentioned above, such as soil pH, row spacing, and seeding rate, are examples of the process variables affecting the optimum yield of the crop if not well managed (Pedersen, 2004).

Deficiency of nutrients such as Nitrogen (N), Phosphorus (P), and Potassium (K), as well as low organic matter content, have a significant impact on soil productivity in Western Kenya, contributing to low crop yield (Wanyama, 2013). The high population growth rate in Western Kenya, particularly in Kakamega County, has put pressure on the farming land (Althof, 2005; Kinyanjui and Wanjala, 2016). As a result, most small-scale farmers practice continuous cropping to meet their food needs, which results in the soil becoming acidic and exchangeable calcium and magnesium levels, as defined by Hossner and Juo (1999) and Mbau *et al.* (2015). Furthermore, Bekunda *et al.* (1997) and Ridder *et al.* (2004) noted that land-use intensification without sufficient nutrient inputs has resulted in low crop yield and increased nutrient removal and deficiencies in many developing countries. As efforts are made by most of the farmers in Western Kenya to restore fertility in the soil, it is clear that both cereals and legumes respond very well to fertilizer with Nitrogen, Phosphorus, and Potassium content applications from a range of sources and rates.

In Africa, natural soil fertility is addressed by applying nitrogen, phosphorus, and potassium fertilizers at low rates. There is always a general expected response of Cereals to NPK fertilizer application at current recommendations. However, the response remains far below the potential level, particularly on-farm due to nutrient deficiencies and imbalances (Althof, 2005; Jaetzold *et al.*, 2016; Tittonell *et al.*, 2008; Wanyama, 2013). However, there has been little investment in research to determine the best method of combining mixture components with SCD of organic fertilizers derived from livestock manure within SPD using a 2^k factorial configuration with a CCD of the process variable. As a result, this study evaluates the impact of MVPD on *Glycine max* production using farm trials in a split-plot pattern.

1.4 Objective

1.4.1 General Objective

The study's main goal was to find an optimal split-plot design for experimenting with a mixture-process variable (MPV) in modeling the yield of soybean (*Glycine max* (L.) Merrill.)

1.4.2 Specific Objectives

The specific objectives of the study were to;

- i) Develop a parsimonious model into consideration with use of split-plot design and MPV.
- ii) Employ the modified MPV model in predicting the yield of *Glycine max* with reasonable split-plot and main plot errors.
- iii) Employ the screening methodology in the framework of a Cox MPV model in modeling and estimating the predicted yield of the specific variety of *Glycine max* using simulation technique.
- iv) Estimate the optimal yield of *Glycine max* in the framework of split-plot structure arrangement, two variety of Blyvoor and R 184.

1.5 Rationale of the Study

Mixture process variable experiments conducted within a split-plot structure have become common in providing solutions to problems encountered in various industrial fields (Hassan *et al.*, 2020). Since it can solve the problem of restricted randomization of mixture-process variable design (Kowalski *et al.*, 2002; Njoroge *et al.*, 2017), several researchers suggest using a split-plot structure if MPV experiments are involved. As a result, MPV architecture for split-plot structure can be used in agriculture to improve food insecurity

faced by small-scale farmers. However, the majority of small-scale farmers lack the expertise, ability, and a desirable design to employ so as to improve on the yield of cereal crops in the region as a result of declining crop productivity due to soil nutrient deficiencies caused by a lack of Nitrogen, Phosphorus, and Potassium, as well as low organic content in the soil. According to Wanyama (2013), the livestock industry consumes around 80% of *Glycine max*, with domestic use responsible for about 20%-30%.

Many experts have predicted that the demand for *Glycine max* will grow to about 150,000 tons per year in the coming decades (Jagwe and Nyapendi, 2004; Karuga and Gachanja, 2004; Song *et al.*, 2009). There is a need to increase *Glycine max* (L.) Merrill production as shown in Figure 1 when this situation is considered in order to supply the deficit which is usually met through imports (MOA, 2006). Currently, the option of increasing the area of land under *Glycine max* (L.) Merrill is not feasible (Wanyama, 2013) because the high Kenyan population growth rate is currently about 47 million people, with most families owning just a small land plot (Kinyanjui and Wanjala, 2016). Further, some parts of Kenya, such as North Eastern and Turkana, where most parts are inhabitant by few people, are regarded as semi-arid; hence a decline in crop production is witnessed within this region. Therefore, appropriate design in conjunction with proper nutrient management under intensive agriculture gives a better option to increase *Glycine max* productivity under limited land resources.

One of the most potent ways to improve productivity is to use micronutrients in *Glycine max* production (Mbau *et al.*, 2015) such as Zinc that is essential for the formation of chlorophyll and growth hormones and improving plant water uptake (Wanyama, 2013).

Also, the process of di-nitrogen fixation and enzyme activation is handled very well with Molybdenum (Zhang *et al.*, 2020).



Plate 1. 1: Typical production of *Glycine max* on scale farm

Other researchers elsewhere have shown that micronutrients balance crop physiology and play a significant role in the gaseous exchange (De Bruin and Peterson, 2008; Isaev *et al.*, 2020). Moreover, apart from organic fertilizer enhancing nutrients in the soil, lime application in *Glycine max* (L.) Merrill fields raise the soil's pH and decrease the toxic concentration of aluminum and manganese, as reported by Gentili *et al.* (2018). Furthermore, this study seeks to evaluate and assess MPV experiments within SPD on soybean production in Western Kenya, particularly in Kakamega County, based on-farm trials.

1.6 Significance of the Study

Overall, the study results are intended to assist farmers in increasing soybean yields, which are one of Kenya's primary sources of protein. As a consequence, this would help to resolve food insecurity as well as protein deficiency.

1.7 Scope of the Study

The scope of the study was limited two varieties of *Glycine max* (Blyvoor and R 184) with four treatment of organic manure (Goat, Cow, Chicken and Sheep) in the presence of variation created by two process variable (seeding rate and soil pH) within split-plot design. The study was carried in Lugari Sub County in Kakamega County which lasted for a period of nine months.

1.8 Assumption of the Study

The following were the assumption of the study:

- (i) Uniform fertility across whole land.
- (ii) Moisture content in the soil uniform across.
- (iii) All legumes crop can be grown on farmland, and that legume production is priority.
- (iv) The effect of NPK and CAN fertilizer is uniform across the farmland.

1.9 Thesis Outline

Chapter one gives the overview of the study, theories behind MPV, statement of the problem, objectives of the study and significance of the study.

In Chapter 2, the study looked at the literature on mixture and MPV, SPD, graphical methods for assessing different design choices, design optimality, screening methodology in mixture design, Cox mixture model, and the impact of soil pH, row spacing, seeding rate, organic manure, and inorganic fertilizer on *Glycine max* productivity.

We implement MPV design within the split-plot framework formulating the model in Chapter 3, and the SPD is built by considering various design options on mixture process

variable designs. Various estimation methods for parameters of the model within SPD are discussed. We also explore and discuss the FDS and sliced FDS plots developed for MPVD within a split-plot structure. FDS plots are used to analyze and compare various design choices according to the prediction capability to develop desirable designs. The design efficiency is also provided for various design choices. In addition, the study looked at statistical designs for experiments involving mixture variables and hard-to-change process variables and restricted randomization for the experimental runs to experiment on *Glycine max*. The study also explored and discussed the material and method involved in experimenting on *Glycine max* (L.) Merrill.

In Chapter 4, the study provides the results and discussion. The results is given based on simulation and actual results obtained experimentally. The study made a comparison between the predicted and actual outcomes. In addition, we compared and contrasted the effects of MPV design within a split-plot structure and the one that was simulated. The influence of process variables on the response of *Glycine max* (L.) Merrill was also discussed. The effect of mixture component design space on the response of *Glycine max* was compared using graphical tools.

Finally, in Chapter 5, the conclusion and recommendation are presented. In addition, the study gives a summary of the critical contribution of MPV design within the split-plot structure on Legume crop (*Glycine max* (L.) Merrill) and suggest potential research areas.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

MPV experiments are encountered in several fields, including agriculture and industry. Cornell (2001) describes the MPV experiment in detail. Not only did Goos and Donev (2007), Goldfarb *et al.* (2003) presented MPV experiments with difficult to modify variables in practice, but they did not recognize randomization problems. As Chung *et al.* (2009) discovered, when the process variable is included in the mixture experiment, the number of runs dramatically increases, making complete randomization impractical. As a result, Cho *et al.* (2009) proposed a split-plot design to cope with restricted randomization.

In mixture experiments, several designs are available as pointed out by Lawson and Willden (2016). For example, according to Sitinjak and Syafitri (2019), the design with the smallest number of experimental runs is often preferred if it offers sufficient details on the model's coefficients. According to Goldfarb and Montgomery (2006), a second feature for design selection is forecasting capability. Conversely, the Scaled Prediction Variance (SPV) is a suggested measure of prediction efficiency that penalizes large designs by considering the total sample size (Liange *et al.*, 2006; Njoroge *et al.*, 2017). Furthermore, when the cost is not the primary concern, an alternative goal is unscaled prediction variance, which compares variance without regard to sample size, as Cho (2010) reported. The critical concern is the estimation variance at a particular position; design efficiency is often a good choice for comparing, analyzing, and assessing various design options. G-, I-, V-, and Q-optimality are architecture optimality parameters that rely on prediction variance. The overall distribution of scaled prediction variance across the design space is

taken into account when evaluating the design's prediction capability instead of evaluating only a single point prediction calculation, such as G-, I-, or V-efficiency because prediction variances vary at different points (Cho, 2010; Goldfarb *et al.*, 2003; Iwundu, 2017). As a result, the preferred design is a relatively constant SPV across the entire design space.

Giovannitti-Jensen and Myers (1989) implemented a graphical method for spherical design space that shows the experimental field's prediction variance properties. The Variance Dispersion Graph (VDG) is the name of this graphical technique. Rozum and Myers (1991) later expanded this approach to include designs of cuboidal regions. This technique for evaluating various design options in mixture design has wowed many researchers (Njoroge *et al.*, 2017; Sijinjak and Syafitri, 2019). Goldfarb *et al.* (2003) also implemented the three-dimensional VDGs for MPV experiments in their paper. As a supplement to the VDG, Zahran, *et al.* (2003) introduced a new graphical approach called Fraction Design Space (FDS) plot. The FDS plots are created by computing the SPV across the design space and then determining the fraction of the design space that is less than or equal to the SPV values (Liange *et al.*, 2006). Goldfarb *et al.* (2004) later proposed using a random sampling approach for FDS plots in mixture design. Adapted from Liange *et al.* (2006) for fraction concept space plot for SPDs, different whole plot shrinkage levels in an SPD over a spherical area were determined using the sliced FDS curves' relative sizes.

Weese (2010) used the Cox Polynomial mixture model to apply screening methods in mixture settings. This approach is used when the number of mixture components is high. There are complex constraints, and the experiment is thought to be challenging to complete the method aids in defining the model coefficients that are relevant to the response.

We will address more complex research ideas and principles for our analysis in the following section. The following is a stepwise procedure that will be adopted by the current study. First, a general understanding of mixture and MPV experiments is outlined, as well as statistical models for data from mixture components, experimental area, and the experimental situation in mixture design, and robust parameter design for MPV with hard-to-modify factors. Secondly, SPDs are discussed, and the Hasse diagram and predicted mean square. Finally, the best design parameters for mixture process variables are provided, along with graphical techniques like FSD plots and VDG for comparing various design options. Thirdly, the methodology of employing the two process variables used in this research is demonstrated. Finally, a screening technique for mixture design using the Cox polynomial mixture model is presented with an application to *Glycine max* (L.) Merrill crop.

2.2 Mixture Design and Statistical Mixture Models

Let x_1, x_2, \dots, x_q be q mixture components. These mixture components act as explanatory variables in designed experiment subject to

$$\sum_{i=1}^q x_i = X' \mathbf{1}_q = 1, \quad (2.1)$$

where $\mathbf{1}_q$ represent a q –dimensional column vector of ones and $X' = (x_1, x_2, \dots, x_q)$.

Goos *et al.* (2016) showed that this mixture restriction produces a Simplex-shaped experimental region that significantly affects the models that can fit. Cornell (1988) points out that a regression model involving linear terms in mixture blends cannot contain the intercept. Otherwise, as many scholars have suggested, we cannot estimate the model's

parameters uniquely (Njoroge *et al.*, 2017). According to many scholars, a second significant implication is that cross-products of proportions and squares of proportions cannot be used in the study because model parameters are not estimable uniquely (Goldfarb *et al.*, 2004b). It is evident that for each proportion x_i , this is the case. The square of a proportion x_i^2 is, in most cases, a linear combination of that proportion and its cross-products with any of the other $q - 1$ mixture blends, and is given as:

$$x_i^2 = x_i \left(1 - \sum_{j=1, j \neq i}^q x_j \right) = x_i - \sum_{j=1, j \neq i}^q x_i x_j, \quad (2.2)$$

Scheffe (1958) proposed the Scheffe mixture models, which take these considerations into account and describe the first order Scheffe model as

$$\zeta(x) = \sum_{i=1}^q \beta_i x_i. \quad (2.3)$$

The second-order Scheffe model, on the other hand,

$$\zeta(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j, \quad (2.4)$$

While a unique cubic model is such that

$$\zeta(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{k=j+1}^q \beta_{ijk} x_i x_j x_k, \quad (2.5)$$

And finally, the full cubic model is:

$$\begin{aligned}
\zeta(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \gamma_{ij} x_i x_j (x_i - x_j) \\
& + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{k=j+1}^q \beta_{ijk} x_i x_j x_k.
\end{aligned} \tag{2.6}$$

where $\zeta(x)$ denotes the predicted response, β_i denotes the regression coefficient linear term while β_{ij} , γ_{ij} and β_{ijk} represents the regression coefficient of interaction terms.

Furthermore, Scheffe (1963), Cornell (2011), Smith (2005) and Goos *et al.* (2016) advocated that the q^{th} polynomial degree model suitable in modeling the interaction effect resulting from mixture components. However, in the literature, this degree model has gotten a lot of coverage. But, due to an increase in the number of unique higher-order terms in special cubic models, which becomes tedious during model parameter estimation, this model is not widely used.

2.2.1 Simplex Lattice Design (SLD)

A $\{q, m\}$ Simple Lattice Design (SLD) for q mixture components entails all possible mixture formulations, each of which has a q unique mixture part that belongs to the set $\{0, \frac{1}{m}, \frac{2}{m}, \dots, 1\}$. As a result, the total number of design points in a $\{q, m\}$ SLD is given as

$$\binom{m+q-1}{m}.$$

A (3, 1) SLD, for example, has three candidate points (1,0,0), (0,1,0), and (0,0,1). Pure mixture components are what these points are called (Goos *et al.*, 2016). For the case a

(3,2) SLD involves 6 candidate points, the points $(0.5, 0.5, 0)$, $(0.5, 0, 0.5)$, $(0, 0.5, 0.5)$ and as well as the pure blends as illustrated in Figure 2.1, 2.2 and 2.3.

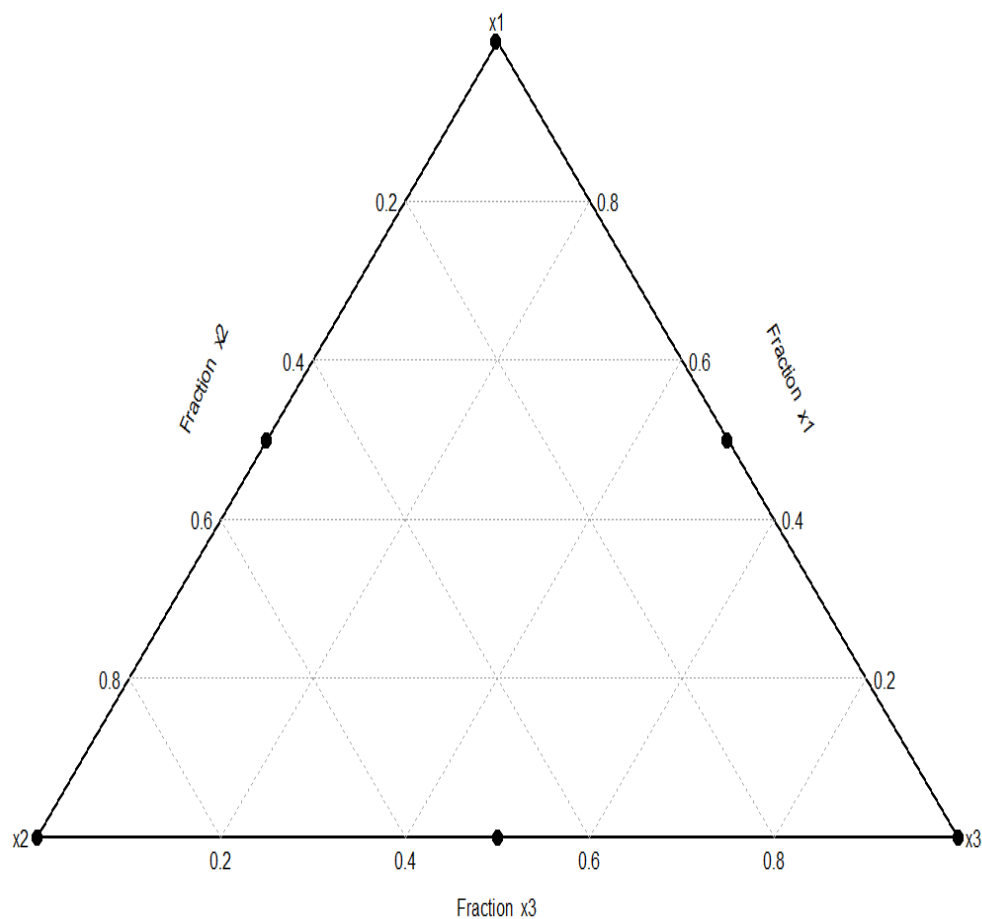


Figure 2.1: SLD with 3 mixture components

The points involving 50% of one mixture blend and 50% of another are commonly referred to as binary mixtures.

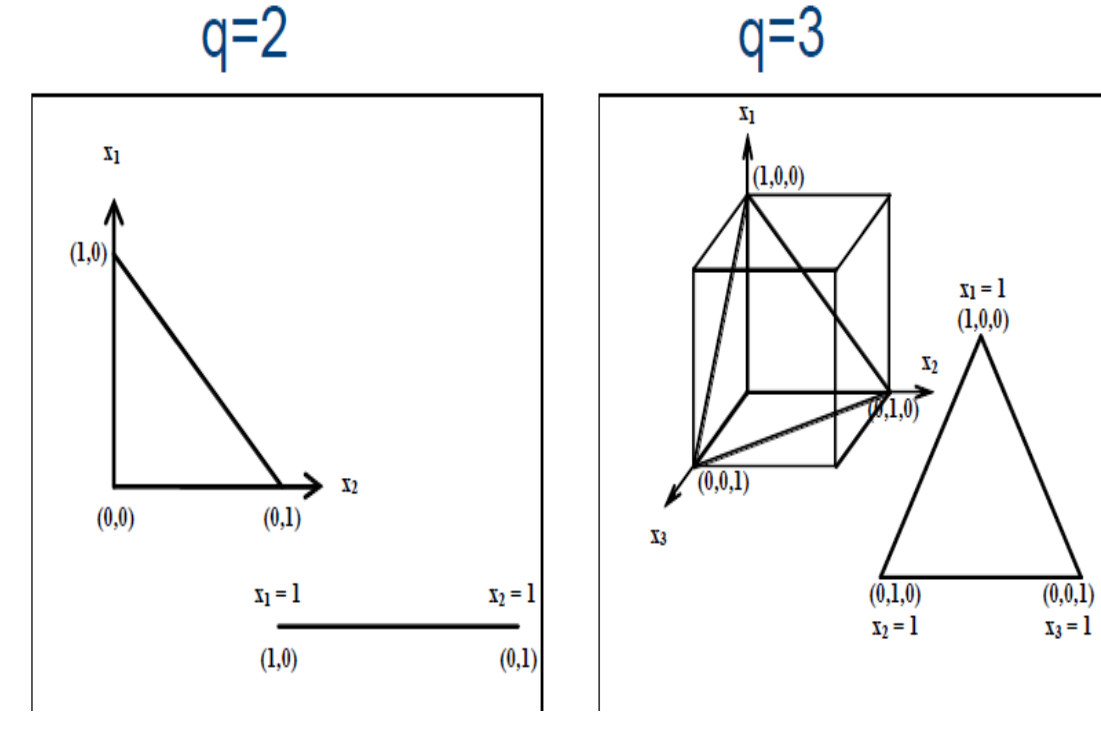


Figure 2.2: The comparison of factor space within simplex lattice design with the case of $\{2, 1\}$ and $\{3, 1\}$ SLD, respectively.

In totality, they are q pure blends and $(q2) = \frac{n(n-1)}{2}$ binary mixture as pointed by Goos *et al.* (2016).

However, double simple lattice design also exists in the form of $\{q, m; q, m\}$ as illustrated in Figure 2.4. According to Cornell (2011), double SLD implies double mixture where each mixture itself is a mixture or a mixture of mixtures, blended with a proportion and $1 - p$ defined by multiple component constraint equalities as $\sum_{i=1}^q X = p$ and $\sum_{i=1}^q Y = 1 - p$.

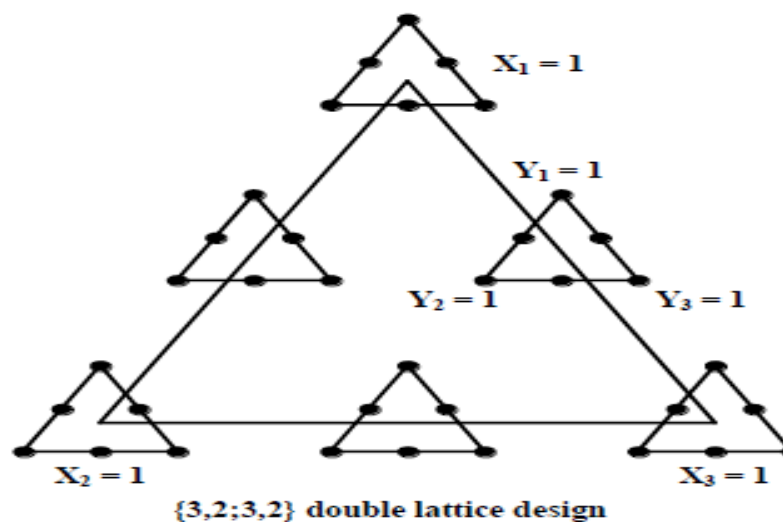


Figure 2.3: Mixtures of mixtures within SLD for three mixture blends in a second order model

The SLD can also be depicted using Figure 2.5 in the case of a four-component mixture.

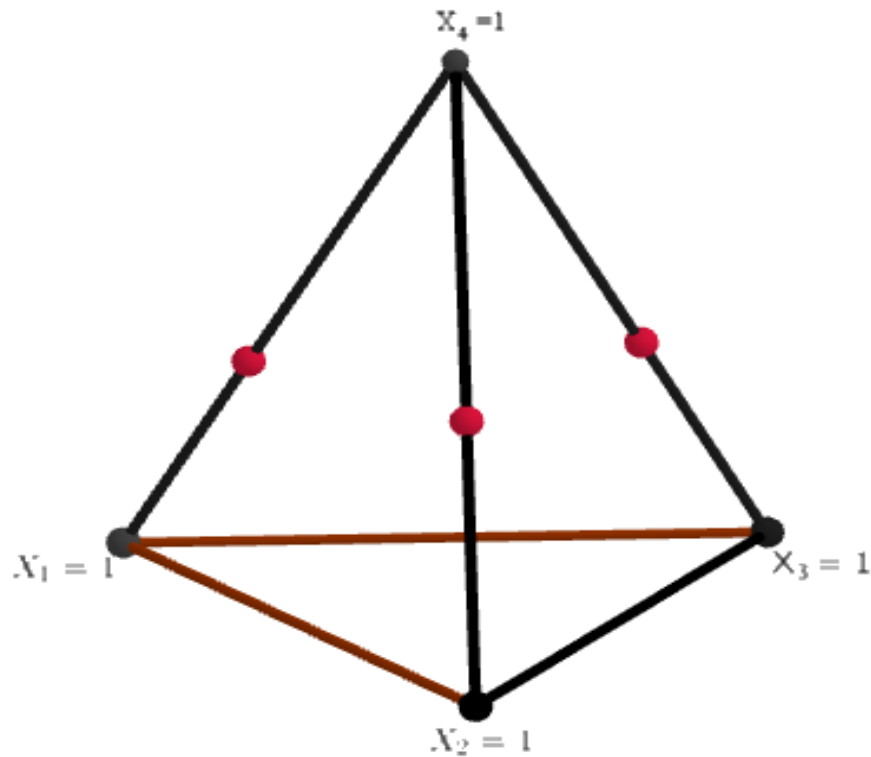


Figure 2.4: A four-component mixture of SLD

The second order model parameters can be estimated using this fraction as shown in Figure 2.7.

2.2.2 Simplex Centroid Design (SCD)

The complete SCD consists of $2^q - 1$ design points: the q pure blends, the $(q, 2)$ binary mixture ingredients, the $(q, 3)$ ternary mixture mixes permutations $\left\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \dots, 0\right\}$, and finally, the q permutations of the mixture ingredients given as $\left\{\frac{1}{q-1}, \frac{1}{q-1}, \frac{1}{q-1}, \dots, \frac{1}{q-1}, 0\right\}$ as provided by Cornell (1988). A $(3,2)$ SCD, for example, includes seven design points, pure

component of mixture, binary mixture points $(0.5,0.5,0)$, $(0.5,0,0.5)$, $(0,0.5,0.5)$, and the centroid points $\left\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right\}$, as illustrated in Figure 2.5.

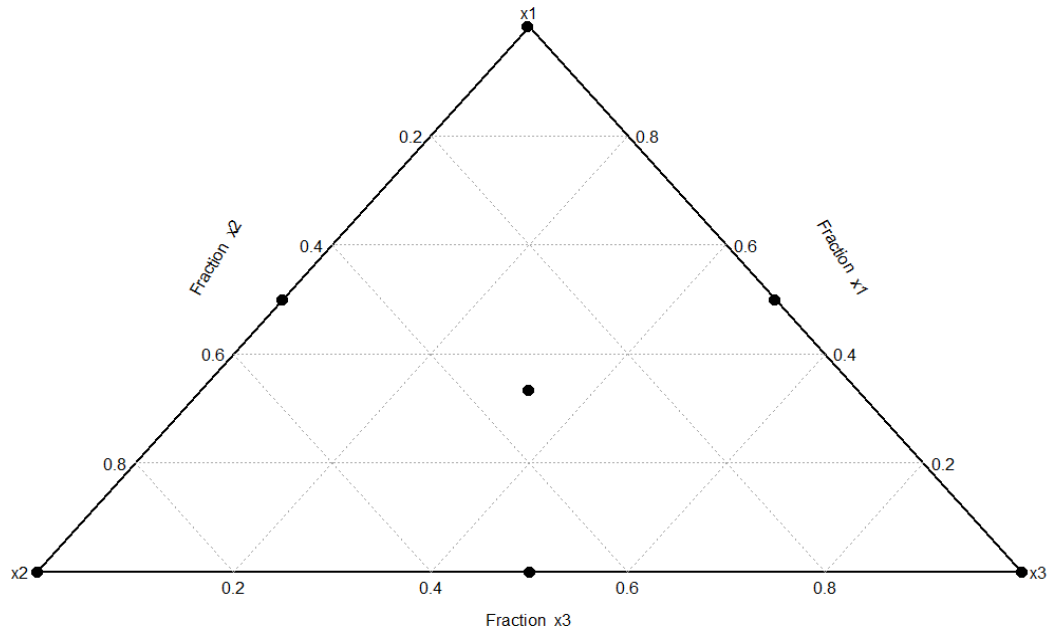


Figure 2.5: A second order model, a SCD of three mixture components.

This example indicates that for every number of mixture ingredient q , there is only one SCD as shown in Figure 2.5 and 2.6 with a black dot at center, but rest is the family of SLD. Moreover, SCD involves the overall centroid given as $\left\{\frac{1}{q}, \frac{1}{q}, \frac{1}{q}, \dots, \frac{1}{q}\right\}$ and $\left\{\frac{1}{q-1}, \frac{1}{q-1}, \frac{1}{q-1}, \dots, \frac{1}{q-1}, 0\right\}$ are the centroids of all lower dimensional simplexes' as in Figure 2.7. An important fraction of the SCD involves the pure blends, the binary and ternary mixtures (Scheffe, 1963). However, the special cubic model is estimated using these fractions. The fraction being referred in this case is as the $(q, 3)$ SCD. Further, a

fraction as the $(q, 5)$ SCD involves quaternary and quinary mixtures because of a larger fraction of the SCD.

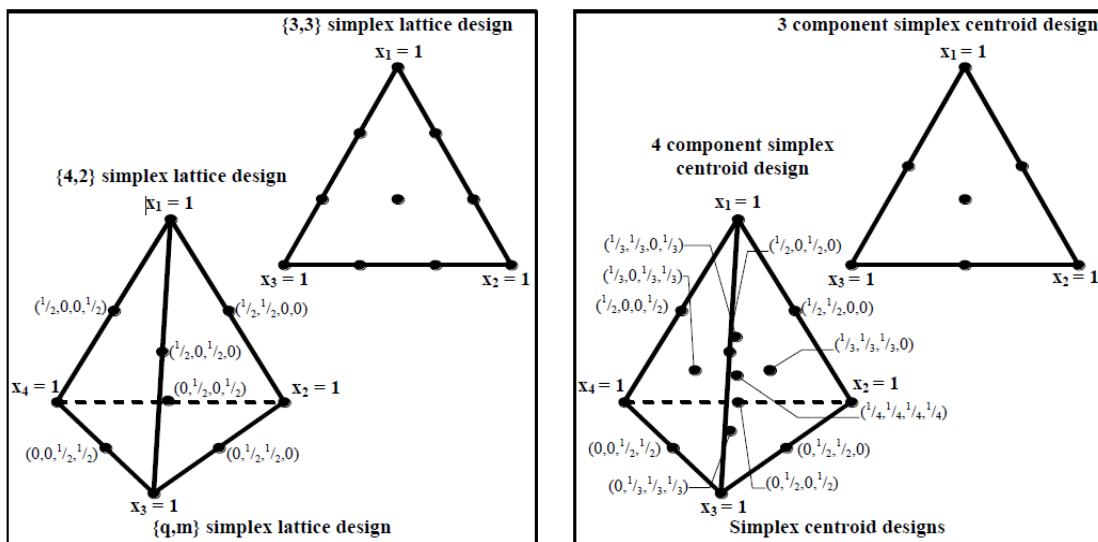


Figure 2.6: The comparison between SCD and SLD for q mixture blends with m degree polynomial model.

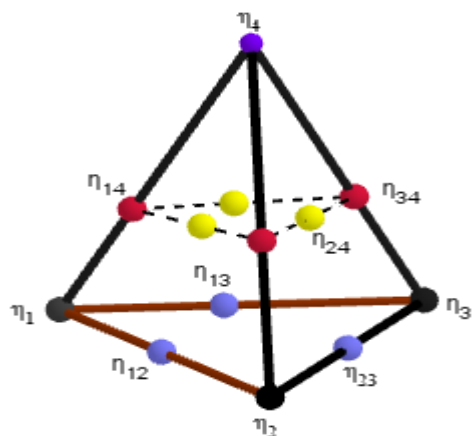


Figure 2.7: The expected response nomenclature for four component mixture in simplex lattice design

2.2.3 Mixture Design with Presence of Process Variable

As Cho (2010) points out, the critical subject of attention in the mixture literature is uncombined mixture product process variable experiment has been on the components' blending properties for decades, with only a secondary interest in process variables' effects. Whereas, since some researchers conclude that in most industrial environments, the process variables' critical problem is at least equal to that in the mixture ingredients, the process variables are often regarded as noise factors (Cornell, 2011; Goldfarb *et al.*, 2004b).

Response surface methodology (RSM) traditionally applies a second-order Taylor series as the appropriate model basis for process optimization (Goos and Donev, 2007; Vining *et al.*, 2005). This assumption relies typically on sufficient background knowledge besides knowing the experimental region that supports an accurate second-order model described by Kowalski *et al.* (2000). The mixture elements, process variable (PV), and MPV interaction are always equal. In reality, according to Goos and Donev (2007) and Cho (2010), the mixture PVs interaction terms also provide considerable insight into optimal operating conditions.

In the polymer experiments proposed by Cornell (2011) and Myers *et al.* (2009), the experimenter may be interested in learning about a particular mixture component that makes the reaction particularly sensitive to the reaction temperature. This experimental situation led some authors (Cornell, 1988; Chung *et al.*, 2009; Cho, 2010) to propose a new model for approaching the MPV experiments simplex centroid design with the split-plot structure experiments.

2.2.4 The Combined MPV Model

The shape of a blended model involving terms in mixture properties and process variables in MPV design is influenced by the mixture components' mixing characteristics, the effect of process variables, and any interactions between the mixture blends and process variables. These models are usually Scheffe second-order models described in Scheffe (1963). As Goos and Donev (2007) point out, they allow for pure quadratic and two-factor interaction terms. Cornell (2011) presented the following general second-order polynomial for q mixture variables, n process variables, and the combined MPV model:

For the q mixture vector, a second order model for n processes variables, z_1, z_2, \dots, z_n , is also given by

$$\zeta(\lambda, z) = \lambda_0 + \sum_{k=1}^n \lambda_k z_k + \sum_{k=1}^{n-1} \sum_{l=k+1}^q \lambda_{kl} z_k z_l \quad (2.7)$$

The Equation (2.4) and (2.7) are merged to create combined MPV model. However, crossing the mixture model terms in Equation (2.4) with any term in Equation (2.7) produces the first type of model. As a result, the combined model is given as.

$$\begin{aligned} \zeta(x, z) = & \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \sum_{i=1}^q \sum_{k=1}^n \gamma_{ik} x_i z_k + \sum_{i=1}^q \sum_{k=1}^{n-1} \sum_{l=k+1}^n \gamma_{ikl} x_i z_k z_l \\ & + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \sum_{k=1}^n \gamma_{ijk} x_i x_j z_k + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{k=1}^{n-1} \sum_{l=k+1}^n \gamma_{ijkl} x_i x_j z_k z_l. \end{aligned} \quad (2.8)$$

This model (2.8) contains parameters for three and four-factor interactions, which provide a measure of the mixture blends' linear and nonlinear blending properties averaged across

the process variable settings and the effects of the process variables on the linear and nonlinear properties depending on the configuration.

However, combining the models in Equation (2.4) and (2.7) without crossing any of the x_i and z_j terms leads to the second type of combine model is given as

$$\zeta(x, z) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \sum_{k=1}^n \alpha_k z_k + \sum_{k=1}^{n-1} \sum_{l=k+1}^n \alpha_{kl} z_k z_l. \quad (2.9)$$

This model (2.9) calculates the quadratic blending of the mixture blends on the response and up to two factor PV interactions on the response. Furthermore, model (2.9) is based on the premise that mixing the mixture ingredients is the same for all factor level combinations of the method variables. This is because this model does not provide any cross-product terminology between the mixture ingredients and the process variables, even though Goldfarb *et al.* (2003) pointed out that this assumption is often impractical in most cases. In fact, in some experiments like the polymer experiment presented by Cornell (2011), the mixture of ingredients by process variables interaction may be the most crucial importance in the model. For instance, the temperature in those experiments acted as a process variable and variation of this process variable determines the end product whether it is of good or poor quality beside the other factor components.

Over a decade, the number of runs with mixture experiments involving process variables has been the major concern by several researchers (Anderson Cook *et al.*, 2004; Cho, 2010; Njoroge *et al.*, 2017). This has been due to a rise in the number of method variables and mixture ingredients, causing Equation (2.8) to require many candidate points automatically. Therefore, this becomes costly when running the experiment due to size and

time required (Sitinjak and Syafitri, 2019). Also, it becomes tedious in estimation parameters due to the large number of runs as described in Wanyonyi *et al.* (2018). This is the reason behind the use of small size of experiments by several industrial situations due to cost, time and/ or other constraints (Njoroge *et al.*, 2017).

However, Cornell (2011) claims that fitting model (2.9) allows for a smaller design than fitting model (2.8) to minimize the experiment's size. Still, it does not discuss the estimation of mixture components by process variable interaction. As a result, some authors (Anderson Cook *et al.*, 2004; Cho, 2010) concluded it is a compromise between these two models, one of which is necessary because the interactions between mixture ingredients and process variables are thought to be crucial.

A Taylor series approximation has been the critical method for most of the models proposed for response surface investigations over the centuries (Goos and Donev, 2007). Following this tradition, Cho *et al.* (2009) proposed the proper model for the n process variables as a replacement for the model in Equation (2.7) given as

$$E(Y) = \lambda_0 + \sum_{k=1}^n \lambda_k z_k + \sum_{k=1}^n \lambda_{kk} z_k^2 + \sum_{k=1}^{n-1} \sum_{l=k+1}^n \lambda_{kl} z_k z_l, \quad (2.10)$$

This model (2.10) has the n pure quadratic terms that give a clear distinction from model (2.7) Furthermore, this model (2.10) is combined with model (2.4) to generate another combined secondary model given as

$$(2.11)$$

$$\zeta(x, z) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \sum_{k=1}^n \lambda_{kk} z_k^2 + \sum_{k=1}^{n-1} \sum_{l=k+1}^n \lambda_{kl} z_k z_l + \sum_{i=1}^q \sum_{k=1}^n \delta_{ik} x_i z_k.$$

This model includes the mixture model and pure quadratic and two-factor interaction effects among process variables, as well as two-factor interactions between the linear blending terms in the mixture ingredients and the process variables' main effect terms. The crossed model in Equation (2.8) and the additive model in Equation (2.9) require a minimum number of design points.

Several authors argue that model (2.11) should be used without the pure quadratic definition, assuming that

$$\sum_{k=1}^n \lambda_{kk} z_k^2 \cong \sum_{k=1}^{n-1} \sum_{l=k+1}^n \lambda_{kl} z_k z_l \quad \text{and} \quad \sum_{k=1}^q \beta_{ii} x_i^2 \cong \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j. \quad (2.12)$$

The fitting model (2.11) also necessitates a design that supports nonlinear mixing of the mixture ingredients as well as the fitting of the complete second-order model method variables, according to Cornell (2011), Kowalski *et al.* (2000), and Goos *et al.* (2016).

As in most RSM investigations, we explore how to choose suitable designs that support the fit of the stated model involving MPV in this section (Demirel and Kayan, 2012). The design method is discussed in this paper by Cornell (1990), Kowalski *et al.* (2000), and Goos *et al.* (2016). According to Goos and Donev, (2007), the best design has at least as many candidate points as the model's number of parameters. As a consequence, this formula specifies the design point needed to support the model (2.11) fitting:

$$\frac{(q + n)(q + n + 1)}{2}, \quad (2.13)$$

Where n stands for the number of process variables.

In recent years, the central composite design has become the most popular response surface design for fitting a second-order mixture model, consisting of a complete 2^n factorial design, $2n$ axial or star points with levels $\pm\alpha$ for one factor and the control points, and at least one center point replicate as defined by Cho (2010), Kowalski *et al.* (2002) and Goos *et al.* (2016).

Several authors recommend that the designed experiment should always begin with CCD in PVs as a method for reducing the number of observations needed in a process variable experiment involving a combined mixture of ingredients (Cornell, 2011; Hassan *et al.*, 2020; Wang *et al.*, 2013). Furthermore, as pointed out by Njoroge *et al.* (2017), a simplex centroid architecture is then implemented at each point in CCD by considering only a fraction of the mixture blends. According to Goos *et al.* (2016), a simplex centroid design consists of $2^q - 1$ candidate points. However, there are S_q centroid points of a $q - 1$ dimensional simplex space inside these candidate points, in addition to the centroids of all

the lower dimensional simplices members of S_q and at q vertices of S_q . Furthermore, the mixture ingredients chosen at various CCD points are usually a subset of full SCD.

As several authors (Cho et al., 2009; Weese, 2010) have pointed out, there is often a need to achieve equilibrium among the mixture ingredients across the process variables when implementing experiments. In reality, the same number of mixtures blend must be present at both high and low levels of each process component. This is because traditionally, at each \pm the factorial level of each process variable normally requires all the mixture components to present. These facts seem very intuitive and often makes researchers apply some mixture components at certain design points and also different mixture ingredients at other candidate points.

Moreover, other researchers elsewhere have established that the use of the simplex vertices should be run at one-half of 2^n factorial points in the PV, while the simplex's mid edge points should be run at the other half. A simplex is created when a design is collapsed over each PV level, with vertices and midpoints at both the low and high levels of remaining PVs.

Furthermore, the PVs' axial points are combined with the simplex's centroid for the centroid to be present even if the geometry is collapsed. As a result, two distinct designs emerge, each differentiated by the number of points positioned at the process variable's base. For instance, suppose one of the designs is ξ_A , and the other design is ξ_B where ξ_A is the full simplex centroid while ξ_B is having the centroid design mixture blends only, which is performed at the mid of the process variable. In other words, this implies that with design (ξ_A), the full simplex centroid is executed at midway whereas design (ξ_B), the centroid

concoction is blended at the center of the variable process. These two designs can be illustrated using an example presented by Kowalski *et al.* (2000) involving three mixture ingredients in order to understand and have clear distinction between ξ_A and ξ_B . The Figure 2.8 shows the proposed design for both model ξ_A and ξ_B . The model (2.11) contains fifteen terms. ξ_A in Figure 2.8 represented with both closed and open circles which tally to twenty-three design points while ξ_B in the same Figure is denoted with open circles which sums up to 17 candidate points. The terms in model (2.11) can be estimated with either ξ_A or ξ_B . The design points in Figure 2.8 can also be represented tabular form as presented in Kowalski *et al.* (2000). In cases where the mixture ingredients include the lower and upper bound limits, the proposed design in Figure 2.9 becomes more difficult to use, resulting in a more complicated mixture region than the simplex. An irregular polygon is usually used as the restricted experimental area. The original components can be converted into L-Pseudo components, which solves the problem. L-Pseudo components are a technique for making design construction and model fitting simpler when a constraint is present in the design space and are described by Kowalski *et al.* (2000) as

$$X'_i = \frac{(x_i - L_i)}{(1 - \sum_{i=1}^q L_i)}, i = 1, 2, \dots, q \quad (2.13)$$

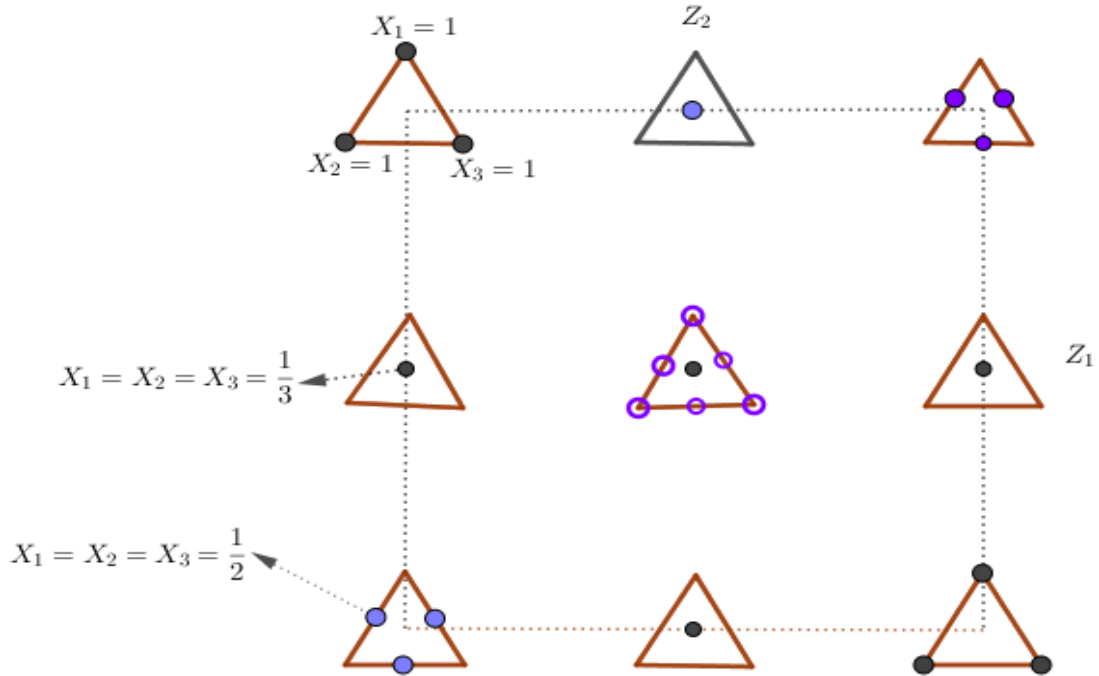


Figure 2.8: Proposed designs for the 3-2 Case (ξ_B is the design with closed circles, ξ_A is ξ_B plus the open circles) (Kowalski *et al.*, 2000)

Where L_i represent the lower bounds in each constraint mixture ingredients. For example, the following constraints given by Cornell, (2011)

$$0.25 \leq x_1 \leq 0.40, \quad 0.25 \leq x_2 \leq 0.40, \quad 0.25 \leq x_3 \leq 0.40$$

can be transformed into L-Pseudo components as

$$X'_i = \frac{(x_i - 0.25)}{(1 - \{0.25 + 0.25 + 0.25\})} = \frac{(x_i - 0.25)}{0.25}, i = 1,2,3. \quad (2.14)$$

In addition, since the candidate points for the two designs, in this case, consist of the six vertices and midpoints of the six edges of the hexagon, as well as the overall centroid at

each of the PV design points, the resulting mixture region for three mixture components is a hexagon. The resulting L- Pseudo components for Figure 2.8 using the restriction example above is shown in the Table described in Kowalski *et al.* (2000). While exploring scenarios involving mixture ingredients without or with restrictions is not meant to be comprehensive, it is used because it encompasses the various options concerning a typical industrial or agricultural experiments (Wang *et al.* 2013); Hassan *et al.*, 2020). Furthermore, based on the assumptions, the designs may be applied to higher dimensions without creating problems, even though they might be challenging to display geometrically.

Cornell (1990) and Kowalski *et al.* (2000) proposed using CCD for design extension. For example, in the case of three mixture ingredients, CCD can extend the case. The centroids of the simplex should be placed at each of the CCD axial point, while the centroid or the seven point's simplex centroid should be placed at the center points.

When the number of design points is less than the total number generated by crossing CCD in the MPV settings, the typical user employs statistical software such as SAS's JMP division. It allows a researcher to generate a design based on a criterion of optimality, such as D- or I-Optimality. As a result, while such generated designs are statistically optimal, according to several researchers (Goos *et al.*, 2016; Njoroge *et al.*, 2017) since there may exist a design that is nearly optimal and more attractive and appealing due to existence of properties such as Symmetry and nearly orthogonality although this computer-generated design should not be accepted blindly as suggested by Snee (1985), Kowalski *et al.* (2000), and Goos *et al.* (2016). Additionally, computer-aided designs are usually helpful when the

experimental area is irregular, and the number of available runs is limited and can also be suitable for non-linear models.

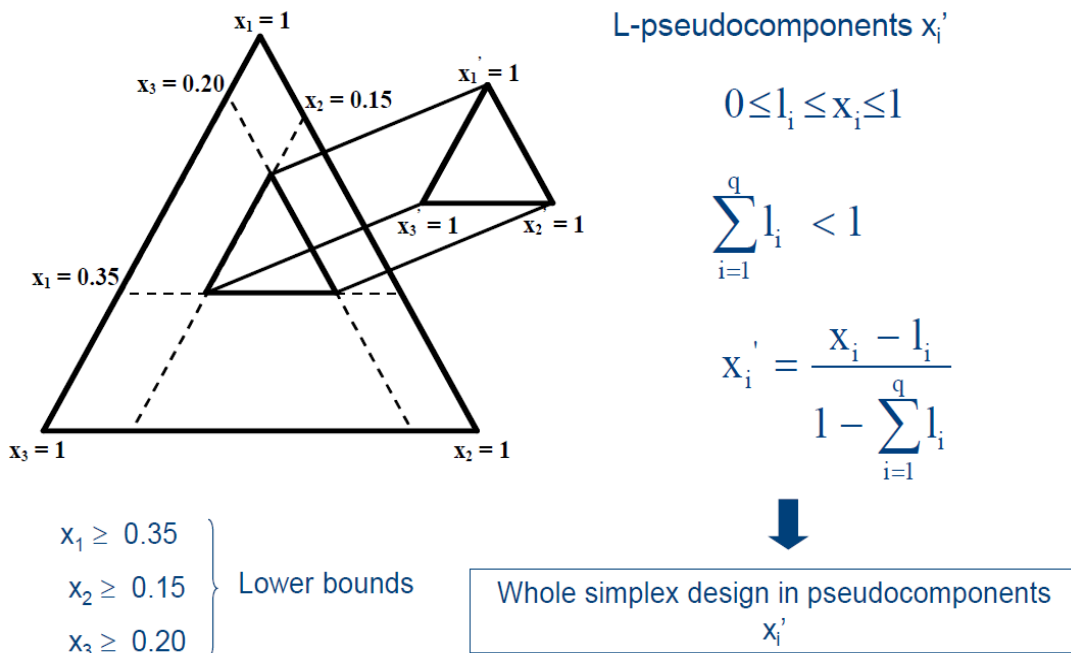


Figure 2.9: Illustration of L- Pseudo components (Cornell, 2011).

2.3 Split-Plot Layout Structure

The split-plot layout structure is an experimental design applied to experiments involving situations where complete randomization is not feasible to address that problem. According to Njoroge *et al.* (2017), randomization is always used in designed experiments when changing some of the factors is difficult or costly to solve. This could be due to physical constraints on the process variable. As a result, some researchers argue that it is necessary to limit the randomization of experimental runs, which leads to the split-plot design described by Cho (2010).

The split-plot experiment is typically randomized in two ways: the Hard-to-Change Factor (HCF) and the Easy-to-Change Factors (ECF) respectively. The HCF combinations are also referred to as whole plot treatment factors, whereas ECF is referred to as split-plot or subplot treatment (factor). The experiments in a split-plot layout structure usually are carried out by first fixing the levels of the HCFs and then running some or all of the combinations of the ECF levels. In the HCFs, a new setting is chosen. The process is then repeated until all of the ECF combinations have been exhausted (Cornell, 1988).

The HCF is frequently assigned at random to the entire plot, subject to the whole-plot design. The ECFs are then restricted and randomly assigned to subplots within each whole plot, with separate complete randomization for each whole plot (Kowalski et al., 2002). Furthermore, the subplot effect is produced due to the primary treatment's sub-treatment interaction with the split-plot factor. Nevertheless, in split-plot design analysis, the response surface methodology is used to determine this factor's optimum. The, pure error estimates for the two variance components, whole plot and the subplot are usually estimated together with robust parameter values for the fixed factors (Box and Jones, 1992; Hassan, 2020). Cho (2010) presented the statistical model for the split-plot design experiment with two factors as

$$y_{ijk} = \mu + \varrho_i + \beta_j + (\varrho\beta)_{ij} + \vartheta_k + (\varrho\vartheta)_{ik} + (\vartheta\beta)_{jk} + (\varrho\beta\vartheta)_{ijk} + e_{ijk} \begin{cases} i = 1, 2, \dots, r \\ j = 1, 2, \dots, a \\ k = 1, 2, \dots, b \end{cases} \quad (2.15)$$

Where,

r = indicate the number of times the experiment is replicated,

a = denotes the levels of the main treatment (whole plot treatment),

b = represent the levels sub plot treatment,

q_i = denotes replicates or blocks of the design,

β_j = represent whole plot treatment (say factor A with a levels),

$(q\beta)_{ij}$ =indicates the total plot error measurements (replicates or blocks A) where

$$(q\beta)_{ij} \sim N(0, \sigma_\beta^2),$$

ϑ_k = represent subplot design with sub plot treatment (say factor B with b levels),

$(q\vartheta)_{ik}$ = denotes the sub plot structure with (replicates or blocks \times B),

$(\vartheta\beta)_{jk}$ = represent sub-treatment interaction between main treatment and split plot factor

(AB) that produces the subplot effect,

$(q\beta\vartheta)_{ijk}$ = denotes the subplot error that originates from three factor interaction of replicates/blocks \times AB),

e_{ijk} = represent the generally sub plot error where $e_{ijk} \sim N(0, \sigma_e^2)$,

μ = Denotes constant mean population.

y_{ijk} = Measured response

Similarly, Cho (2010) derived the expected mean squares (**EMS**) as shown in Table 2.2 for split plot layout structure, with replicates or block random, whole plot treatment and sub plot treatment. Table 2.1 shows that the whole plot treatment (main factor A) is typically tested against the whole plot error. In contrast, the subplot treatment (B) is usually tested against the blocks/ replicates of subplot treatment interaction. Furthermore, the sub-treatment interaction AB is compared to the subplot error.

Table 2.1: The EMS for SPD 1

	Model Terms Subject to Treatment (Factors)	EMS
Whole-plot layout	q_i	$\sigma_e^2 + ab\sigma_q^2$
	β_j	$\sigma_e^2 + b\sigma_{q\beta}^2 + \left(rb \sum \beta_j^2 / a - 1 \right)$
	$(q\beta)_{ij}$	$\sigma_e^2 + b\sigma_{q\beta}^2$
	ϑ_k	$\sigma_e^2 + a\sigma_{q\beta}^2 + \left(ra \sum \vartheta_k^2 / b - 1 \right)$
Subplot structure	$(q\vartheta)_{ik}$	$\sigma_e^2 + a\sigma_{q\vartheta}^2$
	$(\beta\vartheta)_{jk}$	$\sigma_e^2 + a\sigma_{q\beta\vartheta}^2 + \left(r \sum \sum (\beta\vartheta)_{jk}^2 / (b-1)(a-1) \right)$
	$(q\beta\vartheta)_{ijk}$	$\sigma_e^2 + a\sigma_{q\beta\vartheta}^2$
	e_{ijk}	σ_e^2 (not estimable)

Box and Jones (1992) compared the efficiency of a completely randomized design (CRD) and SPD and found out that sub plot error variance was less than the whole plot error variance ($\sigma_e^2 < \sigma_{CRD}^2 < \sigma_{Wpeu}^2$). This demonstrates that SPD is usually more efficient than CRD. Furthermore, Goos and Donev (2007) in their earlier research investigated the use

of SPD in industrial experiments where one factor is known to be difficult to change. However, after the simulation study, their findings confirmed that SPD produced increased precision on subplot factors. According to Cho *et al.* (2009), SPD should always be an excellent alternative source of experiment whenever the experiments to be carried out are impractical, such as in industry, because SPD is much more efficient and easy to run than CRD. The number of runs designed for any experiment tends to increase dramatically in MPV experiments, as do the number of process variables (Kowalski *et al.*, 2002; Njoroge *et al.*, 2017). Conversely, the HCFs are notoriously tricky to adjust and control. As a result, randomization is always a significant condition underlying statistical methods when designing experiments (Hassan *et al.*, 2020). Although it is always impractical to perform the experiments in a completely randomized order, changing the levels of some factors can be difficult or expensive. In this case, limiting the randomization of experiments becomes critical, resulting in the split-plot structure described by Box and Jones (1992). Cho *et al.* (2009) developed graphical evaluation techniques for MPV within SPD, but they did not consider robust parameter design for not easily controlled variables (noise variable).

2.3.1 Hasse Diagrams and Expected Means Squares within SPD

The Hasse Diagrams are graphical representations of models that show nesting and random or fixed structures. This method is similar to ANOVA Tables. It has node M at the top representing the grand mean and anode E at the bottom representing random error.

Oehlert (2010) presented a model for a single whole plot factor (WPF) A with a levels, a single split plot factor B with b levels, and n whole plots for each level of A as

$$y_{ijk} = \mu + \alpha_i + n_{k(i)} + \beta_j + \alpha\beta_{ij} + e_{k(ij)}, \quad (2.16)$$

where $n_{k(i)}$ is the whole plot random error representing for each whole-plot, α_i denotes the main treatment, β_j , $\alpha\beta_{ij}$, and $e_{k(ij)}$ represents the sub-plot treatment, subplot interaction treatment, and split-plot level random error (SPE), respectively. This model is a simple, completely randomized design on the entire plot with the entire plot factor as treatment, where the entire plot's error term is nested within the entire plot treatment. Lohr (1995) proposed the Hasse Diagram as a tool for visualizing two ANOVA Tables. The Hasse diagram has the advantage of displaying the degree of freedom through the denominator test. For example, the denominator test for the whole plot factor A is the whole plot error (WPE).

As shown in Figure 2.10, Oehlert (2010) presented an example of two Hasse Diagrams, one generic and the other a split-plot with $an = 10$ whole plots, whole plot factor A with $a = 2$ levels, and subplot treatment B with $b = 3$ levels.

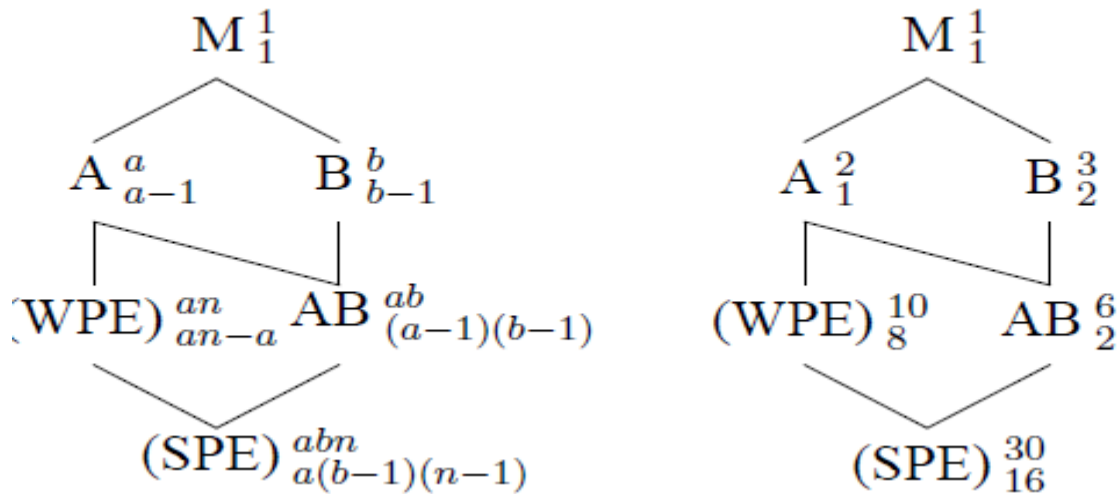


Figure 2.10: Two Hasse Diagrams where the one is generic and the other is a split plot with 10 whole plots

Figure 2.10 and 2.11 shows that the degree of freedom of the whole plot error (WPE), split-plot error (SPE), primary treatment (A), subplot treatment (B), and subplot interaction treatments (AB). As a result, the researcher can efficiently compute the ANOVA Table using this Hasse Diagram.

2.3.2 Analysis of Split-Plot Designs

The split-plot models are usually analyzed using the standard methods with linear mixed effect factorial models. The same result can also be obtained by employing the Heuristic technique as described by Oehlert (2010). The Heuristic technique is based on the idea that a split-plot design typically has two unit sizes and two randomizations. The data variable should be divided into two groups: variation between whole plots and variation within whole plots (between split plots). Using this model with two factors implies that they are whole plots with $N - 1 = abn - 1$ degrees of freedom between all the expected responses. The variation between whole plots can be obtained by treating the whole plots as treatment groups of b units each and performing a standard one-way ANOVA.

As a result, there are $an - 1$ and $an(b - 1)$ degrees of freedom between whole plots, as well as within whole plots and between split plots. This decomposition can be visualized using the heuristic technique and the schema on the partition of total variation. The variation between whole plots is typically made up of effects that completely affect the entire whole plots. If the entire plot is assumed to be blocked, the variation produces the following decomposition. Moreover, the variance in the split plot design is further partitioned as follows:

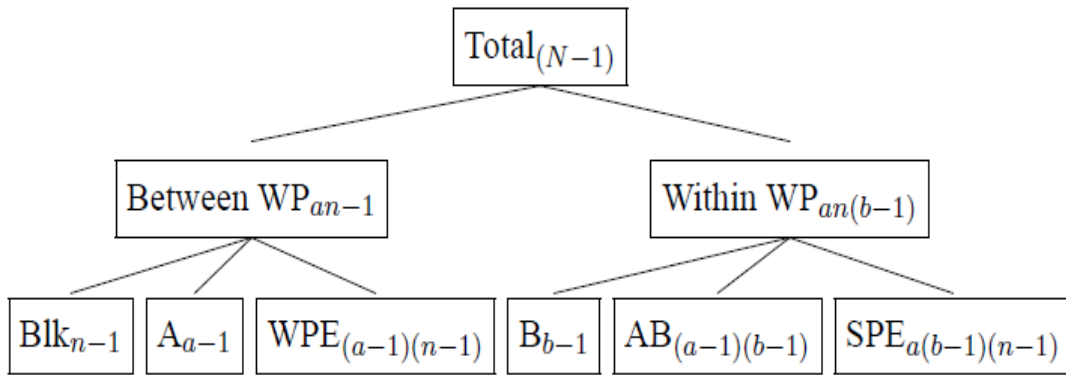


Figure 2.11: The total variation resulting from both and within Whole plot and sub plot

The degrees of freedom for split-plot treatments, subplot treatment interaction, and subplot error are listed respectively as $an(b-1)$, $b-1$, $(a-1)(b-1)$, and $a(b-1)(n-1)$. As a result, the overall decomposition that provides more insight into what is going on in the split-plot analysis. However, sums of squares and estimated treatments of effects are computed in the same way.

2.4 Robust Parameter Design for MPV with Noise Variable

Cho (2010) proposed a robust parameter design (RPD) for variables that are not easily controlled in SPD. In his study, he divided process variables into two categories in the experiment design: controllable variables that directly affect the response and variables that are not controllable and mainly affect the variability of predicted response variability. If the model contains both control factors and noise variables, the factor settings for easy-to-change factors (control factors) make the predicted response resilient subject to the variability transmitted from hard-to-change factors. Furthermore, Steiner and Hamada (1997) were the first to investigate, analyze, and improve the Mixture Process Variable design with difficult to adjust variables.

Cornell (2011) used a split-plot design to analyze data from MPV design experiments where the mixture blends in the process variables. The mixture experiment in his SPD involved making fish parties out of various ingredients from three different fish species, including Mullet, Croaker, and Sheepshead. The development of fish parties was subjected to a three-process variable treatment level mix, with each process variable having two levels. Deep frying time, cooking time, and cooking temperature were the process variables.

2.5 The Design Experimental Region (Design Space)

The experimental region includes the design area, design space, and the region of interest. The experiment area is normally coded with $\mathcal{E} = [-1, 1]^k$ and this is subject to all continuous experimental influences. Thus the experimental region \mathcal{E} is often sensible when there are no restrictions on the factor other than the lower and upper bounds and this is often referred to as a cuboidal region. The spherical experimental region is often a choice when a cuboidal region is not specified (Rozum and Myers, 1991).

$$\mathcal{E} = \left\{ X = (x_1, x_2, \dots, x_k) \mid \sum_{i=1}^k x_i^2 \leq R \in R \right\} \quad (2.17)$$

where for any radius R , the maximum of factors will take on the extreme levels -1 and 1 simultaneously. As a result, the best design is the one that fits the experimental area, making it even more important to choose the best design for a given problem.

2.6 Graphical Techniques for Evaluating Design Optimality

Both the Fraction of Design Space (FDS) plots and the Variance Dispersion graphs (VDG) visualize and show the output of one or more in terms of scaled variance prediction (SPV)

throughout the entire experimental area. Zahran *et al.* (2003) introduced FDS plots as an alternative to the VDGs as suggested by Giovannitti-Jensen and Myers (1989). These prediction-based criteria provide more information and detail than the average or maximum prediction variance that compares numerical efficiency between different experimental designs.

Thus, the FDS plots and VDGs have been the focus of many studies over the last decade, with Anderson Cook *et al.* (2009) offering a summary and discussion of this work. Meyer and Nachtsheim (1995), and Goos *et al.* (2016) addressed the I-optimal design criterion, aiming to eliminate average prediction variance. Rodriguez *et al.* (2010) addressed G-optimal architecture's generation and efficiency, which aims to reduce maximum prediction variance. Many of the authors were primarily concerned with minimizing and optimizing prediction variance. According to Goos and Jones (2011), it is normally appropriate to allow for lower prediction variance over most experimental areas of interest. Atkinson (2008) explored and outlined various optimality criteria for design selection, including the A, D, I, and G optimality criteria, among others. Different design optimality requirements are often present, consistent with the theory of optimal experimental design. According to Goos and Jones (2011), different experimental objectives and problems frequently necessitate different experiments based on optimal experimental design theory. The only way to do this is to apply the design optimality criterion to the experiment. When an experiment has multiple experimental objectives, as defined by Rady *et al.* (2009), designs that perform well in multiple criteria at the same time should be chosen and adopted.

2.6.1 Fraction of Design Space Plots Construction

Vining *et al.* (1993) used the FDS plot to show and explain one or more experimental designs' success in terms of precision of variance prediction. For the FDS map, we define $\varphi(s)$ as a function that considers a vector of factor settings and increases its size to the model terms it corresponds to (Goldfarb and Montgomery, 2006; Ozol *et al.*, 2005). When s is a vector of factor levels that corresponds to one of the experiment design's runs, the $\varphi'(s)$ usually corresponds to the row matrix S . Suppose $S_i = (s_{1i}, s_{2i}, s_{3i})$ are the factor settings for design, then $\varphi'(s_i) = (1, s_{1i}, s_{2i}, s_{3i}, s_{1i}s_{2i}, s_{1i}s_{3i}, s_{2i}s_{3i}, s_{1i}^2, s_{2i}^2, s_{3i}^2)$. According to Goos and Jones (2011), the variance of prediction expectation at the setting S is

$$Var(\hat{y}|s) = \sigma_e^2 [\varphi'(s_i)(S'S)^{-1} \varphi(s_i)]. \quad (2.18)$$

This $Var(\hat{y}|s)$ aid in calculating the expected response variance about the error variance σ_e^2 given as

$$v_{RVP} = \frac{Var(\hat{y}|s)}{\sigma_e^2}, \quad (2.19)$$

where v_{RVP} is the relative variance of prediction. However, when constructing an FDS plot for a given design, a random sample of a large number of points (approximately 10000) within the experimental area (\mathfrak{E}) is normally taken and the v_{RVP} value is computed for each design point. The FDS plot, on the other hand, is generated by plotting the ordered pairs $\left(\frac{i}{(N+1)}, v_i\right)$, where N is the sample's total number of points and v_i is the i^{th} sorted scaled expected variances in the sample.

The FDS plots are often non decreasing curve in nature and joins N points in \mathcal{E} when constructed. The vertical axis normally represents the scaled predicted variance (SPV) that ranges from minimum to maximum at the sampled point in \mathcal{E} whereas the horizontal axis scaled between 0 and 1 and each point on this axis corresponds to FDS or \mathcal{E} .

2.6.2 Algorithm Used to Generate FDS Plots

The algorithms for the implementation of FDS plots using a computer program have been developed in recent years (Cho, 2010). However, according to several authors, this aid in efficient computation of the plots. Goldfarb and Montgomery (2006), suggested the following procedure that aid in developing the algorithms. Usually, the FDS criterion is first defined as

$$FDS = \frac{1}{\phi} \int \dots \dots \dots \chi \cdot \int dx_n \dots dx_1, \quad (2.20)$$

where $\chi = \{(x_1 \dots \dots x_n): V(x) < v\}$, and V denotes any predetermined scaled predicted variance (SPV) value, n the number of component variables, and ϕ the total value of the experimental design region. Furthermore, the integrands of the above integrals are determined using the known elements. This enables the computation of FDS volume. Furthermore, a fine grid is used to estimate and calculate how many numbers of points, say n_χ that can satisfy the condition of the set χ . Then, this, n_χ can also be used to approximately estimate fraction dispersion space as

$$FDS \cong \frac{n_\chi}{N}, \quad (2.21)$$

where N is the total number of points sampled in the design space using the same grid in this case. However, an exact method for obtaining FDS for coded factors is given as

$$FDS = \frac{1}{\phi} \int \dots_E \dots \int I(x) dx_n \dots dx_1, \quad (2.22)$$

given

$$I(x) = \begin{cases} \frac{1}{2}(|\varpi| - \varpi) & \varpi \neq 0 \\ \varpi = 0 & \forall: \varpi = V(x) - v, \end{cases}$$

where E represent the total design space (experimental region). It can be shown that the set χ can be described in terms of ϖ as $\chi = \{(x_1 \dots x_n): \varpi < 0\}$. Thus, the negative value of ϖ becomes the main concern by the researcher as described by Goldfarb and Montgomery (2006). However, as some scholars have pointed out, $I(x)$ is simply an indicator function with two values: zero or one (Ozol *et al.*, 2005; Vining *et al.*, 1993). Disc filtering out all variances greater than the cut-off points allows for convergence across the entire spectrum of variables. For all rotatable designs in a spherical area, this criterion was further simplified as

$$FDS = n \int_0^1 I(x) r^{n-1} dr, \quad (2.23)$$

where r is the experimental region's radius scaled to the unit sphere. However, Jones and Sall (2011) developed and implemented a FORTRAN code related to FDS available from the researchers in some statistical methods such as the JMP software division of SAS to compute the FDS graphs.

2.6.3 Variance Dispersion Graphs (VDGs)

The VDG is a plot that measures or compares various designs based on maximum and minimum values over spheres of radius r from the design center and the spherical average of SPV against the radius in the experimental region of interest. Giovannitti-Jensen and

Myers (1989) were the first to present this variance-based graph approach to studying a design's prediction capability. They defined the spherical average variance as

$$V^r = \vartheta \int_{U_r} V(x) dx, \quad (2.24)$$

where

$$U_r = \left\{ \mathcal{X} : \sum_{i=1}^n x_i^2 = r^2 \right\}, \quad \vartheta^{-1} = \int_{U_r} dx. \quad (2.25)$$

In the case of cuboidal region they are measured over spheres or parts of spheres on or inside the cube. Zahran *et al.* (2003) defined this typically changes the relative importance that can give to the VDGs' interpretation for different numbers of factors n . The degree of rotatability of the SPV at any given radius sphere often demonstrates by comparing the maximum and minimum SPV over a range of values. However, the plot also shows the G-efficiency on a horizontal line at p and $2p$, which are 100 % and 50 %, respectively, as seen in Goldfarb *et al.* (2004b) and Cho (2010). The FORTRAN program builds the VDG for any design.

Cho (2010) demonstrated in the experiment that VDGs are useful summaries for comparing competing designs on fixed design space. Still, they do not include all useful information about a design's prediction capability, so they suggested FDS techniques overcome some of VDG's shortcomings. This is because FDS method quantifies the FDS where the SPV is less than or equal to some pre-specified value, giving the author more precise details. Furthermore, based on the ranges and proportions of potential SPV values, the FDS plots provide more detail about the SPV distribution in the experimental area. The FDS criteria,

on the other hand, complements the use of VDGs by providing the author with additional insight into a design's prediction capability.

2.6.4 Evaluation of FDS plot for MPV designs

The investigation of experiments using mixture process variable designs (MPVD) are well-known in various fields, including agriculture, food, consumer goods, industries, and chemical and pharmaceutical research (Anderson Cook *et al.*, 2004). Cho (2010) points out that the number of blends and constraints always determines the experimental region's form. As a result, the researcher for MPVD is also interested in other variables that can modify one another independently and the mixture blends. Thus, the number of variables as well as the number of mixture ingredients subject to the restrictions determines the form of the experimental regions in these combined designs (Cornell, 2011).

Cornell (2011), Czitrom (1988, 1992), and Kowalski *et al.* (2000) all give an alternative method for MPV experiments. However, Liang *et al.* (2006) proposed techniques for assessing various designs using SPV in MPV. They argued that analyzing scaled prediction variance over single number summaries, as used by many alphabetic optimality criteria, has advantages.

Cornell (1988) was the first to use this technique to map the entire distribution of the SPV for a given shrinkage degree. VDG was introduced by Vining *et al.* (1993) for mixture designs to plot SPV and the Cox directions. They define Cox directions as a ray moving through the unconstrained simplex vertices and the restricted experimental region's centroid.

Goldfarb et al. (2003) and Hassan *et al.* (2020) established three-dimensional VDGs for MPV experiments. They plot the shrinkage in the mixture space, process variable spaces, and averages or maximum SPV along the X -axis, Y -axis, and Z -axis, respectively. Shapes of the prediction variance surface defined by Liang *et al.* (2006) display the prediction variance's behavior in the MPV shrinkage plane. These VDGs normally provide an experimenter with a visual representation of a design's prediction variance properties across the combined MPV space. Zahran *et al.* (2003) introduced FDS plots to counter VDG's drawbacks since SPV computes over the entire design space. Regardless of the proportion of the design space depicted, the VDG usually gives each radius or shrinkage value equal weight on the plot. Furthermore, the FDS plot makes it simple to compare different designs with a single plot. As a result, the distinction between VDG and FDS is obvious. Cho *et al.* (2009) developed and demonstrated these two techniques for spherical and cuboidal designs, but they did not discuss non-regular area designs as defined by Goldfarb et al. (2004b). The FDS and VDG plots demonstrate that these two methods reveal the strengths and limitations of competing designs when used together. Mclean and Anderson (1966), Piepel *et al.* (1993) and Kowalski *et al.* (2000) introduced new techniques to compute VDGs.

2.6.5 The Average Relative Variance

The Averaging relative variances (\mathbf{v}_{RVP}) obtained during the construction of an FDS plots normally leads to the average \mathbf{v}_{RVP} for a given design. The average \mathbf{v}_{RVP} can also be obtained through integrating relative variance prediction over Ξ and dividing the volume of the region as shown

$$\text{average} \mathbf{v}_{RVP} = \left\{ \int_{\Xi} d\mathbf{s} \right\}^{-1} \left\{ \int_{\Xi} [\varphi'(\mathbf{s}_i)(S'S)^{-1}\varphi(\mathbf{s}_i)] d\mathbf{s} \right\}. \quad (2.26)$$

The volume of the experimental region $\int_{\Xi} d\mathbf{s} = \mathbf{2}^k$ if $\Xi \in [-1, 1]^k$ where k denotes the number quantitative experimental factors. Since

$$\begin{aligned} \varphi'(\mathbf{s}_i)(S'S)^{-1}\varphi(\mathbf{s}_i) &= \text{trace} [\varphi'(\mathbf{s}_i)(S'S)^{-1}\varphi(\mathbf{s}_i)], \\ &= \text{trace} [(S'S)^{-1}\varphi(\mathbf{s}_i)\varphi'(\mathbf{s}_i)], \end{aligned}$$

This implies that

$$\begin{aligned} \int_{\Xi} [\varphi'(\mathbf{s}_i)(S'S)^{-1}\varphi(\mathbf{s}_i)] d\mathbf{s} &= \int_{\Xi} \text{trace} [\varphi'(\mathbf{s}_i)(S'S)^{-1}\varphi(\mathbf{s}_i)] d\mathbf{s}, \\ &= \int_{\Xi} \text{trace} [(S'S)^{-1}\varphi(\mathbf{s}_i)\varphi'(\mathbf{s}_i)] d\mathbf{s}, \\ &= \text{trace} \left[(S'S)^{-1} \int_{\Xi} \varphi(\mathbf{s}_i)\varphi'(\mathbf{s}_i) d\mathbf{s} \right], \end{aligned}$$

where $(S'S)^{-1}$ is treated as constant for this integration. This is because the factor level settings of a design experiment are fixed. Therefore,

$$\text{average} \mathbf{v}_{RVP} = \{\mathbf{2}^k\}^{-1} \text{trace} \left[(S'S)^{-1} \int_{\Xi} \varphi(\mathbf{s}_i)\varphi'(\mathbf{s}_i) d\mathbf{s} \right], \quad (2.27)$$

According to Goos and Jones (2011), this expression $\int_{\Xi} \varphi(\mathbf{s}_i)\varphi'(\mathbf{s}_i)d\mathbf{s}$ easily computed for all continuous designs when $\Xi \in [-1, 1]^k$. The result generated from $\int_{\Xi} \varphi(\mathbf{s}_i)\varphi'(\mathbf{s}_i)d\mathbf{s}$ is equivalent to moment matrix (information matrix) denoted with

$$\text{average}\mathbf{v}_{RVP} = \{\mathbf{2}^k\}^{-1} \text{trace} [(S'S)^{-1}M]. \quad (2.28)$$

Letter \mathbf{M} . Since $\int_{\Xi} \varphi(\mathbf{s}_i)\varphi'(\mathbf{s}_i)d\mathbf{s} = \mathbf{M}$ and $\int_{\Xi} d\mathbf{s} = \mathbf{2}^k$, this implies that

Here we consider a full quadratic model with two continuous factors s_1 and s_2 such that $s_1, s_2 \in [-1, 1]$ and $\varphi'(s_1, s_2) = (1, s_1, s_2, s_1s_2, s_1^2, s_2^2)$. Therefore, the information matrix can be obtained as

$$\begin{aligned} \mathbf{M} &= \int_{\Xi \in [-1, 1]^2} \varphi(\mathbf{s}_i)\varphi'(\mathbf{s}_i)d\mathbf{s} & (2.29) \\ &= \int_{\Xi \in [-1, 1]^2} \varphi(s_1, s_2)\varphi'(s_1, s_2)ds_1 ds_2, \\ &= \int_{-1}^1 \int_{-1}^1 \varphi(s_1, s_2)\varphi'(s_1, s_2)ds_1 ds_2, \\ &= \int_{-1}^1 \int_{-1}^1 \begin{pmatrix} 1 & s_1 & s_2 & s_1s_2 & s_1^2 & s_2^2 \\ s_1 & s_1^2 & s_1s_2 & s_1^2s_2 & s_1^3 & s_1s_2^2 \\ s_2 & s_1s_2 & s_2^2 & s_1s_2^2 & s_1^2s_2 & s_2^3 \\ s_1s_2 & s_1^2s_2 & s_1s_2^2 & s_1^2s_2^2 & s_1^3s_2 & s_1s_2^3 \\ s_1^2 & s_1^3 & s_1^2s_2 & s_1^3s_2 & s_1^4 & s_1^2s_2^2 \\ s_2^2 & s_1s_2^2 & s_2^3 & s_1s_2^3 & s_1^2s_2^2 & s_2^4 \end{pmatrix} ds_1 ds_2 \end{aligned}$$

$$= 2^2 \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{9} & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{5} & \frac{1}{9} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{9} & \frac{1}{5} \end{pmatrix}$$

In general, the information matrix for a second order model can be obtained as

$$\mathbf{M} = 2^k \begin{pmatrix} 1 & \mathbf{0}'_k & \mathbf{0}'_{k(k-1)/2} & \frac{1}{3} \mathbf{1}'_k \\ \mathbf{0}_k & \frac{1}{3} \mathbf{I}_k & \mathbf{0}_{k \times k(k-1)/2} & \mathbf{0}_{k \times k} \\ \mathbf{0}_{k(k-1)/2} & \mathbf{0}_{k(k-1)/2 \times k} & \frac{1}{9} \mathbf{I}_{k(k-1)/2} & \mathbf{0}_{k(k-1)/2 \times k} \\ \frac{1}{3} \mathbf{1}_k & \mathbf{0}_{k \times k} & \mathbf{0}_{k \times k(k-1)/2} & \frac{1}{3} \mathbf{I}_k + \frac{1}{9} (\mathbf{J}_k - \mathbf{I}_k) \end{pmatrix}, \quad (2.30)$$

and where \mathbf{I}_k and $\mathbf{I}_{2^{-1}k(k-1)}$ denote identity matrices of dimension k and $2^{-1}k(k-1)$ respectively. $\mathbf{0}_{k \times k}$ and $\mathbf{0}_{k(k-1)/2 \times k}$ denote matrices of zero with $k \times k$ and $k(k-1)/2 \times k$ respectively while $\mathbf{0}_k$ and $\mathbf{0}_{k(k-1)/2}$ represent a column vector containing zeros of k and $k(k-1)/2$ respectively. \mathbf{J}_k and $\mathbf{1}_k$ denotes $k \times k$ matrix of ones and a column vector containing k ones respectively.

2.7 General Optimal Design Criteria for Comparing and Examining Designs

The design optimality criterion, such as A-optimality, I-optimality, E-optimality, V-optimality, G-optimality, Q-optimality, and D-optimality, are typically used to evaluate, and compare performance of the selected designs. When the hypothesized model is of

degree j and $1 \leq j \leq k \leq 8$, Wong (1994) investigated the optimal criterion A, D, E, and G efficiencies for polynomial regression models of degree k . Most optimal designs are model-based and consider different model assumptions since, in practice, a true model is always uncertain (Iwundu, 2017). The robustness of properties of the optimal designs are normally evaluated under different optimality criteria. Wangui (2019) posited that the optimality with respect to any one criterion usually represents an estimate to some vague notion of ‘goodness’ of the model. This sometimes results in a design that meets many optimality requirements without being overly emphasized. Therefore, it is always vital to examine different optimality criteria based on their model assumptions. The efficiencies of various types of optimal designs are typically compared numerically under the assumption that the true model $\varphi_j(x)$ is j^{th} degree polynomial model given $1 \leq j \leq 8$. In the absence of loss in generality, the symbol Ω is used in this paper to denote the design space that lies between -1 and 1. The use of optimal design for $\varphi_k(x)$ is a good idea for the assumed model $\varphi_j(x)$, $k > i$ as described by Kussmaul (1969), Kendall and Stuart (1968), Iwundu (2017), among others. This is because it always enables the researchers at least to perform a lack of fit test to various models being applied in order to determine the best model.

According to the standard optimal design theory by Fedorver, (1972) $M(\xi)$ is also known as information matrix and is defined as

$$M_j(\xi) = \int_{\Omega} \varphi_j(x)\varphi_j'(x)\xi dx, \quad (2.31)$$

In addition, this information matrix also contains the main practical objective of an experiment. Most optimality criteria for various designs have been discussed by various

researchers. Here we look at a few of them on how they contribute in selecting the best design and for instance we start with D- optimal criterion.

2.7.1 D-optimality Criterion

The D-optimality criterion is the determinant of the matrix $M_j(\xi_j)$ for all the set (Ξ) for all continuous designs on Ω .

$$M_j(\xi) = \text{Max}_{\xi} \epsilon \equiv |M_j(\xi)| \quad (2.32)$$

Where $|M_j(\xi)| = \det(M_j(\xi))$ denotes the informational matrix determinant (Wong, 1994). The moment matrix is another name for the information matrix denoted as $M = \frac{X'X}{N}$ where the design matrix is represented by X, and its transpose is represented by X'. This optimality criterion normally focuses on good model parameter estimation. Furthermore, a design is D-optimal if it reduces the overall variance of the model parameter estimates. The criterion function, on the other hand, is defined as:

$$\phi(M_j(\xi)) = \text{Max}_{\xi} \{M_j(\xi)\} = \text{Min}_{\xi} \{ \det \det(M^{-1}(\xi)) \} \quad (2.33)$$

The D- efficiency of any design can also be obtained for the purpose of numerical comparison of various designs. This D- efficiency is given as

$$D(M_j(\xi)) = [\text{Min}_{\xi} \{ \det \det(M^{-1}(\xi)) \}]^{p-1} \quad (2.34)$$

The computation of relative D-efficiency is quite useful when it comes to comparing more than two designs at any given time and only one design needs to be selected (Iwundu, 2017). In addition, the relative D-efficiency (RD) plays a big role in determining best design among other designs with missing observation. Therefore, this leads to computation

of loss in relative D-efficiency due to incomplete observation in any given design. The missing observation in a design experiment often drastically results in relative of D-efficiency as described by Rady *et al* (2009). For the purpose of computing RD, we first start by letting a design with complete observation as $M(\xi_N)$ and the one with missing observations as $M(\xi_{N-m})$ where ξ_N indicates information matrix with full observation while ξ_{N-m} with partial information. The D-efficiency is computed in both cases as

$$D(M(\xi_N)) = [\text{Min}_\xi \{ \det \det(M^{-1}(\xi_N)) \}]^{P-1} \quad (2.35)$$

$$D(M(\xi_{N-m})) = [\text{Min}_\xi \{ \det \det(M^{-1}(\xi_{N-m})) \}]^{P-1} \quad (2.36)$$

In (2.35) and (2.36) represent D-efficiency with complete and incomplete observation respectively. The RD is obtained by using (2.35) and (2.36) as follows

$$RD = \left\{ \frac{[\text{Min}_\xi \{ \det \det(M^{-1}(\xi_{N-m})) \}]^{P-1}}{[\text{Min}_\xi \{ \det \det(M^{-1}(\xi_N)) \}]^{P-1}} \right\} = \frac{D(M(\xi_{N-m}))}{D(M(\xi_N))} \quad (2.37)$$

The RD in (2.37) is used to compare designs, and the better design has the highest D-efficiency value. The RD of a designs lies between 0 and 1 such that $0 \leq RD \leq 1$. Therefore, the relative loss in D-efficiency in the case of missing observation is given as

$$RD_{Loss} = 1 - \frac{D(M(\xi_{N-m}))}{D(M(\xi_N))}. \quad (2.38)$$

2.7.2 G-optimality Criterion

The G-optimality criterion is a criterion that researchers used to reduce the overall variance of the approximated response surface across all variables. The quantity of information matrix becomes G-optimal design when $Max (d_j(X, \xi) X \in \Omega) = h_j'(x) M_j^{-1}(\xi) h_j$ over Ξ is minimized. Wong (1994) and Goos et al. (2016) argued that this design is more useful and important when the main goal is to estimate the entire response using the homoscedasticity assumption. As defined by the criterion function, the G-optimal design is

$$\phi(M(\xi)) = Min\{Max V(x) x \in R\}, \quad (2.39)$$

where the scaled prediction variance is denoted by $V(x)$ in this case. Prior to running the test and taking measurements, the Scaled Prediction Variance (SPV) is used to analyze a planned experiment. This is because it describes and elucidates the error involved with making a prediction using a regression model. Moreover, this optimality criterion considers a design where maximum SPV in the region (\mathcal{E}) of interest is not too large and hence it maximizes the maximum SPV. When there are P parameters in the model and maximum SPV ($V(x)$), the G-efficiency can be calculated as:-

$$G_{eff} = P[V(x)_{Max(\xi)}]^{-1}, \quad (2.40)$$

Furthermore, the G-efficiency can also be computed in the case of missing observation. If we let the expected design be ξ_N and observed design be ξ_{N-m} , then relative G-efficiency (RG) can be obtained for P parameter model. This is done by first determining G-efficiency for a design with partial observation as:-

$$G_{eff}(\xi_{N-m}) = P[V(x)_{Max(\xi_{N-m})}]^{-1}. \quad (2.41)$$

This $G_{eff}(\xi_{N-m})$ enables to compute the RG which is defined as:-

$$RG = \frac{G_{eff}(\xi_{N-m})}{G_{eff}(\xi_N)} = \frac{V(x)_{Max(\xi_N)}}{V(x)_{Max(\xi_{N-m})}}. \quad (2.42)$$

Moreover, this RG aid in comparing design and the best design is known with the largest G-efficiency such that $0 \leq RG \leq 1$ but when $RG \leq 0$ the design ξ_{N-m} with missing observation is better than the design ξ_N with complete observation. Besides calculating RG , the relative loss in G_{eff} due to missing observation is obtained as:-

$$RG_{loss} = 1 - \frac{V(x)_{Max(\xi_N)}}{V(x)_{Max(\xi_{N-m})}} \quad (2.43)$$

2.7.3 A-optimality Design Criterion

Researchers use the A-optimality criterion in planned experiments to reduce the variance of parameter estimates while ignoring model parameter covariance (Cornell, 2011; Iwundu, 2017; Wangui, 2019). Furthermore, the number of the variances of the model coefficient is minimized in this optimality criterion. This criterion for optimality is as follows:

$$\phi(M(\xi)) = Min\{tr[M^{-1}(\xi)]Over \equiv\}, \quad (2.44)$$

where tr represent the trace and \equiv indicates all the set for all continuous designs on Ω . The

A-efficiency in general is defined as

$$A_{eff}(\xi) = \frac{tr[M^{-1}(\xi^*)]}{tr[M^{-1}(\xi)]}, \quad (2.45)$$

where ξ^* in this case indicates A-optimal. However, the A-efficiency for a design with incomplete and complete observation is computed respectively as

$$A_{eff}(\xi_{N-m}) = \frac{tr[M^{-1}(\xi^*)]}{tr[M^{-1}(\xi_{N-m})]} \quad (2.46)$$

and

$$A_{eff}(\xi_N) = \frac{tr[M^{-1}(\xi^*)]}{tr[M^{-1}(\xi_N)]} \quad (2.47)$$

Therefore, the relative A-efficiency is given as

$$RA_{eff}(\xi) = \left(\frac{\left\{ \frac{trace[M^{-1}(\xi^*)]}{trace[M^{-1}(\xi_{N-m})]} \right\}}{\left\{ \frac{trace[M^{-1}(\xi^*)]}{trace[M^{-1}(\xi_N)]} \right\}} \right) = \frac{tr[M^{-1}(\xi_N)]}{tr[M^{-1}(\xi_{N-m})]} \quad (2.48)$$

This $RA_{eff}(\xi)$ aid in comparing the design. The best design is known with the largest $A_{eff}(\xi)$ value where $RA_{eff}(\xi) \in [0, 1]$. If $RA_{eff}(\xi) < 0$ shows that the design ξ_{N-m} is better than ξ_N .

2.7.4 I-optimal Criterion

I -optimal is an optimality criterion that minimizes the average v_{RVP} . Some authors capitalize the I in I -optimal to emphasize that the goal of an I -optimal design is to reduce integrated variance (Goos *et al.*, 2016; Rady *et al.*, 2009; Njoroge *et al.*, 2017). Over centuries ago, G- optimality criterion has often been used since it minimizes the overall prediction variance over the experimental area as a prediction-based criterion for selecting experimental design. Recent research has shown that in more than 90% of experimental areas, reducing the overall prediction variance occurs as a result of increasing the

prediction variance as described by Iwundu. (2017), and Sitinjak and Syafitri (2019). As a result, as Goos and Bradley (2011) point out, most writers prefer I-optimal designs to G-optimal designs. This is evident from the contour plots of the objective function for three criteria (D, I, G) using a very simple model function that was done by Crary *et al.* (1992) shown in Figure 2.12. This simple model function was $Y = \beta_0 + \beta_1 X$ and $n = 2$ experiments where the optimal design for three criteria placed one experiment at -1 and the other at +1 as it is supposed to be. Their findings indicate that the contours for G - optimality have discontinuous slope because of the minimax nature of the criterion.

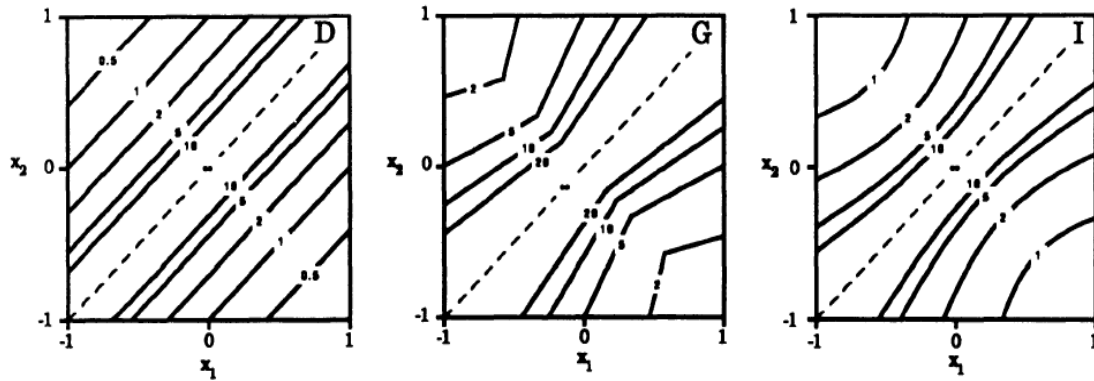


Figure 2.12: The contours plots presented for objective function space for D-, G- and I-optimality for a straight- line regression with $n = 2$ (Crary *et al.*, 1992).

The appropriate objective function for the I-optimality criteria is defined as,

$$\begin{aligned} \text{Min}_Q \int_{x \in \mathcal{E}} E \{ [\hat{Y}(x) - Y(x)]^2 \} d\mu(x) \\ = \text{Min} \int_{x \in \mathcal{E}} [\varphi'(x)(x'x)^{-1}\varphi(x)] d\mu(x), \end{aligned} \quad (2.49)$$

$$\text{Min trace } A(x'x)^{-1} \text{ and } A = \int_{x \in \mathcal{E}} \varphi(x)\varphi'(x)d\mu(x). \quad (2.50)$$

where A is the matrix that contains all of the model's dependencies. The minimization case of the integral over the set of points $s \in \mathcal{E}$ for the experimental design Q is denoted by $\text{Min } Q$. However, the estimated response within different experiment regions can be weighted through the differential Equation $d\mu(x)$ as shown above.

The process for computing I -efficiency for designs is always important since it aids in selecting the appropriate design. The I -efficiency for a design is obtained by the average ν_{RVP} . Therefore, a design with complete observation (ξ_N) , then the I-efficiency is computed as

$$I_{eff}(\xi_N) = \{2^k\}^{-1} \text{trace} [(S'S)^{-1}M(\xi_N)]. \quad (2.51)$$

However, for the case of a design with missing observation, the I-efficiency can be obtained as

$$I_{eff}(\xi_{N-m}) = \{2^k\}^{-1} \text{trace} [(S'S)^{-1}M(\xi_{N-m})]. \quad (2.52)$$

Furthermore, the relative I-efficiency and relative of I-efficiency loss after missing observation is obtained as follows

$$\begin{aligned} RI_{eff}(\xi_{N-m}|\xi_N) &= \frac{\{2^k\}^{-1} \text{trace} [(S'S)^{-1}M(\xi_{N-m})]}{\{2^k\}^{-1} \text{trace} [(S'S)^{-1}M(\xi_N)]}, \\ &= \frac{\text{trace} [(S'S)^{-1}M(\xi_{N-m})]}{\text{trace} [(S'S)^{-1}M(\xi_N)]}, \\ &= \frac{I_{eff}(\xi_{N-m})}{I_{eff}(\xi_N)}. \end{aligned} \quad (2.53)$$

Thus, the relative of I -efficiency loss is given as

$$RI_{loss}(\xi_{N-m}|\xi_N) = 1 - \frac{\text{trace}[(S'S)^{-1}M(\xi_{N-m})]}{\text{trace}[(S'S)^{-1}M(\xi_N)]}, \quad (2.54)$$

and where $0 \leq RI_{eff}(\xi_{N-m}|\xi_N) \leq 1$. If $RI_{eff}(\xi_{N-m}|\xi_N) < 0$. This implies that design ξ_{N-m} is better than this design ξ_N . we also note that if they are two different designs for instance say Q_1 and Q_2 with each having relative I -efficiency RI_{Q_1} and RI_{Q_2} respectively, then relative I -efficiency of Q_1 versus Q_2 is given as $\frac{RI_{Q_2}}{RI_{Q_1}}$. Hence, if $\frac{RI_{Q_2}}{RI_{Q_1}} > 1$ implies that design Q_1 is better than design Q_2 .

2.8 D-Optimal Designs for Construction of Mixture Designs

In this section, we looked at the involvement of exact D- optimal criteria in construction of mixture designs based on pre-existing results on D- optimality. When the design space is the $q - 1$ dimensional simplex, Goos *et al.* (2016) points out that continuous D- optimal parameters for the models (2.3) to (2.4) are known. This criterion generally has two main characteristics. First, in D-optimal designs, the weight of each candidate point is equal to the inverse of the model parameters $\left(\frac{1}{P}\right)$. Secondly, they have minimum support designs, where d the number of distinct candidate is points and P is the number of model parameters. The same $d = P$ design points are used in both continuous and exact D-optimal design. When the budgeted number of runs in a mixture blend experiment, n , is an integer multiple of the D-optimal continuous configuration, then $n/d = n/P$ where the runs are performed at each of the candidate points.

However, if n is not an integer multiple of $d = P$, the situation can be addressed by having as many equireplicated constant $d = P$ design points as possible, as Cornell (2011) advocate. Some authors say that it doesn't matter the design points are repeated the most

since the D-optimality criterion is their only concern (Goos *et al.*, 2016). The D- optimality of the $(q, 1)$ and $(q, 2)$ SLDs for model (1) and (2), respectively, was defined by Kiefer (1961). D-optimal designs, according to Goos *et al.* (2016), include q pure blends, $(q, 2)$ mixtures involving 0.2764 percent of one mixture blend and 0.7236 percent of another mixture blend, with exact proportions given by $\frac{(1+\frac{1}{\sqrt{3}})}{2}$, and $(q, 3)$ ternary mixtures. When it comes to constructing D- optimal designs for complete cubic models, the $\{q, 3\}$ SLDs by 0.2764 and 0.7236 can be replaced by the proportions $\frac{1}{3}$ and $\frac{2}{3}$, respectively, based on Mikaeilli (1993) and Cornell (2011). Mikaeili (1993) proved the complete cubic models with all derivations in a general way. In contrast to other optimality designs, Goos *et al.* (2016) found that D- optimal designs perform remarkably well in terms of the I-optimality criterion.

2.9 I-Optimal Criteria for Construction of Mixture Designs

Research shows that only a small number of theoretical findings on the I-optimal design of mixture experiments have been reported. As Goos *et al.* (2016) point out, all of the outcomes require continuous designs. Some researchers used the terms V- optimal, I- optimal, and all variance design interchangeably to refer to I- optimal designs, citing Sinha *et al.* (2014)'s Theorem 12.1.1, which states that the continuous I- optimal design for a first order model in q mixture blends has a weight of $1/q$ at each stage of the $(q, 1)$ SLD. Furthermore, the q pure Mixture blends are the best candidate points, and each of them should be used equally.

However, candidate points in case $q \geq 3$ a second order model are analytically estimated by Kowalski *et al.*, (2000) based on the assumption that the candidate points are those of

the $(q, 2)$ SLD. As a result, as many authors have argued, Laake's weights are I-optimal if the $(q, 2)$ SLD is used. More specifically, Goos *et al.* (2016) published the analytical expression Goos and Safifri (2014) obtained for the I- optimal weights. The overview of numerical values as obtained for a second order degree model employing SLD techniques by Goos *et al.* (2016) for I _optimal weights for values q that ranges from three to six are given in Table 2.2. This value aids in constructing an Ideal mixture design.

Table 2.2: I-optimal weights for second order mixture model (Goose *et al.*, 2016)

i	$q = 3$			$q = 4$			$q = 5$			$q = 6$		
	k_i	Φ_i	$k_i\Phi_i$	k_i	Φ_i	$k_i\Phi_i$	k_i	Φ_i	$k_i\Phi_i$	k_i	Φ_i	$k_i\Phi_i$
1	3	0.1007	0.3022	4	0.0560	0.2240	5	0.0400	0.2000	6	0.0328	0.1968
2	3	0.2326	0.6978	6	0.1293	0.7760	10	0.0800	0.8000	15	0.0536	0.8032
d	6			10			15			21		

The weight Φ_1 denotes the number of runs that must be completed with each other pure mixture blend, while the weight Φ_2 denotes the number of runs that must be completed with each binary mixture. The number of pure blends (k_1) and (k_2)for binary mixture blends are also mentioned in Table 2.2, as well as the concinnity of experimental runs involving pure mixture blends ($k_1\Phi_1$) and ($k_2\Phi_2$)for binary mixture blends. The number of distinct candidate points is given by $d = (k_1 + k_2 + k_3)$ on the last line of Table 2.2. Furthermore, as Goos *et al.* (2016) points out, each pure blend in Laake's design has a weight of less than $1/d$, while each binary mixture has a weight of more than $1/d$, in comparison to the continuous D- optimal designs for the second order model.

As Goos *et al.* (2016) points out, Laake's proposal for a second order model never considered the case of two ingredients ($q = 2$). Liu and Neudecker (1995) analytically derived continuous I-optimal designs by considering the case of two ingredients ($q = 2$). The I-optimal weights for two pure mixture blends and a binary mixture were $\Phi_2 = 0.3$ and $\Phi_2 = 0.4$, respectively, as a result of their analytical expression.

Many scholars, however, believe that the designs advocated by Cho (2010) and Goos *et al.* (2016) are much superior to Lambrakis' (1968a).

2.10 Comparison of D- And I-optimal Designs

Figure 2.13 shows that I optimal criteria outperforms D-optimal design, based on findings from Laake (1975) and Goos *et al.* (2016) after the two optimality designs are contrasted respectively.

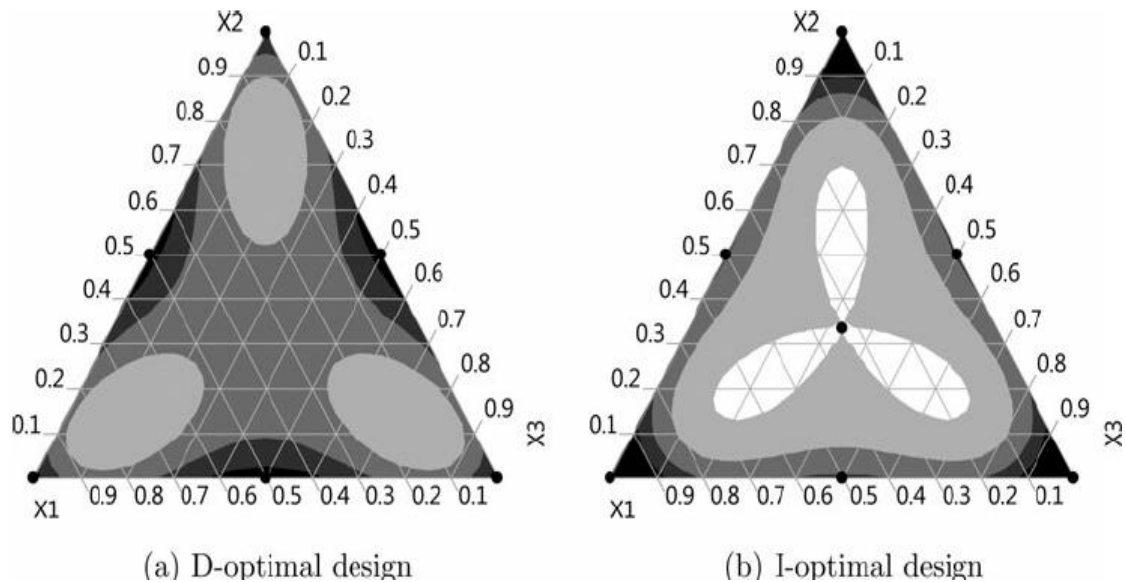


Figure 2.13: Comparison of the SPV created by the two optimal designs (Goos *et al.*, 2016)

The prediction variances provided by the two designs are compared in this Figure 2.13 over the entire experimental area (design space). Figure 2.13's white, dark gray, and black areas

correspond to an SPV of less than 2.5 to 4.5 to 5.5, and above 5.5, respectively. However, as demonstrated by Goos *et al.* (2016), I-optimal designs result in lower prediction variance over the majority of the design space as compared to D-optimal designs.

Furthermore, as shown in Figure 2.13, D- optimal designs collect information in the center of the experimental design, while I-optimal designs do not. However, according to the findings of Iwundu (2017), D- efficiency of I- optimal designs relative to D- optimal designs is approximately 89.02 percent, while I- efficiency of optimal designs criteria relative to I- optimal designs is symmetric and unique. In addition, Njoroge *et al.* (2017) applied these two optimal criteria to construct two mixture process variable designs within a split plot design in which they found out that the initial model constructed using D-optimal design outperforms I-optimal designs. Therefore in study we employed D-optimal to create mixture dataset in the presences of process variable.

2.11 Central Composite Design

To determine the optimum conditions for the critical factors, the central composite design (CCD) is used. This is because, as many authors have pointed out, CCD includes an embedded factorial or fractional factorial architecture with center points that is supplemented with a group of star points that allows estimation of curvature (Cho *et al.*, 2009; Goos and Donev, 2007; Njoroge *et al.*, 2017). Cornell (1988) points out that if the distance from the experimental region's (design space) center to a factorial point is one unit for each factor, then the distance from the experimental region's center to the factorial point is $|\alpha| > 1$. However, certain properties desired for the design, as well as factors involved in the experiment, decide the precise value a .

Salehi *et al.* (2012) found that a CCD includes twice as many star points as there are component factors in the design. As shown in Figure 2.14, the new extreme values (low and high) for each factor in the design are frequently represented by star points.

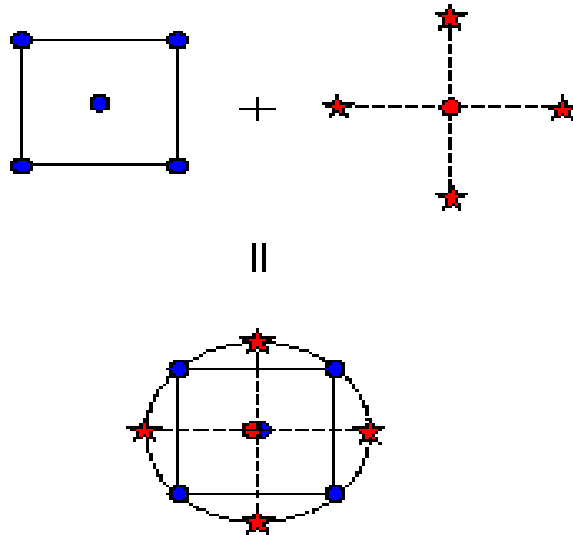


Figure 2.14: Diagram of CCD generation for two factor components (Salehi *et al.*, 2012).

This value of α is obtained as

$$\alpha = (2^k)^{\frac{1}{4}} \quad (2.55)$$

where 2^k is the number of factorial runs and k represent the number of factors. The total of experimental in CCD is given as

$$N = 2^k + 2k + n_c, \quad (2.56)$$

where n_c is the number of central points as shown in Figure 18. CCD is commonly used to fit the general second-order model in the form

$$y = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^k \sum_{j=1}^k \beta_{ij} X_i X_j + \sum_{i=1}^k \beta_{ii} X_i^2 + \varepsilon, \quad (2.57)$$

Where y represent the response variable, $\beta_0, \beta_i, \beta_{ii}$, and β_{ij} indicates the regression coefficient of variable intercept, linear, quadratic and interaction terms, respectively.

Exploration of CCD within SPD

Assume that the model used to suit the SPD is a second order model. As a result, the model for the n whole plot factor is as follows:

$$g_w(Z_i)' \beta_{wp} = \beta_0 + \sum_{i=1}^n \beta_i Z_i + \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} Z_i Z_j + \sum_{i=1}^n \beta_{ii} Z_i^2 \quad (2.58)$$

and the model for the k split-plot treatments is

$$g_s(Z_i, X_j)' \theta_{sp} = \theta_0 + \sum_{i=1}^k \theta_i X_i + \sum_{i=1}^k \sum_{j=1}^k \theta_{ij} X_i X_j + \sum_{i=1}^k \theta_{ii} X_i^2 \quad (2.59)$$

$$+ \sum_{i=1}^k \sum_{j=1}^n \gamma_{ij} X_i Z_j$$

where $g_w(Z_i)' \beta_{wp}$ and $g_s(Z_i, X_j)' \theta_{sp}$ represents the expected response for the main treatment and sub plot treatment, respectively. θ_0 and β_0 , θ_i and β_i , θ_{ii} and β_{ii} , θ_{ij} , β_{ij} , and γ_{ij} are variables that represent the regression coefficients for the intercept, linear, quadratic, and interaction terms, respectively. When the model inside an SPD is presumed to be a second order model, the CCD, as defined by Goldfarb *et al.* (2003) is a design that will help in the model fit and estimation of the mean response along the curvature.

The normal coding in a CCD is ± 1 for factorial portion. In addition, permutation of $(\pm\alpha, 0, 0, \dots, 0)$ for the axial points and $(0, 0, 0, \dots, 0)$ for the center points respectively. However, if the real values of the independent variable (whole plot factors, w_i) for instance the PH of Soil can be coded to Z_i according to Equation (2.60) by setting the lowest values as -1 and highest values as $+1$.

$$Z_i = \frac{w_i - w_0}{\Delta w_i} \quad (2.60)$$

Where w_i and w_0 denote the WPF's real value at the axial and center points, respectively.

Gunaraj *et al.* (1999) and Cho (2010) define Δw_i as the phase change and Z_i as the dimensionless value of an explanatory variable.

However, if any property of the design is needed, such as rotatability or the ability to accommodate a cuboidal area, the value of Z_i should be carefully chosen. A well-known choice for a totally random CCD is always $a = n_f^{0.25}$ where $n_f = 2^k$. Thus, the CCD within SPDs lends itself to pure error estimates of the variance elements. The center point of each subplot is supposed to be repeated n times and the point of the center point is replicated r times. According to Goldfarb *et al.* (2003) the pooled estimate of the sub plot error variance given by Equation (2.61) is then considered as follows

$$S_{sp}^2 = \frac{\{(n-1)S_1^2 + (n-1)S_2^2 + \dots + (n-1)S_{2k}^2 + r(n-1)S_c^2\}}{\{(n-1) + (n-1) + \dots + r(n-1)\}} \quad (2.61)$$

Where

$$S_l^2 = \frac{1}{(n-1)} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2, \quad l = 1, 2, \dots, 2k,$$

$$S_c^2 = \frac{1}{r} \sum_{j=1}^r (y_{ij} - \underline{y}_i)^2, \quad l = 1, 2, \dots, 2k,$$

From above Equations represents the split plot error variance components estimated from the axial points and the replicate at the middle of WPFs, respectively. It is also noted that

$$E[S_{sp}^2] = E\left(\frac{\{(n-1)S_1^2 + (n-1)S_2^2 + \dots + (n-1)S_{2k}^2 + r(n-1)S_c^2\}}{\{(n-1) + (n-1) + \dots + r(n-1)\}}\right), \quad (2.62)$$

$$= \frac{\{2k(n-1) + r(n-1)\}\sigma^2}{\{2k(n-1) + r(n-1)\}} = \sigma^2. \quad (2.63)$$

Therefore, $E[S_{sp}^2] = \sigma^2$ is an unbiased approximation of $\hat{\sigma}^2$ with degree of freedom $(n-1)(2k+r)$. The whole plot variance components are considered next if the pooled estimation of the variability is based on whole plot averages of the repeated whole plot core rather than subplots as

$$S_{wp}^2 = \frac{1}{(r-1)} \sum_{i=1}^r (\underline{y}_i - \underline{y}_{..})^2, \quad (2.64)$$

where $\underline{y}_{..}, \underline{y}_i$ denotes the overall mean as well as the mean of the values in the i^{th} whole plot.. It can also be noted that

$$E(S_{wp}^2) = E\left[\frac{1}{(r-1)} \sum_{i=1}^r (\underline{y}_i - \underline{y}_{..})^2\right], \quad (2.65)$$

$$= E\left[\frac{1}{(r-1)} \sum_{i=1}^r \{[\delta_i - \underline{\epsilon}_i] - [\underline{\delta}_i - \underline{\epsilon}_{..}]\}^2\right],$$

$$\begin{aligned}
&= \frac{1}{(r-1)} \left[E \left(\sum_{i=1}^r [\delta_i - \underline{\delta}_i]^2 - \sum_{i=1}^r [\underline{\epsilon}_i - \underline{\epsilon}_i]^2 \right) \right], \\
&= \frac{(r-1)\sigma_\delta^2}{(r-1)} + \frac{(r-1)\sigma^2}{n(r-1)}, \\
&= \sigma_\delta^2 + \frac{1}{n}\sigma^2. \tag{2.66}
\end{aligned}$$

Therefore, $E(S_{wp}^2) = \sigma_\delta^2 + \frac{1}{n}\sigma^2$ is a $r-1$ degree of freedom unbiased estimation that is often referred to as pure-error estimates. This is because they are used to estimate the model's variance elements, which are model free and unaffected by model misspecification, as Njoroge *et al.* (2017) pointed out.

2.12 The Ecology and Significant Production of Soybean Globally

The *Glycine max* crop has been ranked first among various oilseed crops globally, contributing an estimated 25% of total oil and fat production. *Glycine max* production in the world's farming soybean sector, which is about 101.81 ha, is about 253.38 million tons, with a productivity rate of 2.5 tons per ha (Singh *et al.*, 2012).

Glycine max production in Kenya is still inadequate, according to Mahasi *et al.* (2011), averaging 2000-5000 metric tons (MT) per year. Furthermore, according to FAOSTAT (2011), industrial demand for *Glycine max* products increased from 5000 MT in 2008 to approximately 120000 MT in 2019, with about 150 MT of soybean protein concentrates and textured soybean protein.

According to Mbembe (2020), human consumption in Kenya accounts for 10-15% of total output, or 1000-15000 MT per year, implying that *Glycine max* imports are currently meeting a portion the domestic demand. However, Western, Nyanza, Rift Valley, Central,

and Eastern regions of Kenya have become famous for their high *Glycine max* output, which covers around 25000 ha and yields 0.8 t/ha on average. According to Mahasi *et al.* (2011), a capacity of 1.5-3.0 t/ha exists depending on the region. In 1999, the Ministry of Agriculture (MOA) projected that 135800 hectares would need to generate over 108000 metric tons to achieve self-sufficiency. Due to increased understanding of the health benefits and nutritional value of soybean, international agriculture organizations such as the International Center for Tropical Agriculture (CIAT) and the International Institute of Tropical Agriculture (IITA) have seen improvements *Glycine max* sub-sector in collaboration with the government. Research shows that *Glycine max* has excellent adaptability towards a broad range of soils and varying climatic conditions (Zhang *et al.*, 2020)

However, most researchers advocate the soil with a PH of 6.0 as best suitable for growing *Glycine max*. Soybean are used to make various Kenyan dishes, including bread, chapatti, milk sweets, and pasties. For most SDA church members in Kenya, it is one of the most well-known beverages. They use as a fermented commodity and a pulsed seed. Despite its popularity in Kenya, the farmers give very little priority for its cultivation on a large scale. Like other leguminous crops, the requirement of nitrogen, as reported by Kamble *et al.* (2009), is substantially fulfilled from symbiotic nitrogen fixation through Rhizobium. *Glycine max* is a source of high-protein foods for children. They use it widely in the industrial development of various antibiotics, as described by Singh *et al.* (2012), it is. According to Chianu *et al.* (2008), soybean improves soil fertility by fixing significant atmospheric nitrogen levels by root nodules and leaf fall on the ground at maturity.

According to Wanyama (2013), *Glycine max* responds well to organic and inorganic fertilizers. They aid in nodule formations in legume crops that, in turn, fixed atmospheric nitrogen needed for their growth, as Singh *et al.* (2012) described. The agro-climatic conditions prevailing in Kakamega County parts in western Kenya are highly favorable for *Glycine max* cultivation.

2.12.1 Effect of organic and inorganic fertilizer on *Glycine max* Crop

According to Kimetu *et al.* (2008), organic fertilizer is a relatively heterogeneous category of products that includes anything from single chemicals like urea to highly complex and variable mixtures of organic and inorganic compounds like municipal wastes in urban areas. Further, wastes from animal production, raw sewage sludge, and products processed from sewage are of greater interest for agricultural plant production.

Organic and inorganic fertilizers both have positive and negative effects on plant growth and soil. According to Han *et al.* (2016), inorganic fertilizers are relatively cheap and have the high nutrient content need by the plants. Furthermore, as defined by Kamble *et al.* (2009), excessive inorganic fertilizer use can result in nutrient loss, surface water and groundwater pollution, soil and acidification, the decline in critical microbial organisms, and increased susceptibility to harmful insects. As opposed to chemical fertilizers, organic fertilizers have a range of disadvantages, including low nutrient content, slow decomposition, and different nutrient compositions based on the same organic materials (Wanyama, 2013).

According to Han *et al.* (2016), organic fertilizer derived from livestock manure has been used in agriculture fields to combat environmental pollution and crop productivity losses caused by the constant improper use of inorganic fertilizer. Several authors (Kamble, 2009;

Wanyama, 2013) advocated recycling livestock waste to avoid pollution and minimize treatment costs. At the same time, this improves soil quality and increases agricultural production. However, using inorganic and organic fertilizers at the same time has yielded a range of results depending on the type of the crop, soil and soil properties. Using a combination of nitrogen (N) phosphorus (P), potassium (K) chemical fertilizer, and organic manure derived from animal products increases the mean growth of *Glycine max* by 46 % and soil N, P, and K concentrations by 26 %, 129 %, and 65 %, respectively.

However, Hans *et al.* (2016) researched a short rotation Willow plantation in the Middle East region of North America to increase biomass production. They used slow-acting inorganic fertilizer and organic manure made from animal waste. The organic manure treatments significantly increased Willow growth, PH at a soil depth of 0-10 cm by 2, and soil K, P, and magnesium concentrations (Mg). In the case of a poplar plantation in clay soil, Wanyama (2013) found that inorganic fertilizers encourage higher growth and root production than organic fertilizer derived from livestock manure because of its availability and good nutritional value to the soil.. Since plant roots play a critical role in individual plants' functionality, Zhang *et al.* (2020) discovered that organic manure and biochar amendments increase soil fertility and crop productivity. They also found that organic manure and biochar amendments improved root morphology and physiology by increasing soil nutrients, with a substantial increase in root physiology associated with an increase in soil nutrient content at the crop's bud level.

Kimetu *et al.* (2008) found that raising the phosphorus levels and farmyard manure derived from animal waste increased *Glycine max* dry matter yield, nitrogen, phosphorus, and potassium content. Kamble *et al.* (2009) examined the response of Soybean on the

application of organic fertilizer obtained from animal products together with rice straw in the proportion of 1:1 at 0, 5 and 10 t/ha, Phosphorus (37.5 and 56.25 Kg P_2O_5 /ha and noticed that at the peak growth of 60 days after sowing. However, with 5, and 10 t/ha of organic manure, the leaf area index increased substantially from 2.57 in the control to 3.41 and 4.05 in the organic manure treatments. Similarly, Soybeans seed yield increased drastically from 20.7 to 23 and 26.9 q/ha with steadily rise from 0 to 5 and 10 t/ha organic fertilizer. In addition, Furthermore, Wanyama (2013) reported increased dry matter output at different growth stages as a result of increased dry matter production at 22.7 to 24.1 q/ha with 37.5 to 56.25 Kg P_2O_5 /ha.

Isaev *et al.* (2020) performed a field study on the reaction of Soybeans to Rhizobium inoculation and discovered a large increase in response with Rhizobium treatments over no inoculation at all stages. They also discovered that seed inoculation increases seed yield substantially more than inoculation alone when a 30 kg N/ha initial dose is applied. According to Tittonell *et al.* (2008), this is attributable to a rise in the number of nodules per plant and nodule dry weight, as well as their subsequent translocation to the seed.

Furthermore, Wanyama (2013) stated that among the different treatments, the soil with enriched Farmyard manure contributes to a substantial increase in plant height, number of branches per plant stem, and number of pods per plant. Their study also revealed that the yield of *Glycine max* harvested depends with the seasons which implies that the productivity of Soybean is affected with the season at which it is planted. This is depicted with their results that was obtained during summer and Kharif where grain yield of 1259 Kg/ha and 1499 Kg/ha, respectively. They also looked at the effects of Farmyard manure (FYM, 0 and 10 t/ha) and Sulphur fertilizer (10, 20, 40, and 60 Kg/ha) on the

growth and yield of *Glycine max* cultivars Js-335 and NRC-12, and found that all of the parameters increased FYM and Sulphur fertilizer rates. However, Sulphur fertilizer at 40 Kg/ha recorded the highest seed about 15.59 q/ha and straw yield as described by Singh *et al.* (2012).

According to Zhang *et al.* (2020), *Glycine max* yield and N and P uptake were quadratically raised with increasing broiler inorganic fertilizer application rates. They revealed that Soybean grain yield and N uptake from broiler litter application were significantly greater than those from inorganic fertilizers. Furthermore, Kamble *et al.* (2009) and Wanyama (2013) examined the influence of organic and commercial fertilizers on nodulation in *Glycine max* and reported that the number of root nodules at 60 for Diammonium Phosphate (DAP). The highest number of root nodules per plant was observed under bio-digested slurry combined with 30:120:40 Kg/ha NPK.

2.12.2 Effect of Row Spacing and Seeding Rate on *Glycine max* Variety of Seeds

Glycine max drilled in rows produced more than the one planted in rows with a row crop planter and more than Soybean planted in rows, according to Kimetu *et al.* (2008). Row spacing research in the Central and Southern United States, on the other hand, found little difference in yield between narrow and broader rows, according to Wanyama (2013). According to some reports, row spacing and seeding rate interactions with *Glycine max* yield greater with higher seeding rates and narrow rows compared to wide rows (Zhang *et al.*, 2020).

Swain *et al.* (2006) posted a \$ 30 per ha profit with a seeding rate of 420000 seeds/ha in 19 cm rows versus 321,000 seeds/ha in 76 cm rows due to higher yields outweighing seed costs. On the other hand, according to Mbembe (2020), other studies have found similar

optimum seeding rates in narrow and long rows, indicating that there is no relationship between row spacing and seeding rate. On the other hand, glycine max plants compensate for space in the canopy by increasing more branches, but increased seeding rates had no impact on yield. Because Glycine max plants compensate for lower seeding rates with biomass, pods, and seeds per plant stem, but they also compensate for wider rows (Kamble *et al.*, 2009). The *Glycine max* compensates for lower seeding rates and broader rows. It is less successful at compensating for broader rows than for lower seeding rates, meaning that row spacing has a more significant impact on yield than seeding rates. Furthermore, more seeding rates at early vegetative stages can boost *Glycine max* growth, leading to increased yield. Again, the above seeding rate does not increase vegetative growth or follow a pattern of rising or decreasing yield (Chianu *et al.*, 2008). The fungus *Sclerotinia sclerotium* causes Sclerotinia stem rot, also known as White mold. This disease often causes low soybean production. As a result, management practices like narrow spacing, increased *Glycine max* plant populations, early planting season, well-drained soil, and high soil fertility, as defined by Tittonell *et al.* (2008), can increase yield. Raising white growth inside the Soybean canopy may often have unintended consequences. While fungicides are available to combat white mold, it is generally impossible to completely eradicate the disease using only chemical management. Other than fungicides, more management strategies are required to eliminate white mold in soybeans, such as cultivar selection and management practices. Furthermore, planting with a large row spacing or a lower plant population delays canopy formation, resulting in a decrease in canopy density and thus preventing the growth of white mold.

2.12.3 Effect of Soil pH on *Glycine max* (L.) Merrill Yield

Soil PH has an essential effect on the soil biogeochemical process in the natural world. It is known as the "master soil component" because it affects the various biological, chemical, and physical properties in the soil that affect *Glycine max* growth and biomass yield. According to Zhang *et al.* (2020), soil pH is only essential for soil chemistry and fertility. Furthermore, the recognition of soil functions beyond plant nutrient supply and the importance of soil as a medium for plant growth necessitated a multidisciplinary analysis of soil and its characteristics in light of broader ecosystem functions. Many biogeochemical processes are affected by soil pH, according to decades of study. Recent research has revealed some fascinating facts about the importance of soil pH in various soil processes. According to Neina (2019), soil pH regulates the solubility, mobility, and bioavailability of trace elements, determining their translocation in plants. As a result the content of dissolved organic matter increases with soil pH, which accounts for the strong effects of alkaline soil pH conditions on dissolved organic carbon and nitrogen leaching observed in many soils containing significant amounts of organic matter. However, Neina (2019) found that soil PH conditions needed for microbial activity range from 5.5-8.8. As a result, soil respiration often rises to an optimum level as soil pH rises. However, this is also related to higher organic carbon and nitrogen levels in microbial biomass above pH 7. Furthermore, soil pH influences biodegradation through its impact on microbial behavior, microbial population and diversity, enzymes that aid in the degradation process, and the characteristics of the substances to be degraded, just as it does many other soil biological processes (Gentili *et al.*, 2018). Alkaline or slightly acid soil pH aids in biodegradation. Acidic conditions, on the other hand, restrict biodegradation and cause a decrease in

Glycine max yield (Jagwe and Nyapendi, 2004). According to Gentili *et al.* (2018) and Neina (2019), pH values of 5.4 to 8.0 are optimal for oil degradation, making them suitable and adequate for growing cereal and legume crops.

The soil's pH rises to a peak and then drops when they combine organic matter from livestock or raw plant residues with unburned soil. The initial soil pH also increases when they apply a mixture of sludge derived from a bleach factory, urban solid waste, and a mixture of sludge extracted from a purification plant or residual agro-food industries to the soil.

Most micronutrients, for example, are more accessible to plants in acidic soils than in neutral-alkaline soils, favoring plant growth, according to Ayalew (2011). Also, as Gentili *et al.* (2018) point out, pH influences various plant characteristics (i.e., traits) such as height, lateral spread, biomass, flower sizes, pollen development, and many others.

2.13 Screening Methodology in Mixture design with application to *Glycine max*

A mixture test is an experiment where the descriptive variable (factors) and response rely only on the mixture's relative ratio in the mix but not its composition. For example, the yield of crops may be the maximum number of *Glycine max* per stem or the number of seeds per stem. In the most basic mixture design test, the q component in the compound meets the following barriers.

$$\sum_{i=1}^q x_i = 1 \qquad 0 \leq x_i \leq 1 \qquad (2.67)$$

The proportion of each blend must be between 0 and 1. Also, the proportions of the q blends in the mixture must total up to unity. The factor components space for an experiment with

constraints in Equation (2.67) is a $q - 1$ dimensional simplex that may include the design space's edge and interior.

However, experiment with mixtures was officially formalized by Henry Scheffe in 1958, where the simplex lattice design (SLD) and corresponding Scheffe canonical polynomial model was formally introduced (Goos *et al.*, 2016). Scheffe defines a (q, m) lattice to fit the design where q and m represent the number of components in the mixture and the polynomial model's degree, respectively. They are $\binom{m+q-1}{m}$ candidate points in a simplex lattice design. The proportions applied for each component have $m + 1$ equally spaced values from 0 to 1 of $x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1$. One-to-one correspondence of candidates points to the polynomial model parameters, as pointed out by Weese (2010). For instance, in a $(q, 1)$, SLD is the form:

$$\eta = \sum_{i=1}^q \gamma_i x_i \quad (2.68)$$

subject to the substitution

$$x_q = 1 - \sum_{i=1}^q x_i \quad (2.69)$$

into the standard polynomial model form:

$$\eta_1 = \gamma_0 + \sum_{i=1}^q \gamma_i x_i \quad (2.70)$$

According to Cornell (1988) and Goos et al. (2016), the polynomial coefficient has a one-to-one correspondence with the points in the design. As mentioned above, there are q candidate points in a $(q, 1)$ SLD; hence for a three-mixture blend, there are three candidate points, and in the corresponding polynomial, three parameters to be approximated that enables for the coefficients to be compared employing least squares (MLS) regression, Maximum likelihood method (MLM), restricted maximum likelihood (REML), and ordinary least squares (OLS). Scheffe defines a second-order polynomial model for mixtures where the anticipated response to take on a nonlinear form as:

$$\eta = \sum_{i=1}^q \gamma_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \gamma_{ij} x_{ij} \quad (2.71)$$

In this polynomial model, the pure quadratic terms are combined with the two-factor quadratic terms owing to the substitution Equation 2.72 to model 2.71.

$$x_i^2 = x_i \left(1 - \sum_{i=1, j \neq i}^q x_j \right) \quad (2.72)$$

As described by Goos *et al.* (2016) and Weese (2010), in addition to the substitution used in the model (2.70). However, with this substitution, the polynomial degree remains unchanged, and the number of terms $\binom{m+q-1}{m}$ maintaining the one-to-one correspondence of design points and parameters in the model.

In mixture design experiments, the interaction terms in the model are commonly known as nonlinear blending terms. However, the nonlinear blending terms, response to binary and

ternary or quinary mixtures, can be perceived and illustrated as being either a synergistic effect or an antagonistic effect. These interpretations of binary, ternary, and quinary mixture terms are broadly used in describing the impact of components on the characteristics of a mixture.

Gorman and Hinman (1962) protruded the Scheffe' polynomial model up to the third and fourth degree. They were the first to note a caution regarding Scheffe's polynomials when the design space is small. Furthermore, Lambrakis (1968) generalized Scheffe's polynomials to order m . Scheffe traced his pioneering work within the realm of mixture experiments to introduce SCD. This design encompasses $2^q - 1$ mixtures consisting of q pure component blends, $\frac{q}{2}$ binary ingredients with equal proportions, and $\frac{q}{3}$ ternary mixture blends with equivalent ratios up to the q - nary combination with similar proportions as described by Goos *et al.* (2006). The simplex centroid consists of observations on mixtures containing every (non-empty) mixture that the components present appear in equal proportions as reported by Weese (2010). Similarly, with the SLDs, the SCD has a one-to-one correspondence with Scheffe's polynomial model. The coefficients are approximated by utilizing linear combinations of the responses at each of the candidate points. Later, in 1963, Scheffe also introduced process variables in the mixture experiment design as a factorial experiment.

Several researchers, including Scheffe', recognized some of the limitations of the SLD and SCDs (Cornell, 1975; Goos *et al.*, 2016; Weese, 2010). Further, one of the drawbacks realized by Scheffe' was that not all of the components in some situations are allowed to vary from zero to unity. Therefore, this led him to introduce pseudo-elements for a case where one of the blends has an upper bound. However, his polynomial coefficients become

non-trivial during interpretation, as pointed out by Weese (2010). Additionally, the lower order designs do not include many if any interior points in the experimental region since, in some cases, the only feasible experimentation is with the internal issues. Kowalski *et al.* (2002) acknowledge this paucity and regard designs for three and four factors where all the candidate points are interior. Their prime objective of experimentation is to fit a response surface model to the experimental area. These points are established by applying the design criterion proposed by Njoroge *et al.* (2017)

Myers *et al.* (2009) also proposed a design to approximate the response surface mixture by applying an ellipsoidal region centered at a point of maximum interest to the user. The ellipsoid is determined by the researcher in these types of designs.

Lambrakis proposed multiple lattice designs where significant and minor mixture components exist in Scheffe's mixture problem. Further, the generalized Scheffe canonical Polynomial and advocated analyzing these numerous lattice designs. Cornell (2011) later introduced the alternative model to the canonical Polynomial. The model he proposed was a homogenous model of degree one for fitting mixtures to interpret an inert or additive component in the mix. Also, he discussed regression procedures for mixture variables in his subsequent paper. Various types of models, such as the extended contrast model (Aitchison *et al.*, 1984) and a model including inverse terms (Cho, 2010), were developed to account for possible extreme response changes as a component approaches zero. Despite the massive amount of literature on mixture experiments, the issue of screening in a mixture environment, as defined by Weese (2010), has only put it in a better form, one division in mixture experiments, and deodorized in formulation research setup. The entire literature on experimental screening concepts organizes around evaluating overall linear effects

using the first-degree Scheffe polynomial model, closely resembling the description of a factorial screen design defined by some researchers (Cornell, 2011; Weese, 2010). As specified by Box and Hunter (1957) a Factorial screen design's primary goal is to decide which of a large number of component factors are of critical importance. We can also identify the prime component variables and then reexamine the experiment. However, as Cornell (1975) described, the first-order model should be used to evaluate a screening experiment to determine the most significant factors out of a large number based on their first-order effect estimates. Snee and Marquardt (1976) were the first to address screening principles with mixture experiments. They stated that the only difference between the screening theory for mixture variables and ordinary independent variables is in the screening principle in mixture settings. They use candidate points lying on what they refer to as components axes to explicitly translate the theory of looking for significant linear effects in a screening experiment. They effectively research directions through a factor region where the measured response remains unchanged or nearly unchanged to evaluate the critical mixture variables in this manner. They advocated using the Scheffe canonical polynomial to model the mixture's linear mixing and then checking the results in contrasts to see which effects are equal or not, as described by Weese (2010). As a consequence, they are a decline in the number of components needed to perform follow-up experiments. The screening philosophy is currently adapting from dependent variables (response) to mixture environments. However, Cornell (2011) discussed the Cox polynomial model and the use of the principle of effect path to find candidate points, the first order Scheffe polynomial to model, test, and assess the results. However, according to Njoroge *et al.* (2017), some researchers did not say anything about having five or fewer components.

However, they urged that even if the designs are for six or more mixture components, it would be easier to collect enough data to match a quadratic model. They also suggested using simplex screening designs where the components can vary across the entire composition spectrum or the design space can be expressed in pseudo components. These designs contain all pure component ingredients, the centroid, internal points on the component effect axes, and end effect points in cases where it assumes the complete absence of a component would negatively affect the response. In this case, the edge's midpoint refers to the end impact points where the primary concern component is absent from the mixture, but the other components are present.

Expanding the architecture of these factorial points to the mixture setting by adding and subtracting $+\Delta_1$ or $-\delta_2$ and $+\Delta_2$ or $-\delta_1$ where 1 and 2 are the components whose relationship is to be approximated to the mixture setting as defined by Weese (2010). To approximate the interaction, the researchers use the Cox effect instructions to include and exclude these quantities and the regular mixture. Cornell (2011) goes into greater depth regarding screening in a mixed setting. It becomes possible to conduct experimental runs and resolve the essential components based on the size of their effects when there are six or more component factors. He describes screening as reducing the number of component factors to the point that only the most essential components are needed further. As several authors have pointed out, the construction of screening designs and the establishment of screening models often begin with the Scheffe first degree model (Cornell, 2011; Weese, 2010). Cornell (2011) also recommends bringing the ranges of the various component variables as close together as possible so that the components' relative effects can be designed using the effect approximate to its standard error ratio. Indeed, if the component

factors' ranges are close, the magnitude of the beta estimates can infer the more significant impact on the response.

Several authors propose using the Cox mixture model for mixture screening design, arguing that the Cox model offers a natural way to integrate this current formulation into the experiment. Also, use the standard response as a basis for comparison when adding components to the mixture. Usually, the parameters in a Cox mixture model denote the slope between a mixture point x and the response at a standard mixture, as defined by Cornell (2011).

Furthermore, the experimental region's labyrinthine constraints render interpreting individual Scheffe parameters a waste of time. In this case, the Scheffe model can only be used to select significant effects or to create a complete model of the surface, not to understand individual actions on the expected response, as Cornell (2011) describes. Furthermore, when the number of mixture components is high, as it is in industrial settings, it is not practical to analyze more than linear results, as stated by Weese (2010). For smaller, simpler mixtures, the new approach works well. However, as Cornell (2011) and Weese (2010) point out, fitting a model to the entire space is not always the most realistic way to experiment with large mixtures with complicated constraints.

Furthermore, as indicated by Weese, they are likely to be a current environment, a standard mixture, and a consistently beginning point in this form of experimentation. We also agree that it is not always appropriate to investigate or research all of the components' effects. In chapter four, we show numerical examples of this. When it comes to agricultural mixture screening, the most critical components are discovered not by quantifying the most significant linear effect but by calculating the concern response changes at the new point

and comparing it to the previously agreed response. However, we assume that the Cox mixture model type structure is the most important for screening a mixture in an agriculture setting while growing varieties of crops subject to various treatments.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This section outlines the methodology on the split-plot designs and the concept of mixture experiments and also how each of the study objectives will be achieved.

3.2 Developing a Parsimonious Model for Use of Split-Plot Design and MPV

The SPD was made up of a simplex centroid design (SCD) of four mixture blends and a 2^2 factorial design with a central composite design (CCD) of the process variable. The SPD comprised 54 treatment combinations. The four mixture blends were denoted as X_1, X_2, X_3, X_4 and set up in SCD with the following eleven blends;

$$\begin{aligned}
 & (x_1, x_2, x_3, x_4) \\
 & = (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1), (0.5, 0.5, 0, 0), \quad (3.1) \\
 & (0.5, 0, 0.5, 0), (0.5, 0, 0, 0.5), (0, 0.5, 0.5, 0), (0, 0.5, 0, 0.5), (0, 0, 0.5, 0.5), \\
 & (0.25, 0.25, 0.25, 0.25).
 \end{aligned}$$

The two process variables were coded as Z_{SP} and Z_{WP} had two levels each plus additional point of CCD as shown in the Equation (3.2) and Figure 3.1 where $Z_{SP} = Z_1, Z_{WP} = Z_2$ is sub-plot and whole-plot, respectively.

$$\begin{aligned}
 & (Z_{SP}, Z_{WP}) \\
 & = (1, 1), (1, -1), (-1, 1), (-1, -1), (1.414, 0), (-1.41, 0), (0, 0), \\
 & (0, 1.414), (0, -1.414), (1.414, 1.414), (-1.414, -1.414). \quad (3.2)
 \end{aligned}$$

These initial model as described by Njoroge et al. (2017), was proposed and extended from 3 to 4 mixture components as shown Figure 3.2 and 3.3. Their model 1 consisted of seven mixture blend set up at each of the four points of the factorial arrangement. In model 2,

they set up four points of the factorial design at each of the seven mixture blends of the simplex centroid design. Still, they found that model 1 was more efficient. It also provided more concise parameter estimates in terms of A-, D- and E-optimality criteria because it had more sub-plots than whole-plots since SPD provides room to measure the effect of change of process variable the different mixture ingredients. We extended model 1 by looking at six other alternative arrangements of design points in a split-plot design as discussed in section 3.2. The process variables were the whole-plots and the mixture ingredient the split-plots as shown in Figure 3.1. Our split-plot design consisted nine whole-plot with each having six sub-plot for all the six alternative arrangement of the candidate points in a SPD. We created the six different design option using D-optimal as discussed section 3.2 purposely to assess the best design can suitably fit model (3.4) using the proposed SPD shown in Figure 3.2 and 3.3.

3.2.1 New Design for Split-Plot Layout to Support Fitting MPV

The Figure 3.1 shows Proposed Design for Split-Plot layout to Support Fitting mixture process variable with CCD of second order polynomial model. Figure 3.3 depicts the newly developed Design for Split-Plot Structure to Support Fitting the Combined Second-Order MPV model after extending Model 1 proposed by Njoroge *et al.* (2017). The center point $[z_1, z_2] = \{0,0\}$, v times is replicated, and each time the centroid $(X_1, X_2, X_3, X_4) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$, k times is replicated. Also, the centroid $(X_1, X_2, X_3, X_4) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ at each axial setting is replicated k times. We formulated of the model within a split plot design as follows

$$Y = g(x, Z_{SP}, Z_{WP}) = g_w(Z_{WP})' \beta_{WP} + g_s(x, Z_{SP}, Z_{WP})' \beta_{SP} + \varrho + \varepsilon \quad (3.3)$$

$$\begin{aligned}
Y = & \sum_i \beta_i x_j + \sum_{i < j} \sum \beta_{ij} x_i x_j \\
& + \sum_i \sum_{wp} \vartheta_{ip} x_i Z_{wp} \\
& + \sum_{i < j} \sum_{wp} \sum \vartheta_{ijp} x_i x_j Z_{wp} + \sum_i \sum_{sp} \mu_{isp} x_i Z_{sp} \\
& + \sum_{i < j} \sum_{sp} \sum \mu_{isp} x_i \mu_{isp} x_i x_j Z_{sp} \\
& + \sum_i \sum_{wp} \sum_{sp} \gamma_{ijspwp} x_i Z_{wp} Z_{sp} \\
& + \sum_{i < j} \sum_{wp} \sum_{sp} \sum \gamma_{ijspwp} x_i x_j Z_{wp} Z_{sp} + \varrho + \varepsilon
\end{aligned} \tag{3.4}$$

where β_{wp} is a vector representing the coefficient terms drawn from the Whole-plot variable, β_{sp} is a vector containing the coefficient terms resulting from the sub-plot variable, $\varrho \sim N(0, \sigma_{wp}^2)$, represent the random error associated with the entire plot factor by itself during the randomization level, and $\varepsilon \sim N(0, \sigma_{sp}^2)$ indicate the random error that is associated with sub-plot randomization level. However, σ_{wp}^2 and σ_{sp}^2 are assumed to be statistically independent and distributed. This model (3.3) can still be simplified by omitting $g_w(Z_{wp})' \beta_{wp}$ because whole-plot factor affects only the response through the interaction mixture component variable. Therefore, simplified model reduces to

However, considering the four mixture components and two process variable using the Scheffe technique in the framework of CCD within SPD, then model (3.4) reduced to

$$\begin{aligned}
Y(x, z) = & \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 \\
& + \beta_{14} X_1 X_4 + \beta_{23} X_2 X_3 + \beta_{24} X_2 X_4 + \beta_{34} X_3 X_4 \\
& + \vartheta_{11} X_1 Z_1 + \vartheta_{12} X_1 Z_2 + \vartheta_{21} X_2 Z_1 + \vartheta_{22} X_2 Z_2 \\
& + \vartheta_{31} X_3 Z_1 + \vartheta_{32} X_3 Z_2 + \vartheta_{41} X_4 Z_1 + \vartheta_{42} X_4 Z_2 + \varrho + \varepsilon
\end{aligned} \tag{3.5}$$

where Z_1 the control is process variable and Z_2 is the noise process variable, β_i is the vector of fixed effect resulting from mixture blend of the vertices of component X_i , β_{ij} is the vector of random effect resulting from the interaction between mixture components and ϑ_{ij} is the vector of random effect resulting from the interaction between mixture components and process factors. The model (3.5) is an empirical model that corresponds well with the experience and plots of the data. The random component effect of the model has a whole plot and split-plot contribution. The whole plot error is nested under x_1, x_2, x_3 , and x_4 , while the subplot error is the standard residual error term. However, the model (3.5) under split plot design was further simplified to

$$Y_{jk} = X_{jk}\beta + d_{jk}\delta_j + \varepsilon_{jk} \begin{cases} j = 1, 2, \dots, n_w \\ k = 1, 2, \dots, n_j \end{cases} \tag{3.6}$$

where Y_{jk} represents whole plot j at k^{th} measurement response variable subject to split-plot factors and process variable. n_w denotes the number whole plot while n_j number of measurements in whole plot j . d_{jk} indicates a covariate vector of j^{th} whole plot at k^{th} measurement for random effects $\delta_j \in \mathbb{R}^q$ associated with whole plot effect where q is the number of factor components applied in split plot layout experiment.

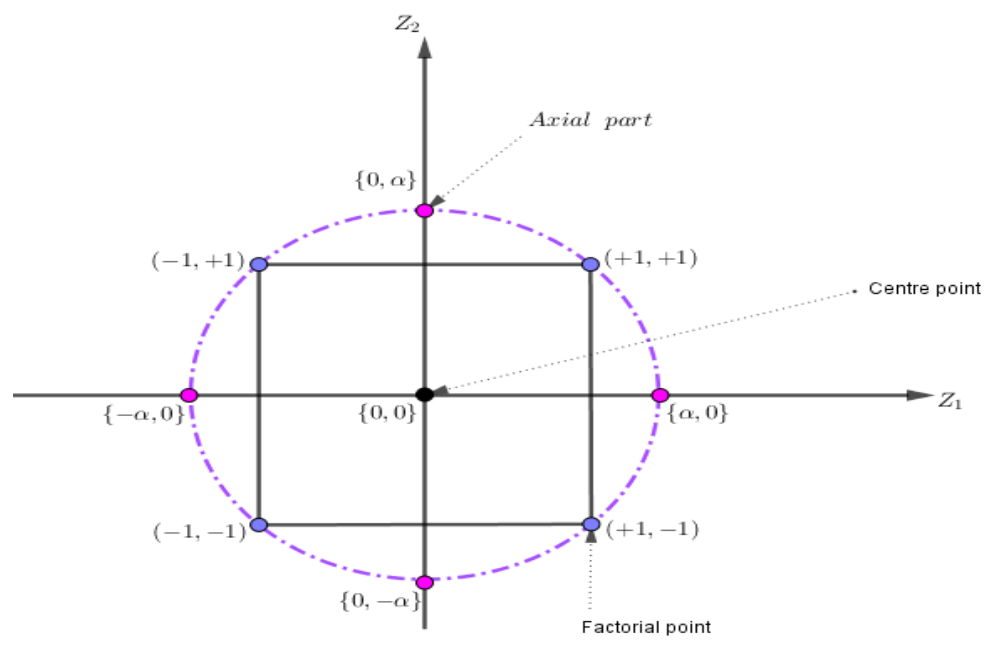


Figure 3.1: A 2^2 factorial design with CCD of process variable

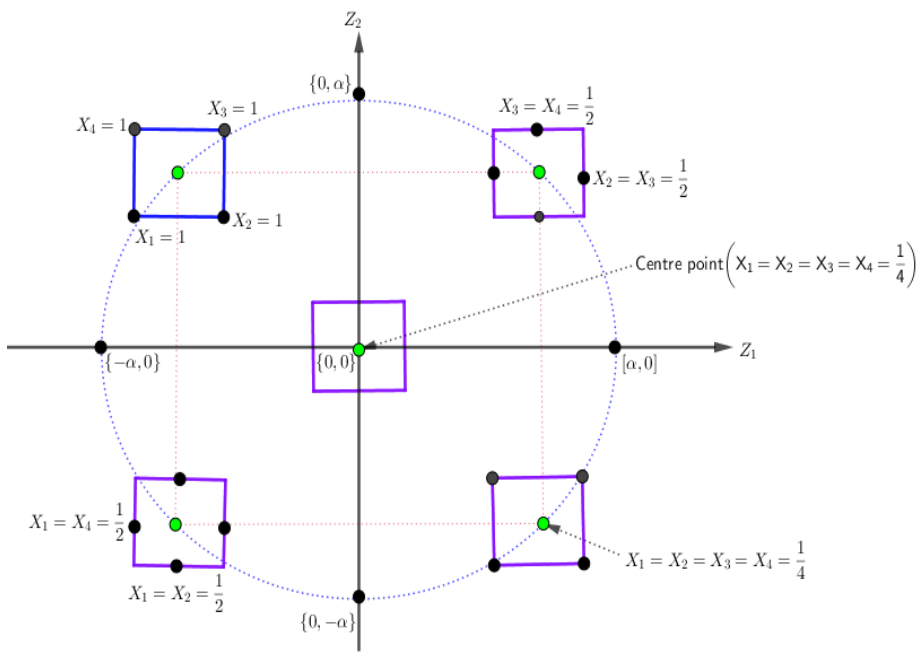


Figure 3.2: A Proposed Design for Split-Plot layout to fit MPV model with CCD

$$\delta = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_b^2 \end{pmatrix} \in \mathbb{R}^{n \times q}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_b \end{pmatrix} \in \mathbb{R}^b,$$

$$U = \begin{bmatrix} \varepsilon_1 & 0 & \dots & 0 \\ 0 & \varepsilon_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \varepsilon_b \end{bmatrix} \in \mathbb{R}^{n \times N} = \sigma^2 I_n,$$

Therefore, the statistical linear model matrix formulation can be written as

$$Y = X\theta + D\delta + \varepsilon \quad (3.7)$$

$$\text{Where } \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix} \sim N_{bq+n} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} Z & 0 \\ 0 & U \end{bmatrix} \right) \text{ and } Z = \sigma_b^2$$

where Z is the Whole Plot Error (WPE) and U is the split-plot error (SPE).

However, if we let

$$Y = X\theta + \varepsilon^* \text{ where } \varepsilon^* = D\delta + \varepsilon = [D \quad I_{n \times n}] \begin{bmatrix} \delta \\ \varepsilon \end{bmatrix} \quad (3.8)$$

This implies that

$$\varepsilon^* \sim N(0, V), \quad (3.9)$$

$$Y = X\theta + P \begin{bmatrix} \delta \\ \varepsilon \end{bmatrix} \text{ since } P = [D \quad I_{n \times n}],$$

$$\text{Var}(Y) = \text{Var} \left(P \begin{bmatrix} \delta \\ \varepsilon \end{bmatrix} \right),$$

$$\text{Var}(Y) = P \text{Var} \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix} P',$$

$$\begin{aligned} \text{Var}(Y) &= [D \quad I_{n \times n}] \begin{bmatrix} Z & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} D' \\ I_{n \times n} \end{bmatrix} = DZD' + U \\ &\Rightarrow V = DZD' + U \end{aligned} \quad (3.10)$$

where

From Equation (3.8) and (3.9) are termed as marginal model. Therefore from this the statistical model can be written as two-level hierarchical model if the random effect in the whole plot is known.

From Equation (3.11) implies that $\delta \sim N_n(0, U)$, the parameters δ and θ can estimated

$$Y|\delta \sim N_n(X\theta + D\delta, U), \quad (3.11)$$

using Ordinary Least Square (OLS), maximum likelihood (ML), restricted maximum likelihood (REML) and Bayesian method but in this case, we restrict to ML and REML.

3.2.2 Construction of MPV Design for Split Plot Structure Using D – Optimal Designs

The design algorithms are required to find D-optimal SPDs for MPV designs. In the literature on finding the optimal design of experiments, the most popular algorithms are either the point transfer algorithms or the candidate-set free integration transmission algorithms implemented statistical software JMP (Goos *et al.*, 2016). Furthermore, the D-optimal design used in the study was calculated using the point algorithms described in Sitinjak and Syafitri (2019). The algorithm is developed primarily to calculate A-, G-, I-, and D-optimal efficiency with given numbers and sizes of whole plots.

The FORTRAN code of the algorithm and the input files needed to compute the designs are executed in SAS's JMP software section. This algorithm implemented in JMP software requires a specification of observations and split-plot configurations, including the total

number of whole plot, n_{WP} , and the number of b_i observations in each complete plot. Furthermore, a prior guess of the variance component ratio $d = \sigma_\delta^2 / \sigma_\varepsilon^2$ has to be given. It is always good to assume that $d = 1$ in many practical cases as pointed out Goos and Donev (2007). Still, it turns out that the generated designs may not be sensitive to a particular d value. Referring to another leads to the same designs.

However, Cho (2010) define the algorithm as a classical point exchange algorithm that requires user-specific candidate design points. A simple way to create a good candidate set when designing a split-plot for a compound process variable test is to, as Njoroge *et al.* (2017) points out, the response to a factorial design or process variable is to bypass the design of the MPV with the surface design. However, conditions involving unrestricted simplex-shaped composite design spaces can do this by passing SLDs or SCDs for the factorial arrangement of MPV by Cornell (2011) and others. Goos *et al.* (2016) suggested the use of fringe centroids and verticals to create better test designs in case of handling irregularly shaped mixture design region. Therefore, for examining and evaluating different design options in terms of G- and V- efficiency, the candidate points should also include interior points other than the overall centroid as reported by Njoroge *et al.* (2017). The simplex check points as described by Hassan et al. (2020) can be used as interior points for the case a simplex shaped region whereas for a constrained design region, pairwise averages of the overall centroid $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ and other points in the candidate set can be used as interior points.

However, to create the desired design, the algorithm begins with generation of a starting design with the specified number of whole plots, $n_{WP} = 9$ and whole plot sizes, $b_1 = 6, b_2 = 6, b_3 = 6, b_4 = 6, b_5 = 6, b_6 = 6, b_7 = 6, b_8 = 6, b_9 = 6$. Part of this is done at

random as stated by Sitinjak and Syafitri (2019). The initial design is completed by adding consecutive candidate points with the largest prediction variations. Therefore, the algorithm explores and evaluates all possible exchanges of design points and candidate points and possible transfers of design points across different whole-plots, as Goos and Vanderbroek (2003) stated. Better transfer or swap is done for each iteration. However, the search will stop when further improvement is not possible. Several startup designs was developed to increase the probability of finding the best overall D-optimized design.

Furthermore, the Candidate set free coordinate algorithm described in Jones and Goose (2007) was implemented allowing the creation of D-optimal SPDs in the absence of a candidate package. Except for the candidate set, the algorithm's input is similar to that required for Goos and Vandebrook's (2003) algorithm. Furthermore, Goose et al. (2016) showed that D-optimized designs could act as building blocks in the construction of new designs that will require duplication and additional points for the absence of fit tests in the presence of sample uncertainty.

D- Optimal Designs for Split-Plot Design

The set of candidate design points in Equation (3.1) and (3.2) were used as an input to the design construction method described in Sitinjak and Syafitri (2019) to determine the D-optimal design test. This set includes all combinations of all points of the SCD and two checkpoints for the four composite components, including the permutation of the binary compound (mixture) and the overall centroid point. On the one hand, the two process variable 2^2 factorial design arrangements with central composite design (CCD), plus the center point for the two process variables. We use Goos *et al.* (2016) algorithm for construction A1, A2, A3, A4, A5, and A6 because FORTRAN code was freely available

and could easily be modified to solve and handle nonstandard problems. In this design creation, we utilized d as discussed in section 3.2.2 since we needed the alluring plan to fit demonstrate (3.5) from the distinctive design. The alternative design was proposed since the relative D- and A- proficiency does not depend exceptionally much on the d esteem, but as it were, the relative G- and V- productivity (efficiency) diminish with d , and this concurring to discoveries detailed by Goos and Donev (2007). They also noted that D-optimal designs outperform the designs initially proposed by Kowalski *et al.* (2002). They called benchmark design in terms of the G and V optimality criterion with the value of d that ranges from 0.1 to 10.

Furthermore, during this design generation, we increased the center points in design A4 and A6 compared to the rest. According to the literature review, additional center points allow for extra other boundary points in the D- optimal designs that provide an opportunity to improve the efficiency of the methods substantially (Cho, 2010; Goos *et al.*, 2016). Since lack of center points in the optimal design is, however, criticized by several researchers (Hassan *et al.*, 2020, Njoroge *et al.*, 2017) and would probably cause the D- optimal designs to be biased, this is attributed to modifying different design options until found desirable. It is possible to construct designs that are substantially more efficient than those without or contain several center points. Design A1 to design A6 was also built using the candidate set free algorithm. We reported the D-, A-, G-, and I- efficiency together with a sliced FDS plot for each design relative to each other to select a desirable that supports and fits combined second-order MPV with CCD for split-plot layout structure.

3.2.3 Construction of SPD for Combined MPV with CCD Formulated

They were six design namely A_1, A_2, A_3, A_4, A_5 and A_6 extended from the initial model created by Njoroge *et al.* (2017) by considering the set of SCD design point at different settings of 2^2 factorial arrangement plus additional points of CCD in order to find the best MPV settings as discussed in section 3.1. We subjected designs to various optimality criterion and FDS plot techniques to select the best design. The design in Tables 3.1 to 3.6 were generated using the candidate set free algorithm based on the design proposed by Cornell (2011) Vinyl thickness experiment involving three mixture components and two process variables. But in this case, this design A1 involves four mixture components (x_1, x_2, x_3 and x_4) and two process variables. However, the data set for mixture components for the six different design options can also be generated by genetic algorithms in conjunction with process variables in a designed split-plot experiment as described in Cho (2010).

Table 3.1 shows the proposed design A_1 obtained using JMP at different combination mixture components at 2^2 factorial arrangement of process variable with CCD. We created the design using the D-optimal criteria discussed in section 3.2.2.1. A simplex centroid design was used in this design runs at both low and high levels of the remaining process variables since it allows for identifying component factors that are deemed unimportant.

Further, this design includes replicates of the center point ($Z_1 = Z_2 = 0, x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$) that can be used to compute pure error estimates. This design also includes replicates at the axial point ($\{Z_1 = 0, Z_2 = -1\}, \{Z_1 = 0, Z_2 = 1\}, \{Z_1 = -1.414, Z_2 = 0\}, \{Z_1 = 1.414, Z_2 = 0\}, x_1 = x_2 = x_3 = x_4 = 0.25$),

that makes the created design different form the one proposed by Cho (2010) and Njoroge *et al.* (2017).

Table 3.1: The proposed design A1 obtained using JMP version 15 at different combination mixture component at 2^2 factorial arrangement of process variable with CCD

Run	Whole plot	X_1	X_2	X_3	X_4	Z_1	Z_2
1	1	0.25	0.25	0.25	0.25	-1	1
2	1	0	0	0	1	-1	1
3	1	0	1	0	0	-1	1
4	1	0.25	0.25	0.25	0.25	-1	1
5	1	0	0	1	0	-1	1
6	1	1	0	0	0	-1	1
7	2	0.25	0.25	0.25	0.25	1	-1
8	2	0.25	0.25	0.25	0.25	1	-1
9	2	0	0	1	0	1	-1
10	2	0	1	0	0	1	-1
11	2	0	0	0	1	1	-1
12	2	1	0	0	0	1	-1
13	3	0.5	0.5	0	0	1	1
14	3	0.5	0	0.5	0	1	1
15	3	0.5	0	0	0.5	1	1
16	3	0	0.5	0	0.5	1	1
17	3	0	0.5	0.5	0	1	1
18	3	0	0	0.5	0.5	1	1
19	4	0	0.5	0.5	0	-1	-1
20	4	0.5	0.5	0	0	-1	-1
21	4	0	0	0.5	0.5	-1	-1
22	4	0.5	0	0.5		-1	-1
23	4	0.5	0	0	0.5	-1	-1
24	4	0.25	0.25	0.25	0.25	-1	-1
25	5	0.25	0.25	0.25	0.25	0	0
26	5	0.25	0.25	0.25	0.25	0	0

27	5	0.25	0.25	0.25	0.25	0	0
28	5	0.25	0.25	0.25	0.25	0	0
29	5	0.25	0.25	0.25	0.25	0	0
30	5	0.25	0.25	0.25	0.25	0	0
31	6	0.25	0.25	0.25	0.25	0	1
32	6	0.25	0.25	0.25	0.25	0	1
33	6	0.25	0.25	0.25	0.25	0	1
34	6	0.25	0.25	0.25	0.25	0	1
35	6	0.25	0.25	0.25	0.25	0	1
36	6	0.25	0.25	0.25	0.25	0	1
37	7	0.25	0.25	0.25	0.25	0	-1
38	7	0.25	0.25	0.25	0.25	0	-1
39	7	0.25	0.25	0.25	0.25	0	-1
40	7	0.25	0.25	0.25	0.25	0	-1
41	7	0.25	0.25	0.25	0.25	0	-1
42	7	0.25	0.25	0.25	0.25	0	-1
43	8	0.25	0.25	0.25	0.25	-1.414	0
44	8	0.25	0.25	0.25	0.25	-1.414	0
45	8	0.25	0.25	0.25	0.25	-1.414	0
46	8	0.25	0.25	0.25	0.25	-1.414	0
47	8	0.25	0.25	0.25	0.25	-1.414	0
48	8	0.25	0.25	0.25	0.25	-1.414	0
49	9	0.25	0.25	0.25	0.25	1.414	0
50	9	0.25	0.25	0.25	0.25	1.414	0
51	9	0.25	0.25	0.25	0.25	1.414	0
52	9	0.25	0.25	0.25	0.25	1.414	0
53	9	0.25	0.25	0.25	0.25	1.414	0
54	9	0.25	0.25	0.25	0.25	1.414	0

Table 3.2 shows the proposed design A_2 obtained using JMP at different combination mixture components at 2^2 factorial arrangement of process variable with CCD. We formulated the design using the D-optimal criteria. In this design A_2 , a SCD also runs at both low and high level of the remaining process variable as in the case of design A_1 in

Table 3.1. Further, this design includes the pure mixture blend and only two replicates of centroid point $(x_1 = x_2 = x_3 = x_4 = \frac{1}{4})$ at center point $(Z_1 = 0, Z_2 = 0)$ of design. The inclusion of four pure mixture components at the center point of the design is what distinguishes design A_1 from design A_2 .

Table 3.2: The proposed design A2 obtained using JMP version 15 at different combination mixture component at 2^2 factorial arrangement of process variable with CCD

Run	Whole plot	X_1	X_2	X_3	X_4	Z_1	Z_2
1	1	0.25	0.25	0.25	0.25	-1	1
2	1	0	0	0	1	-1	1
3	1	0	1	0	0	-1	1
4	1	0.25	0.25	0.25	0.25	-1	1
5	1	0	0	1	0	-1	1
6	1	1	0	0	0	-1	1
7	2	0.25	0.25	0.25	0.25	1	-1
8	2	0.25	0.25	0.25	0.25	1	-1
9	2	0	0	1	0	1	-1
10	2	0	1	0	0	1	-1
11	2	0	0	0	1	1	-1
12	2	1	0	0	0	1	-1
13	3	0.5	0.5	0	0	1	1
14	3	0.5	0	0.5	0	1	1
15	3	0.5	0	0	0.5	1	1
16	3	0	0.5	0	0.5	1	1
17	3	0	0.5	0.5	0	1	1
18	3	0	0	0.5	0.5	1	1
19	4	0	0.5	0.5	0	-1	-1
20	4	0.5	0.5	0	0	-1	-1
21	4	0	0	0.5	0.5	-1	-1
22	4	0.5	0	0.5		-1	-1
23	4	0.5	0	0	0.5	-1	-1

24	4	0.25	0.25	0.25	0.25	-1	-1
25	5	1	0	0	0	0	0
26	5	0	1	0	0	0	0
27	5	0	0	1	0	0	0
28	5	0	0	0	1	0	0
29	5	0.25	0.25	0.25	0.25	0	0
30	5	0.25	0.25	0.25	0.25	0	0
31	6	0.25	0.25	0.25	0.25	0	1
32	6	0.25	0.25	0.25	0.25	0	1
33	6	0.25	0.25	0.25	0.25	0	1
34	6	0.25	0.25	0.25	0.25	0	1
35	6	0.25	0.25	0.25	0.25	0	1
36	6	0.25	0.25	0.25	0.25	0	1
37	7	0.25	0.25	0.25	0.25	0	-1
38	7	0.25	0.25	0.25	0.25	0	-1
39	7	0.25	0.25	0.25	0.25	0	-1
40	7	0.25	0.25	0.25	0.25	0	-1
41	7	0.25	0.25	0.25	0.25	0	-1
42	7	0.25	0.25	0.25	0.25	0	-1
43	8	0.25	0.25	0.25	0.25	-1.414	0
44	8	0.25	0.25	0.25	0.25	-1.414	0
45	8	0.25	0.25	0.25	0.25	-1.414	0
46	8	0.25	0.25	0.25	0.25	-1.414	0
47	8	0.25	0.25	0.25	0.25	-1.414	0
48	8	0.25	0.25	0.25	0.25	-1.414	0
49	9	0.25	0.25	0.25	0.25	1.414	0
50	9	0.25	0.25	0.25	0.25	1.414	0
51	9	0.25	0.25	0.25	0.25	1.414	0
52	9	0.25	0.25	0.25	0.25	1.414	0
53	9	0.25	0.25	0.25	0.25	1.414	0
54	9	0.25	0.25	0.25	0.25	1.414	0

Table 3.3 shows the proposed design A_3 obtained using JMP, at different combination mixture components at 2^2 factorial arrangement of process variable with CCD. We also

created the design using the D-optimal criteria discussed in section 3.2.1. In this design A_3 , a simplex centroid design also runs at both low and high level of the remaining process variable as in the case of design A_1 and A_2 . This design consists of all set of combination of the eleven point of the SCD plus the four simplex checkpoints for the four mixture blends. On the other hand, they are 2^2 factorial design with CCD plus center point $(0, 0)$, axial point $((1,0), (0, 1), (-1, 0))$ for the two process variables that makes it different from design A_1 and A_2 .

Table 3.3: The proposed design A3 obtained using JMP version 15 at different combination mixture component at 2^2 factorial arrangement of process variable with CCD

Run	Whole plot	X_1	X_2	X_3	X_4	Z_1	Z_2
1	1	0.25	0.25	0.25	0.25	-1	1
2	1	0	0	0	1	-1	1
3	1	0	1	0	0	-1	1
4	1	0.25	0.25	0.25	0.25	-1	1
5	1	0	0	1	0	-1	1
6	1	1	0	0	0	-1	1
7	2	0.25	0.25	0.25	0.25	1	-1
8	2	0.25	0.25	0.25	0.25	1	-1
9	2	0	0	1	0	1	-1
10	2	0	1	0	0	1	-1
11	2	0	0	0	1	1	-1
12	2	1	0	0	0	1	-1
13	3	0.5	0.5	0	0	1	1
14	3	0.5	0	0.5	0	1	1
15	3	0.5	0	0	0.5	1	1
16	3	0	0.5	0	0.5	1	1
17	3	0	0.5	0.5	0	1	1

18	3	0	0	0.5	0.5	1	1
19	4	0	0.5	0.5	0	-1	-1
20	4	0.5	0.5	0	0	-1	-1
21	4	0	0	0.5	0.5	-1	-1
22	4	0.5	0	0.5		-1	-1
23	4	0.5	0	0	0.5	-1	-1
24	4	0.25	0.25	0.25	0.25	-1	-1
25	5	1	0	0	0	0	0
26	5	0	1	0	0	0	0
27	5	0	0	1	0	0	0
28	5	0	0	0	1	0	0
29	5	0.25	0.25	0.25	0.25	0	0
30	5	0.25	0.25	0.25	0.25	0	0
31	6	1	0	0	0	0	1
32	6	0	1	0	0	0	1
33	6	0	0	1	0	0	1
34	6	0	0	0	1	0	1
35	6	0.25	0.25	0.25	0.25	0	1
36	6	0.25	0.25	0.25	0.25	0	1
37	7	1	0	0	0	0	-1
38	7	0	1	0	0	0	-1
39	7	0	0	1	0	0	-1
40	7	0	0	0	1	0	-1
41	7	0.25	0.25	0.25	0.25	0	-1
42	7	0.25	0.25	0.25	0.25	0	-1
43	8	0.25	0.25	0.25	0.25	-1.414	0
44	8	0.25	0.25	0.25	0.25	-1.414	0
45	8	0.25	0.25	0.25	0.25	-1.414	0
46	8	0.25	0.25	0.25	0.25	-1.414	0
47	8	0.25	0.25	0.25	0.25	-1.414	0
48	8	0.25	0.25	0.25	0.25	-1.414	0
49	9	0.25	0.25	0.25	0.25	1.414	0
50	9	0.25	0.25	0.25	0.25	1.414	0
51	9	0.25	0.25	0.25	0.25	1.414	0
52	9	0.25	0.25	0.25	0.25	1.414	0

53	9	0.25	0.25	0.25	0.25	1.414	0
54	9	0.25	0.25	0.25	0.25	1.414	0

Table 3.4 shows the proposed design A_4 obtained using JMP version at different combination mixture components at 2^2 factorial arrangement of process variable with CCD. We developed created the design using the D-optimal criteria. In this design A_4 , a simplex centroid design also runs at both low and high level of the remaining process as in the case of design A_1 , A_2 , and A_3 . This design includes four pure mixture blends plus two replicates of $(x_1 = x_2 = x_3 = x_4 = \frac{1}{4})$ at center point $(Z_1 = 0, Z_2 = 0)$ of design, permutation of binary mixture $(0.5, 0.5, 0, 0)$ at axial point $((1,0), (0, 1), (0, -1))$ for the two process variables and additional runs of overall SCD $(0.25, 0.25, 0.25, 0.25)$ at axial point $((-1.414,0), (1.414, 0))$ for one of the process variable (Z_1) and this makes it different from design A_1 , A_2 , and A_3 .

Table 3.4: The proposed design A4 obtained using JMP version 15 at different combination mixture component at 2^2 factorial arrangement of process variable with CCD

Run	Whole plot	X_1	X_2	X_3	X_4	Z_1	Z_2
1	1	0.25	0.25	0.25	0.25	-1	1
2	1	0	0	0	1	-1	1
3	1	0	1	0	0	-1	1
4	1	0.25	0.25	0.25	0.25	-1	1
5	1	0	0	1	0	-1	1
6	1	1	0	0	0	-1	1
7	2	0.25	0.25	0.25	0.25	1	-1
8	2	0.25	0.25	0.25	0.25	1	-1
9	2	0	0	1	0	1	-1
10	2	0	1	0	0	1	-1
11	2	0	0	0	1	1	-1
12	2	1	0	0	0	1	-1
13	3	0.5	0.5	0	0	1	1
14	3	0.5	0	0.5	0	1	1
15	3	0.5	0	0	0.5	1	1
16	3	0	0.5	0	0.5	1	1
17	3	0	0.5	0.5	0	1	1
18	3	0	0	0.5	0.5	1	1
19	4	0	0.5	0.5	0	-1	-1
20	4	0.5	0.5	0	0	-1	-1
21	4	0	0	0.5	0.5	-1	-1
22	4	0.5	0	0.5		-1	-1
23	4	0.5	0	0	0.5	-1	-1
24	4	0.25	0.25	0.25	0.25	-1	-1
25	5	1	0	0	0	0	0
26	5	0	1	0	0	0	0
27	5	0	0	1	0	0	0

28	5	0	0	0	1	0	0
29	5	0.25	0.25	0.25	0.25	0	0
30	5	0.25	0.25	0.25	0.25	0	0
31	6	0.5	0	0.5	0	0	1
32	6	0.5	0.5	0	0	0	1
33	6	0.5	0	0	0.5	0	1
34	6	0	0.5	0.5	0	0	1
35	6	0	0.5	0	0.5	0	1
36	6	0	0	0.5	0.5	0	1
37	7	0.5	0	0.5	0	0	-1
38	7	0.5	0.5	0	0	0	-1
39	7	0.5	0	0	0.5	0	-1
40	7	0	0.5	0.5	0	0	-1
41	7	0	0.5	0	0.5	0	-1
42	7	0	0	0.5	0.5	0	-1
43	8	0.25	0.25	0.25	0.25	-1.414	0
44	8	0.25	0.25	0.25	0.25	-1.414	0
45	8	0.25	0.25	0.25	0.25	-1.414	0
46	8	0.25	0.25	0.25	0.25	-1.414	0
47	8	0.25	0.25	0.25	0.25	-1.414	0
48	8	0.25	0.25	0.25	0.25	-1.414	0
49	9	0.25	0.25	0.25	0.25	1.414	0
50	9	0.25	0.25	0.25	0.25	1.414	0
51	9	0.25	0.25	0.25	0.25	1.414	0
52	9	0.25	0.25	0.25	0.25	1.414	0
53	9	0.25	0.25	0.25	0.25	1.414	0
54	9	0.25	0.25	0.25	0.25	1.414	0

Table 3.5 shows the proposed design A_5 obtained using JMP version 15 at different combination mixture components at 2^2 factorial arrangement of process variable with CCD. We developed created the design using the D-optimal criteria discussed in chapter two. In this design A_5 , a simplex centroid design also runs at both low and high level of

the remaining process as in the case of design A_1, A_2, A_3 and A_4 . This design includes four pure mixture blends plus eight replicates of $(x_1 = x_2 = x_3 = x_4 = 0.25)$ at center point $(Z_1 = 0, Z_2 = 0)$ of design, permutation of binary mixture $(0.5, 0.5, 0, 0)$ at axial point $(0, 1)$ for the two process variables and additional runs of overall SCD $(0.25, 0.25, 0.25, 0.25)$ at axial point $((-1.414, 0), (1.414, 0))$ for one of the process variable (Z_1) which makes also different from the case of design A_1, A_2, A_3 and A_4 .

Table 3.5: The proposed design A5 obtained using JMP version 15 at different combination mixture component at 2^2 factorial arrangement of process variable with CCD

Run	Whole plot	X_1	X_2	X_3	X_4	Z_1	Z_2
1	1	0.25	0.25	0.25	0.25	-1	1
2	1	0	0	0	1	-1	1
3	1	0	1	0	0	-1	1
4	1	0.25	0.25	0.25	0.25	-1	1
5	1	0	0	1	0	-1	1
6	1	1	0	0	0	-1	1
7	2	0.25	0.25	0.25	0.25	1	-1
8	2	0.25	0.25	0.25	0.25	1	-1
9	2	0	0	1	0	1	-1
10	2	0	1	0	0	1	-1
11	2	0	0	0	1	1	-1
12	2	1	0	0	0	1	-1
13	3	0.5	0.5	0	0	1	1
14	3	0.5	0	0.5	0	1	1
15	3	0.5	0	0	0.5	1	1
16	3	0	0.5	0	0.5	1	1
17	3	0	0.5	0.5	0	1	1
18	3	0	0	0.5	0.5	1	1
19	4	0	0.5	0.5	0	-1	-1

20	4	0.5	0.5	0	0	-1	-1
21	4	0	0	0.5	0.5	-1	-1
22	4	0.5	0	0.5		-1	-1
23	4	0.5	0	0	0.5	-1	-1
24	4	0.25	0.25	0.25	0.25	-1	-1
25	5	1	0	0	0	0	0
26	5	0	1	0	0	0	0
27	5	0	0	1	0	0	0
28	5	0	0	0	1	0	0
29	5	0.25	0.25	0.25	0.25	0	0
30	5	0.25	0.25	0.25	0.25	0	0
31	6	0.5	0	0.5	0	0	1
32	6	0.5	0.5	0	0	0	1
33	6	0.5	0	0	0.5	0	1
34	6	0	0.5	0.5	0	0	1
35	6	0	0.5	0	0.5	0	1
36	6	0	0	0.5	0.5	0	1
37	7	0.25	0.25	0.25	0.25	0	0
38	7	0.25	0.25	0.25	0.25	0	0
39	7	0.25	0.25	0.25	0.25	0	0
40	7	0.25	0.25	0.25	0.25	0	0
41	7	0.25	0.25	0.25	0.25	0	0
42	7	0.25	0.25	0.25	0.25	0	0
43	8	0.25	0.25	0.25	0.25	-1.414	0
44	8	0.25	0.25	0.25	0.25	-1.414	0
45	8	0.25	0.25	0.25	0.25	-1.414	0
46	8	0.25	0.25	0.25	0.25	-1.414	0
47	8	0.25	0.25	0.25	0.25	-1.414	0
48	8	0.25	0.25	0.25	0.25	-1.414	0
49	9	0.25	0.25	0.25	0.25	1.414	0
50	9	0.25	0.25	0.25	0.25	1.414	0
51	9	0.25	0.25	0.25	0.25	1.414	0
52	9	0.25	0.25	0.25	0.25	1.414	0
53	9	0.25	0.25	0.25	0.25	1.414	0
54	9	0.25	0.25	0.25	0.25	1.414	0

Table 3.6: The proposed design A6 obtained using JMP version 15 at different combination mixture component at 2^2 factorial arrangement of process variable with CCD

Run	Whole plot	X_1	X_2	X_3	X_4	Z_1	Z_2
1	1	0.25	0.25	0.25	0.25	-1	1
2	1	0	0	0	1	-1	1
3	1	0	1	0	0	-1	1
4	1	0.25	0.25	0.25	0.25	-1	1
5	1	0	0	1	0	-1	1
6	1	1	0	0	0	-1	1
7	2	0.25	0.25	0.25	0.25	1	-1
8	2	0.25	0.25	0.25	0.25	1	-1
9	2	0	0	1	0	1	-1
10	2	0	1	0	0	1	-1
11	2	0	0	0	1	1	-1
12	2	1	0	0	0	1	-1
13	3	0.5	0.5	0	0	1	1
14	3	0.5	0	0.5	0	1	1
15	3	0.5	0	0	0.5	1	1
16	3	0	0.5	0	0.5	1	1
17	3	0	0.5	0.5	0	1	1
18	3	0	0	0.5	0.5	1	1
19	4	0	0.5	0.5	0	-1	-1
20	4	0.5	0.5	0	0	-1	-1
21	4	0	0	0.5	0.5	-1	-1
22	4	0.5	0	0.5		-1	-1
23	4	0.5	0	0	0.5	-1	-1
24	4	0.25	0.25	0.25	0.25	-1	-1
25	5	0.25	0.25	0.25	0.25	0	0

26	5	0.25	0.25	0.25	0.25	0	0
27	5	0.25	0.25	0.25	0.25	0	0
28	5	0.25	0.25	0.25	0.25	0	0
29	5	0.25	0.25	0.25	0.25	0	0
30	5	0.25	0.25	0.25	0.25	0	0
31	6	0.5	0	0.5	0	0	-1
32	6	0.5	0.5	0	0	0	-1
33	6	0.5	0	0	0.5	0	-1
34	6	0	0.5	0.5	0	0	-1
35	6	0	0.5	0	0.5	0	-1
36	6	0	0	0.5	0.5	0	-1
37	7	0.25	0.25	0.25	0.25	0	0
38	7	0.25	0.25	0.25	0.25	0	0
39	7	0.25	0.25	0.25	0.25	0	0
40	7	0.25	0.25	0.25	0.25	0	0
41	7	0.25	0.25	0.25	0.25	0	0
42	7	0.25	0.25	0.25	0.25	0	0
43	8	0.25	0.25	0.25	0.25	-1.414	0
44	8	0.25	0.25	0.25	0.25	-1.414	0
45	8	0.25	0.25	0.25	0.25	-1.414	0
46	8	0.25	0.25	0.25	0.25	-1.414	0
47	8	0.25	0.25	0.25	0.25	-1.414	0
48	8	0.25	0.25	0.25	0.25	-1.414	0
49	9	0.25	0.25	0.25	0.25	1.414	0
50	9	0.25	0.25	0.25	0.25	1.414	0
51	9	0.25	0.25	0.25	0.25	1.414	0
52	9	0.25	0.25	0.25	0.25	1.414	0
53	9	0.25	0.25	0.25	0.25	1.414	0
54	9	0.25	0.25	0.25	0.25	1.414	0

Table 3.6 shows the proposed design A_6 obtained using JMP at different combination mixture components at 2^2 factorial arrangement of process variable with CCD. We developed created the design using the D-optimal criteria. In this design A_6 , a simplex

centroid design also runs at both low and high level of the remaining process as in the case of design A_1, A_2, A_3, A_4 and A_5 . This design includes twelve replicates of $(x_1 = x_2 = x_3 = x_4 = 0.25)$ at center point $(Z_1 = 0, Z_2 = 0)$ of design, permutation of binary mixture $(0.5, 0.5, 0, 0)$ at star point $(0, -1)$ for the two process variables and additional runs of overall SCD $(0.25, 0.25, 0.25, 0.25)$ at axial point $((-1.414, 0), (1.414, 0))$ for one of the process variable (Z_1) as in the case of design A_1, A_2, A_3, A_4 and A_5 .

3.2.4 Evaluation of MPV Design with Split Plot Structure

The analysis of the MPV design experiment within SPD is addressed in this section. When selecting the appropriate design, FDS plots for an MVP design within an SPD are developed and demonstrated for visual examination and evaluation. Besides, sliced fraction design space plots demonstrate the effect of mixture and process variables on prediction variance over the experimental area.

3.2.4.1 Prediction Variance for MPVD with a Split Plot Structure

The predicted expected response at any location x_0 as described by Goldfarb *et al.* (2004) is given by

$$\zeta(x_0) = x_0' \hat{\beta}^* \quad (3.12)$$

where x_0 is the point of interest in the experimental region, $\hat{\beta}^*$ denotes the vector of fixed effects resulting from mixture process variable settings and $\zeta(x_0) = \hat{y}(x_0)$. Therefore, prediction variance at x_0 now given as

$$Var(\zeta(x_0)) = x_0' (X'V^{-1}X)^{-1}x_0 \quad (3.13)$$

Furthermore, Cho (2010) pointed out that when the design is completely randomized, the covariance matrix $V = \sigma^2 I$ is used because the best design for predicting variance is determined solely by the design space. Furthermore, because of the different sources of error in the SPD, the covariance matrix becomes more complex than the general form of V described by Cornell (2011). SPD prediction variance is a function of the variance component ratio given by whole plot space error variance and split plot space error variance, as well as the experimental region x .

We take prediction variance as an objective to examine and evaluate the design. The prediction variance is scaled by the variance observation error to make the quantity scale-free and, by design, size to penalize larger design. According to Liange et al. [15], the scaled predicted variance (SPV) for the split-plot structure is calculated by multiplying the prediction variance by the total number of runs, N , and then dividing by the observational error variance. As a result, the scaled prediction variance for SPDs is

$$SPV = \frac{Nx'_0 (X'V^{-1}X)^{-1}x_0}{\sigma_\delta^2 + \sigma_\varepsilon^2} = x'_0 (X'D^{-1}X)^{-1}x_0 \quad (3.14)$$

where $D = \text{diagonal} \{D_1, \dots, D_n\}$. D_i represents the correlation matrix of observations within plot i as a whole.

The size of the design in split-plot designs is not nearly related to the cost because the number of observations in SPDs is not the number of setups required to collect the data described by Cho (2010). The variance of the approximated means response divided by the variance of observational error ($\sigma_\delta^2 + \sigma_\varepsilon^2$) is modeled as given by

$$\text{Predicted Variance} = \frac{x_0' (X'V^{-1}X)^{-1}x_0}{\sigma_{\delta}^2 + \sigma_{\varepsilon}^2} = x_0' (X'D^{-1}X)^{-1}x_0, \quad (3.15)$$

Furthermore, in a split-plot design, unscaled variance is a valid alternative to scaled prediction variance, as reported by Cornell (2011).

3.2.4.2 Fraction Design Space Plots for MPV Design within SPD

Various methods involving prediction variance have been proposed in MPV Design to examine and evaluate a design's prediction performance. Zahran et al. (2003) proposed a fraction design space plot. When constructing the FDS plots, the scaled predicted variance is computed throughout the design space. Furthermore, the fraction of the experimental region that is less than or equal to a given SPV value is calculated. We develop FDS plots for mixture and mixture process variable designs. Design points are generated at random within the experimental region's constraints. The minimum scaled predicted variance is then plotted at a design space fraction of zero, while the maximum is plotted at a fraction of one. A desirable design starts with a small SPV and has a relatively flat slope on a cross-section of the fraction design space plot.

3.2.4.3 Sliced Fraction Design Space Plots for MPVD within SPD

In this section we discuss how sliced FDS plots were developed. We created Sliced fraction design space plots to examine the prediction variance distribution across the subplot region at different whole plot shrinkage levels. Furthermore, random points were generated throughout the subplot space at each whole plot shrinkage level from to one (in steps of 0.1) depending on the variance component ratio, d .

However, sliced FDS plots can also be constructed at different subplot shrinkage levels if one is interested in trends in the entire plot region for a given subplot shrinkage level. Furthermore, the sliced fraction design space plot provides the maximum scaled predicted variance value at each shrinkage level for the whole plot or split-plot used to build sliced FDS. It is preferable to keep the scaled predicted variance as small as possible. These sliced FDS also show how two regions, whole plot or split-plot, contributed to the change scaled predicted variance value.

The sliced FDS plot was employed with different whole plot shrinkage levels since the whole plot location has a more significant impact on the SPV value than the split-plot location. Typically, this occurs when the fraction design space "slices" are spaced far apart, even though changing the entire plot location is insufficient to affect the SPV value significantly. However, if the sliced fraction design space plots start from a similar minimum scaled predicted variance and are evenly distributed, changes in the SPV values balance the two regions' contribution. In other words, the length of the plots represents the effect of the entire plot region on the SPV values, whereas the slope of the fraction design space plot represents the effect of the split-plot space on the SPV values.

Furthermore, we analyzed the sliced FDS plot by constructing the FDS plot with a different variance component ratio, d , at the desired shrinkage level. These sliced fraction design spaces show SPV trends by changing the variance component ratio at a specific shrinkage of the whole plot or split-plot.

3.2.6 Evaluation of a Desirable Design for MPV within SPD

Using design criteria as discussed in Chapter two in this research is to find an appropriate experimental design that allows for efficient parameter vector estimation in the model 3.6.

We use the D- optimality criterion to find such a desirable design to fit the model, which seeks a design that minimizes the parameter estimates' generalized variance. Normally, the D- optimal criterion relies on ratio, $d = \frac{\sigma_\delta^2}{\sigma_\varepsilon^2}$, of the two observational variance components (whole plot error variance denoted by σ_δ^2 and split plot error variance represented by σ_ε^2) through covariance matrix V . To find the best appropriate design, we compare the alternative different design option (A_1, A_2, A_3, A_4, A_5 and A_6) in this research and report relative D-, A-, G-, I- or V- efficiency where A_1, A_2, A_3, A_4, A_5 and A_6 denotes the model matrices of six different designs option. We evaluate and compare SPD options based on D-, A-, G-, I- or V- optimality criterion performance. In this case, however, the A-optimal criterion seeks to reduce the mean-variance of the parameter estimates. On the other hand, as mentioned above, G-optimal design seeks to reduce forecast variability,

$$\text{Max}_{(Z, a) \in \chi} h'(Z, a)(X'V^{-1}X)^{-1}h(Z, a), \quad (3.16)$$

Over the region of interest χ where Z , and a represents the two process variables, and four mixture components (x_1, x_2, x_3, x_4), respectively. However, I- or V- optimal in this case minimizes the average forecast variance of that test region:

$$\text{avg}_{(Z, x) \in \chi} \text{Max} h'(Z, x)(X'V^{-1}X)^{-1} h(Z, x). \quad (3.17)$$

Therefore, we report A, G and V relative efficiency of six designs with model matrices A_1, A_2, A_3, A_4, A_5 and A_6 are then computed as

$$\frac{\text{trace}(A_n'V^{-1}A_n)^{-1}}{\text{trace}(A_{n-1}'V^{-1}A_{n-1})^{-1}}, \quad (3.18)$$

$$\frac{\text{Max}_{(Z, a) \in \chi} h'(Z, a)(A'_n V^{-1} A_n)^{-1} h(Z, a)}{\text{Max}_{(Z, a) \in \chi} h'(Z, a)(A'_{n-1} V^{-1} A_{n-1})^{-1} h(Z, a)}, \quad (3.19)$$

and

$$\frac{\text{avg}_{(Z, a) \in \chi} \text{Max}_{(Z, a) \in \chi} h'(Z, a)(A'_n V^{-1} A_n)^{-1} h(Z, a)}{\text{avg}_{(Z, a) \in \chi} \text{Max}_{(Z, a) \in \chi} h'(Z, a)(A'_{n-1} V^{-1} A_{n-1})^{-1} h(Z, a)}, \quad (3.20)$$

respectively, where $n = 1, 2, \dots, 6$. Furthermore, G- and I- efficiency are calculated by exploring and evaluating the predictive variance at the design space's point. However, for an accurate evaluation of the different options competing for test designs, the grid must often cover the boundaries of the test area and its interior, as described by Goose and Donev (2007).

Furthermore, the reported relative D-, A-, G-, V- or I-performance multiple values are defined to indicate progress in design with the sample matrix X_n . This is because the relative efficiency depends on the value of $d = \sigma_\delta^2 / \sigma_\varepsilon^2$. We, therefore, employed d values, $d = 0.5$, $d = 1.0$ and $d = 1.5$ to evaluate the different design options in this thesis where design A1 and A2 used $d = 0.5$, design A3 and A4 applied $d = 1.0$, and finally design A5 and A6 $d = 1.5$, but with modification of runs at axial point and center point of each design in order to make them unique and have clear distinction from each design created though all the six designs have some combination of mixture components that both runs at both low and high level of the remaining process variables.

Therefore, with these facts we report the relative efficiencies using d value 0.5, 1 and 1.5 in order to evaluate design option (A1, A2, A3, A4, A5, A6). The relative efficiencies of D-

, A-, G-, I- optimality criteria were computed using the formula described in Iwundu (2017) which also implemented in JMP software. We believe that SPDs often cause such small or large variance component ratings with few whole plot structure, and as a result, the worst estimate of the whole plot error variance. Therefore, for this reason, we have increased the number of total number of whole plot to nine compared to the seven whole plot used by Njoroge *et al.* (2017) when evaluating two different design options.

However, design A4 was found desirable in fitting model 3.5 based on the results reported on FDS plots and design optimality criteria discussed in Chapter four. We therefore, used design A4 in conjunction with model 3.6 in modeling the yield of *Glycine max* as discussed in section 3.3.

3.3 Employing the Modified MPV Model in Predicting the Yield of *Glycine Max* Reasonable Split-Plot And Main Plot Errors

This section outlines the steps involves in modeling the yield of *Glycine max* within an optimal split-plot design in the context MPV setting taking into account SCD and how two sources of errors that arise from SPD are evaluated.

3.3.1 Data Source

They were two sources data where one was primarily collected from the field of experiment in consideration with use of SPD in the context of MPV settings and the other was simulated data based on (4, 2) SCD with the split-plot structure experiments. The data consist of eight response measurements obtained from *Glycine max* (L.) Merrill after the soybean varieties were subjected to different types of MPV treatments subject to SPD layout developed MPV design. These eight responses measured include the number of

branches per plant (Y_1), number of pods on branches (Y_2), pods per branch (Y_3), pods on the main stem (Y_4), entire pods per plants (Y_5), number of seeds per plant (Y_6), seeds per pod (Y_7), and yield of seeds in grams per plant (Y_8). The mixture settings included four components x_1 , x_2 , x_3 , and x_4 , derived from different organic matter varieties, which represent goat manure, cow manure, chicken manure, and sheep manure, respectively. The process variable in the model were Z_1 and Z_2 where Z_1 the control is process variable (seeding rate (seeds per acre) at a constant row spacing of *Glycine max* seed) and Z_2 is the noise process variable (soil PH). The mixtures were the subplots and process variables the whole plots.

3.3.2 Description of Experimental Sites

The study was conducted in Spande and Munge's villages in Kakamega County, Mautuma Ward, Lugari Sub-District, and Western Kenya as shown in Figure 3.4. Both sites are about 8 km apart. The two regions lie between ($0.706373^{\circ} N, 35.0722^{\circ} E$) and ($0.695366^{\circ} N, 35.028022^{\circ} E$), with an elevation of between 1800 and 1900 m above sea level, respectively. The region receives bimodal rains with an annual mean precipitation of about 1971 mm, and an annual mean temperature of about $20.4^{\circ} C$, as reported by Althof (2005) and Mbau *et al.* (2015). Additionally, prolonged rain usually occurs between April and July, while short precipitation occurs between August and December, as described by Mbau *et al.* (2015). Further, the reliability growth period for *Glycine max* (L.) Merrill lies between 75 and 140 days (Jaetzold *et al.*, 2005).

Further, as Isaev *et al.* (2020), the best period for sowing *Glycine max* is when the temperature in the 0 – 10 cm layer of soil is about $12 - 14^{\circ} C$. According to Tsikhungu (2016), the Lugari sub-county grounds are predominantly well-drained deep red to dark,

sandy loams to sandy clays that are not very fertile. Still, well-drained soils, with moderately to slightly condition with soil PH, lie between 5.3 to 5.9. However, some part of this region contains low inherent fertility as evidenced by low amounts of Nitrogen, soil organic carbon and exchangeable base as described by Mbau *et al.* (2015). The experimental site encompasses farmlands adjacent to the Lugari forest. The area was initially inhabited by a sparse population of former forest residence communities who practiced shifting cultivation, hunting, and gathering. The study sites have a settlement history dating more than a hundred years with relatively intensive sedentary mixed subsistence agriculture as reported by Kimetu *et al.* (2006) for over the last sixty years. Pender *et al.* (2006) found that landholding per household has reduced drastically because of the high population growth rate and immigration into the area.

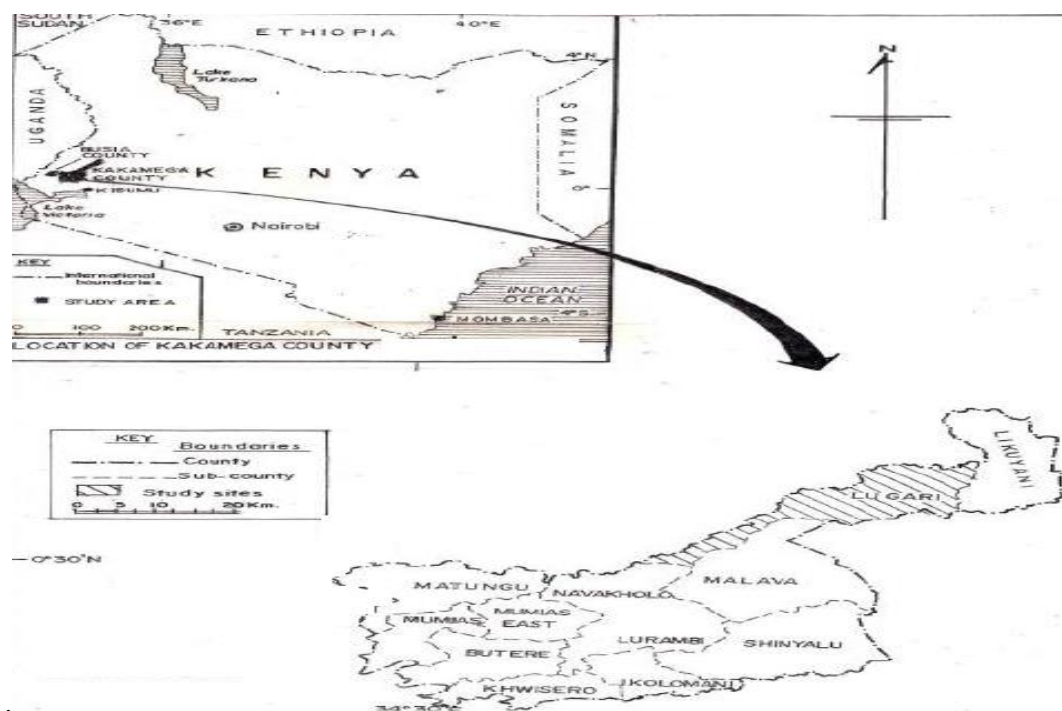


Figure 3.4: The map of Lugari Sub County in Kakamega County where the experiment was carried out (Tsikhungu, 2016).

Currently, most agricultural land is characterized by low soil fertility, low crop yields, and low farm income (Mbau *et al.*, 2015). However, cereals (maize), legumes (beans), and sugarcane have become the primary crops, with most fields described by Tsikhungu (2016).

3.3.3 Composting Farmyard Manure through the Framework of Mixture Design

Four agro-organic wastes commonly found in the test sites were selected for our study. The selection criteria for the required farm manure (FYM) obtained from livestock were based on the region's availability of material. FYM includes goat manure, chicken manure, sheep manure, and cow manure derived from the Spande farm. Compound composting was done using the pit method using the structure of the mixture design under control

$$x_1 + x_2 + x_3 + x_4 = 1$$

in reference to Figure 3.3. They were 11 composite pit sites as shown in Figure 3.5, 3.6, 3.7, 3.8, 3.9, and 3.10. The first four pit sites comprised only one major organic waste applied at permutation of (0,1,0,0) from the mixture components either x_1, x_2, x_3 or x_4 while the rest of materials aiding for formation of pit remaining unchanged across all the 11 composite pit sites. Secondly, the 6 other pit sites formed at permutation of binary mixture of (0.5, 0.5, 0, 0) of x_1, x_2, x_3 and x_4 with other material used remaining constant. Finally, the composite pit in site 11 as shown in Figure 3.10 generated using the overall simplex centroid design (SCD) of (0.25, 0.25, 0.25, 0.25). The total organic manure from the four major components applied in each 11 compost pit sites was 100 Kg ($x_1 + x_2 + x_3 + x_4 = 100$). Therefore, this implies that the first four pit consists 100 % of either x_1, x_2, x_3 or x_4 which means that if $x_4 = 100$ Kg, then $x_1 = x_2 = x_3 = 0$. Secondly, the other six pits in Figure 24, 25, and 26 implies that each pit site constitute two the

permutation of the major organic waste of either x_1, x_2, x_3 or x_4 at each 50 %, which therefore, means if $x_4 = 50 \text{ Kg}$, $x_1 = 50 \text{ Kg}$, then $x_2 = x_3 = 0$. Lastly, the 11 pit sites imply that all the four components are applied at each 25 % of the total manure for the ingredients present and in this case is ($x_1 = x_2 = x_3 = x_4 = 25 \text{ Kg}$).

According to the literature, we used the pit method (Agromisa et al., 2005, Mbau et al., 2008). Further, 11 pits, each measuring $3 \times 3 \times 2 \text{ m}$, were dug in the ground with a garden shovel in preparation for composting. Lining the hole was made using masonry lining and packing in the sides and bottom of the hole to avoid the sides crumbling down into the pit. We first shredded large and bulky dry materials applied at equal proportion in all the 11 composite pit sites to smaller pieces about 3 cm long to enhance decomposition rate. Layering within each of the eleven composite pits was done to a thickness of 20 cm. Ingredients were stacked in a mixed pile starting with dry grass, followed by a layer of vegetable waste and a layer of organic manure such as soil, goat, poultry, and sheep, all mixed with dirt and ash at the above relative ratio. Ash, in this case, helps to maintain the neutrality of the compost by decomposing the material in the compost pile, which may become somewhat acidic, so wood ash can help compensate for this as it is more alkaline. On the other hand, the soil was added to a decomposed compost pile to help the pile break down faster as it is rich in microbial activity. For example, a teaspoon may contain 100 million bacteria and 400 to 800 feet of fungal thread, so adding soil also stimulates microbes to speed up the process and help keep insects at bay while the fruit is flying. Then flies can become a problem in a compost pile, especially in late spring and summer (Mbau et al., 2008). The compost pile was created basing on Jeavon's (2001) suggestion, who proposed a compost pile with 45 % dry material, 45 % green material, and 10 % soil.

However, three liters of water were sprinkled at the top of each pit to moisten the materials since composite need moisture between 40-60 % to multiply the beneficial microbes within the pile. This process was repeated for each of the eleven pits till there were about ten layers. Furthermore, dry grass was used to cover each of the 11 pits to ensure proper aeration and reduce moisture loss. The heaps for FYM were composited for not less than eight weeks. Turning decaying material in each of the 11 pits was done after every two weeks using a shovel to enhance and facilitate even decomposition. Approximately 10 liters of water were sprinkled during the turning of each cavity to keep the pile moist. These compost pits were then cured for another period of four weeks and sieved using 5 mm in readiness for planting and did this after all the composts become brown and crumbly with an earthy smell, which implies that the composite manure was ready to plant.

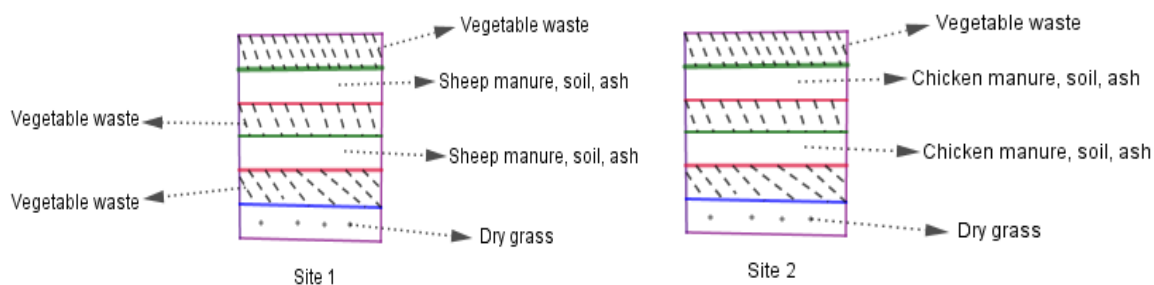


Figure 3.5: Only Sheep and Chicken manure considered for site 1 and 2, respectively

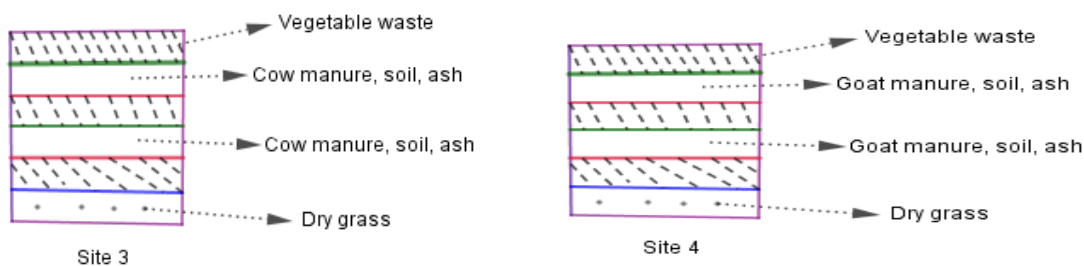


Figure 3.6: Only Cow manure and Goat manure considered for site 3 and 4, respectively

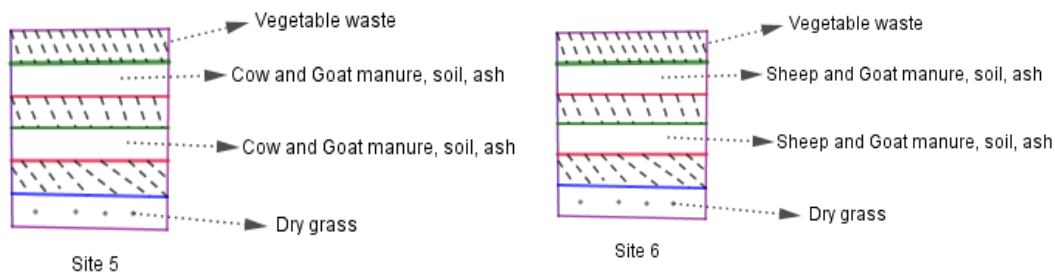


Figure 3.7: Both cow and Goat, Sheep and Goat manure were considered for site 4 and 6, respectively

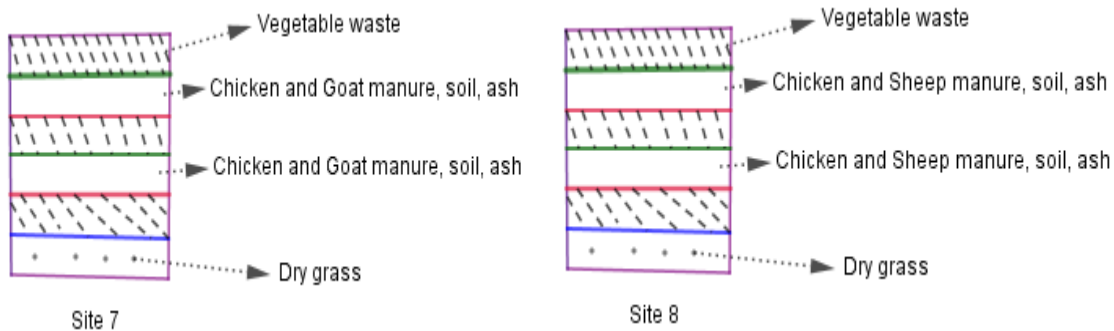


Figure 3.8: Both chicken and Goat manure considered in this case

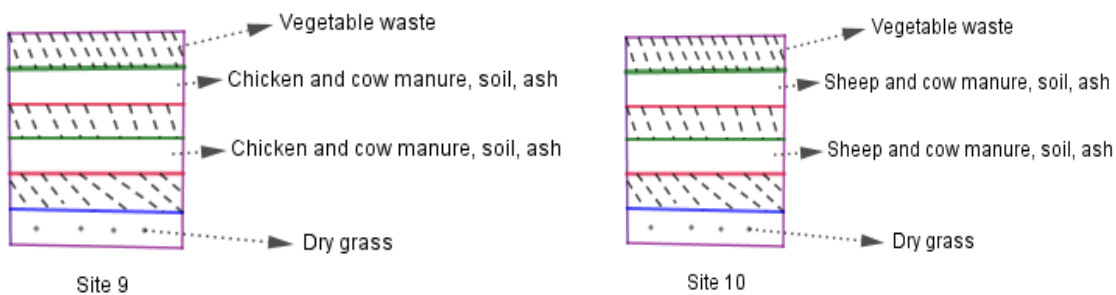


Figure 3.9: both chicken and Cow, Sheep and Cow manure were considered for site 9 and 10, respectively

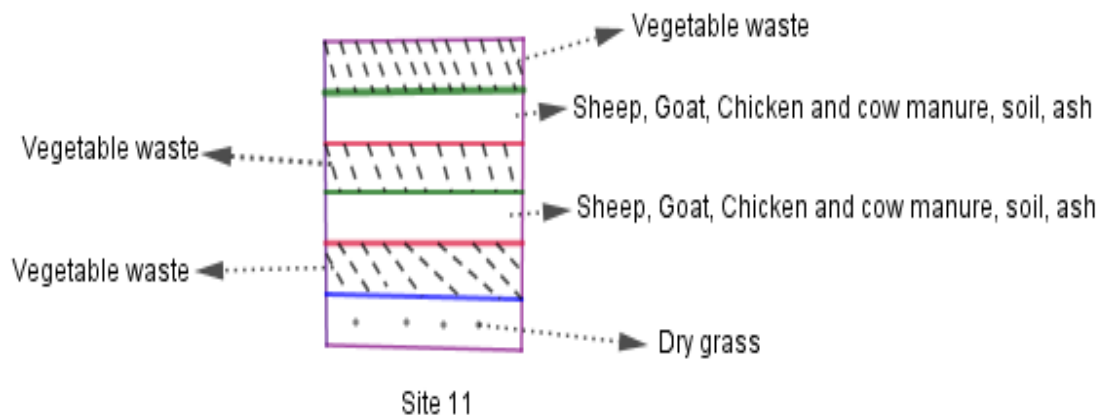


Figure 3.10: All the four-manure component considered

3.3.4 Treatment combinations Manure through of MPV within SPD Context

Field trials were performed on two farms. The experiment was carried out using a well randomized complete block in a split-plot arrangement with replication, as shown in Figure 3.3. The split-plot structure comprised nine whole plots, with each field having six subplot treatments. Each plot's plot size was $95.5 \text{ ft} \times 170 \text{ ft}$, while each experimental subplot unit was $15.5 \text{ ft} \times 50 \text{ ft}$. Split plot treatments were applied based on the proposed design in Figure 3.2 using composite compost manure from the 11 compost pits. They were four lime treatments (0, 1.7, 5, and 15 – ton *aglime /acre*) being applied to 9 main plots with at least twice at axial part as shown below with correspondence of soil pH obtained after the application.

Table 3.7: The scaled seeding rate (Z_1) and soil PH (Z_2) according to 2^2 factorial design with CCD

Whole plot	Lime application (tons/ acre)	Row spacing (inches)	Un coded Z_1	Un coded Z_2	Coded Z_1	Coded Z_2
1	15	25	125000	7.0	-1	1
2	1.7	15	225000	6.0	1	-1
3	15	15	225000	7.0	1	1
4	1.7	25	125000	6.0	-1	-1
5	5	20	175000	6.5	0	0
6	15	20	175000	7.0	0	1
7	1.7	20	175000	6.0	0	-1
8	0	30	100000	5.4	-1.414	0
9	0	10	275000	5.4	1.414	0

After initial testing of soil pH at the farm was 5.4, we prepared five different soils for selected plant growth at optimal pH values, as shown in Table 3.7, as shown in Figure 3.11 of the guided soil pH chart. PH chosen deals from the initial 5.4 pH of the soil using a control method according to the literature (Brown et al., 2008; De Bruin & Peterson, 2008; Thompson *et al.*, 2016):

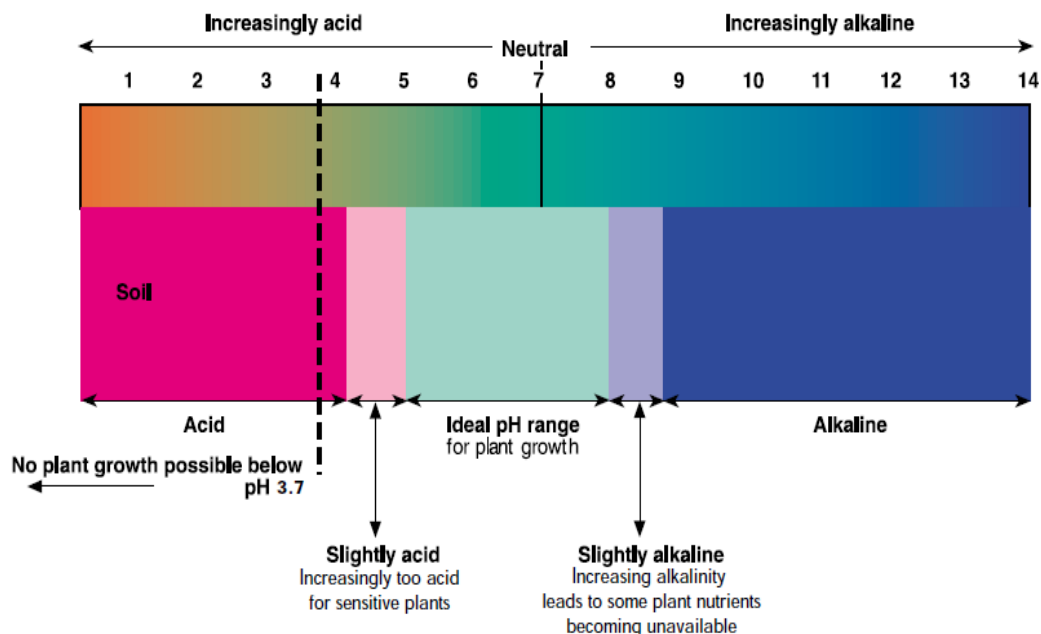


Figure 3.11: The soil pH chart indicating the ideal PH range for plant growth (Acid Soil Management, 2000)

The whole plots consisted of five primary seeding rates (100000, 125000, 175000, 225000, and 275 000 seeds per acre) applied to sub-plot experimental units taking into account row spacing whole plot as shown in Table 3.7. According to the literature review, we used the seeding rate and row spacing (De Bruin and Pedersen, 2008). The *Glycine max* varieties were R 184, and Blyvoor was planted on April 24th, 2020. The seeds were first inoculated with *Bradyrhizobium Japonicum*, and each subplot was grown using row spacing specified in Table 3.7 and 1- inch depth. Plots were harvested on August 27th, 2020. Grain yields obtained were then adjusted to 13 % moisture. .

3.3.5 Soil Sample Analysis

We randomly picked soil samples from four points on each of the two farms: Spande and Munge using soil augers to a depth of 20 cm. The two farms were labeled Farm A and B,

respectively, with one reference GPS, co-ordinate N 0°42'20.064" E 35° 0'4.74", defining only Munge village Mautuma ward, Lugari sub-county, Kakamega County. The soil picked was mixed correctly for each site. Two composite samples were taken for laboratory analysis at Kenya Agricultural and Livestock Research Organization (KALRO) located at Alupe, Busia County, and Western Kenya. Soil sample analysis from the two experimental sites gave the following result shown in Table 3.8. The pH of Soil at Spande and Munge Farm is 5.4 and 6.0, respectively. The PH of 6.0 indicates that the soil at Munge is weakly acidic and is within the optimum ranges. In contrast, a pH of 5.4 for Spande means that the ground is at the critical value for acidic soil and is within the optimum ranges for pH values for most crops to grow since most crop nutrients are available between pH of 5.4 and 7.0. Also, 0.1 % N means the soil is deficient in the nutrient since the deficiency level is any value below 0.2% at Munge (Mbau *et al.*, 2008). On the other hand, 9.33 – 10.1 ppm indicates that phosphorous is deficient at the site since minimum level 30 ppm. 1.1 – 1.3 me % for potassium means the nutrient is adequate since any value below 0.24 is deficient. 2.2 me % for Calcium means the nutrient is adequate since any values below 2.0 for Calcium are deficient. Therefore, the Calcium nutrient in Spande Farm is deficient. 6.2 me % for Magnesium means Mg is adequate. Any values below 1.0 me % mean inadequacy. 0.2 % for Organic carbon level means the soil has deficient organic carbon. Any values below 1.33 are low. Therefore, according to the literature: Liming is recommended since both magnesium and calcium levels are below the critical levels needed. We should apply the mixture of Agricultural Calcitic and dolomitic lime at the rate of 300 kg per acre. Further, phosphate fertilizers like D.A.P during planting while

nitrogenous fertilizers like C.A.N should be used for topdressing to introduce nitrogen to the soil.

Table 3.8: Soil analysis from the two sites (KALRO June 17th 2020, Maddo, Lab. No. 556/20 (A), 557/20 (B))

Study site	PH (1:2.5)	N (%)	P (ppm)	K (me %)	Ca (me %)	Mg (me %)	Organic Carbon	% Sand	% Silt	% Clay	ST
Spande farm A	5.4	0.5	9.3	1.1	1.2	0.2	0.4	5	57	38	SL
Munge farm B	6.0	0.1	10.1	1.3	2.2	6.2	0.2	4	42	54	SL

*SL= Silty Loam, ST= Soil type

Lastly, we should apply farmyard manure, green manure, compost, and green manure cover crops to raise the organic matter to the soil (Mbau *et al.*, 2008; Wanyama, 2013). Therefore, the development of design A4 is shown in Table 3.4 and Figure 3.3 based on these two farms' soil sample analyses

3.3.6 Exploration and Estimation of Parameters for MPV within Split Plot Layout

This section outlines how the parameters in model (3.7) can be estimated using Ordinary Least Square (OLS), Maximum likelihood (ML), Restricted Maximum likelihood, Bayesian method and Expectation Maximization Algorithm (EMA) thus enabling to evaluate the two sources errors (WPE and SPE) arising from split-plot design.

3.3.6.1 Maximum Likelihood and Ordinary Least Square Method

Estimation of the two parameters in model 3.7 using OLS and ML is based on following two cases

Case 1: Known covariance (Σ) for estimation of θ and δ

(i) Known covariance (Σ) for estimation of θ , the parameter can be obtained as

$$\theta = (X'V^{-1}X)^{-1}X'V^{-1}Y, \quad (3.21)$$

$$\hat{\theta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y, \quad (3.22)$$

(ii) Known covariance (Σ) for estimation of δ , the parameter δ can be obtained as

$$Y \sim N_n(X\theta, V), \quad \delta \sim N_{bq}(0, Z), \quad (3.23)$$

where

$$\begin{aligned} \text{Cov}(Y, \delta) &= \text{Cov}(X\theta + D\delta + \varepsilon, \delta), \\ &= \text{Cov}(X\theta, \delta) + \text{Cov}(D\delta, \delta) + \text{Cov}(\varepsilon, \delta) \\ &= \text{Cov}(D\delta, \delta), \\ &= DCov(\delta, \delta) = DVar(\delta), \\ &= DZ \text{ Since } Var(\delta) = Z. \end{aligned}$$

Therefore,
$$\begin{pmatrix} Y \\ \delta \end{pmatrix} \sim N_{bq+n} \left(\begin{pmatrix} X\theta \\ 0 \end{pmatrix}, \begin{bmatrix} V & DZ \\ ZD' & Z \end{bmatrix} \right) \quad (3.24)$$

The conditional expectation of $\delta|Y$ is shown to be $E(\delta|Y) = ZD'V^{-1}(Y - X\theta)$ the best linear unbiased predictor (BLUP) of δ . Therefore, the empirical best linear unbiased predictor (EBLUP) estimator of $\hat{\delta}$ can be shown to be

$$\hat{\delta} = ZD'V^{-1}(Y - X\hat{\theta}) \quad (3.25)$$

Now for the case of known covariance (Σ), then EBLUP of $\hat{\delta}$ is given as

$$\hat{\delta} = ZD'\Sigma^{-1}(Y - X\hat{\theta}) \quad (3.26)$$

Case 2: Unknown Covariance (Σ) for Estimation of δ and θ .

For estimation of δ and θ when Σ is unknown we employ the joint optimum maximization of log likelihood of (Y', δ) with respect to (θ, δ) . We let

$$\begin{aligned} h(Y, \delta) &= h(Y|\delta) \cdot h(\delta) \\ &\propto \exp\left\{-\frac{1}{2}(Y - X\theta - D\delta)'V^{-1}(Y - X\theta - D\delta)\right\} \exp\left\{-\frac{1}{2}\delta'Z^{-1}\delta\right\} \end{aligned} \quad (3.27)$$

Taking natural logarithm on both sides of Equation (3.39) we obtain

$$\begin{aligned} \ln(Y, \delta) &= \mathcal{L}(\theta, \delta), \\ \mathcal{L}(\theta, \delta) &= -\frac{1}{2}(Y - X\theta - D\delta)'V^{-1}(Y - X\theta - D\delta) - \frac{1}{2}\delta'Z^{-1}\delta + \vartheta \end{aligned} \quad (3.28)$$

Where ϑ indicates a constant independent of (θ, δ) and $-\frac{1}{2}\delta'Z^{-1}\delta$ represent the penalty term of δ . For estimation θ and δ using ML we take

$$\frac{\partial}{\partial \theta} \mathcal{L}(\theta, \delta) = 0,$$

$$\frac{\partial}{\partial \theta} \mathcal{L}(\theta, \delta) = -2X'U^{-1}Y + 2X'U^{-1}D\delta + 2X'U^{-1}X\theta,$$

$$\frac{\partial}{\partial \delta} \mathcal{L}(\theta, \delta) = 2\theta'XU^{-1}D - 2D'U^{-1}Y + 2D'U^{-1}D\delta + 2Z^{-1}\delta,$$

$$X'U^{-1}Y = X'U^{-1}X\theta + X'U^{-1}D\delta,$$

$$D'U^{-1}Y = \theta'XU^{-1}D + (D'U^{-1}D + Z^{-1})\delta,$$

For further simplification we obtain

$$\begin{aligned} \begin{bmatrix} X'U^{-1}X & X'U^{-1}D \\ D'U^{-1}X & D'U^{-1}D + Z^{-1} \end{bmatrix} \begin{bmatrix} \theta \\ \delta \end{bmatrix} &= \begin{bmatrix} X'U^{-1}Y \\ D'U^{-1}Y \end{bmatrix}, \\ \begin{pmatrix} \hat{\theta} \\ \hat{\delta} \end{pmatrix} &= \begin{bmatrix} X'U^{-1}X & X'U^{-1}D \\ D'U^{-1}X & D'U^{-1}D + Z^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'U^{-1}Y \\ D'U^{-1}Y \end{bmatrix}, \end{aligned} \quad (3.29)$$

Solving Equation 3.29 using sage Math 9.0 software, we find the ML estimates of $\hat{\delta}$ and $\hat{\theta}$ to be

$$\hat{\delta} = ZD'\hat{V}^{-1}(Y - X\hat{\theta}) \quad (3.30)$$

$$\hat{\theta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y \quad (3.31)$$

3.3.6.2 Estimation of θ and δ Using REML Method

REML method is always employed whenever estimating parameters subject to unknown covariance structure. This method is usually preferred to ML by most researchers because the estimates parameters obtained is unbiased (Goldfarb *et al.* 2003; Kowalski *et al.*, 2002; Njoroge *et al.*, 2017). To apply this method, we consider marginal model of Equation (3.8) with $V = DZD' + U$ under assumption that Z and U are both known to the variance parameter φ . For instance, we write Equation 3.10 in terms φ as

$$V(\varphi) = DZ(\varphi)D' + U(\varphi) \quad (3.32)$$

We now find maximum log likelihood for profile log likelihood given as (θ, φ) as

$$L(\theta, \varphi) = -\frac{1}{2}\{\ln|V(\varphi)| + (Y - X\theta)'V(\varphi)^{-1}(Y - X\theta)\} + \gamma \quad (3.33)$$

where γ is a constant independent of (θ, φ) . If we take

$$\frac{\partial}{\partial \theta} L(\theta, \varphi) = 0, \text{ we obtain}$$

$$\hat{\theta}(\varphi) = (X'V(\varphi)^{-1}X)^{-1}X'V(\varphi)^{-1}Y. \quad (3.34)$$

But this estimate $\hat{\theta}_{ML}$ obtained is biased and therefore REML is required for that case. We now employ REML method on Equation (3.34) by taking natural logarithms on both sides

$$L_{REML}(\varphi) = \ln \left[\int L(\theta, \varphi) d\theta \right] \quad (3.35)$$

But
$$\int L(\theta, \varphi) d\theta = \int (2\pi)^{\frac{-n}{2}} |V(\varphi)|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(Y - X\theta)'V(\varphi)^{-1}(Y - X\theta)\right\} d\theta, \quad (3.36)$$

For further simplification we consider the following cases

$$\begin{aligned} & (Y - X\theta)'V(\varphi)^{-1}(Y - X\theta) \\ &= \theta'X'V(\varphi)^{-1}X\theta - 2Y'V(\varphi)^{-1}X\theta + Y'V(\varphi)^{-1}Y, \\ & C_1(\varphi) = X'V(\varphi)^{-1}X, \\ & C_2(\varphi) = C_1(\varphi)^{-1}X'V(\varphi)^{-1}, \\ & C_2(\varphi)'C_1(\varphi) = V(\varphi)^{-1}XC_1(\varphi)^{-1}C_1(\varphi) = V(\varphi)^{-1}X, \end{aligned}$$

Therefore, using these four cases we have

$$\begin{aligned} \int L(\theta, \varphi) d\theta &= (2\pi)^{\frac{-n}{2}} |V(\varphi)|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(Y'[V(\varphi)^{-1} \right. \\ &+ C_2(\varphi)'C_1(\varphi)C_2(\varphi)]Y)\right\} \cdot \int \exp\left\{-\frac{1}{2}(\theta \right. \\ &\left. - C_2(\varphi)Y)'C_1(\varphi)(\theta - C_2(\varphi)Y)\right\} d\theta, \end{aligned} \quad (3.37)$$

but

$$\int \exp\left\{-\frac{1}{2}(\theta - C_2(\varphi)Y)'C_1(\varphi)(\theta - C_2(\varphi)Y)\right\} d\theta = (2\pi)^{\frac{n}{2}} |C_1(\varphi)^{-1}|^{-1}.$$

$$\begin{aligned} \text{Using } \hat{\theta}(\varphi) &= (X'V(\varphi)^{-1}X)^{-1}X'V(\varphi)^{-1}Y = \\ & C_1(\varphi)^{-1}XV(\varphi)^{-1} = C_2(\varphi)Y \end{aligned}$$

$$\begin{aligned} \text{and } C_2(\varphi)'C_1(\varphi)C_2(\varphi) &= V(\varphi)^{-1}XC_1(\varphi)^{-1}C_1(\varphi)C_2(\varphi) = \\ & V(\varphi)^{-1}XC_2(\varphi) \end{aligned}$$

as the intermediate variable. We can therefore rewrite 3.37 as

$$\begin{aligned}
\int L(\theta, \varphi) d\theta &= (2\pi)^{\frac{-n}{2}} |V(\varphi)|^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} (Y - X\theta)' V(\varphi)^{-1} (Y \right. \\
&\quad \left. - X\theta) \right\} (2\pi)^{\frac{n}{2}} |C_1(\varphi)^{-1}|^{\frac{1}{2}}, \\
\Rightarrow L_{REML}(\varphi) &= \ln \left[\int \mathcal{L}(\theta, \varphi) d\theta \right] \tag{3.38} \\
&= -\frac{1}{2} \left\{ \ln |V(\varphi)| + (Y - X\hat{\theta}(\varphi))' V(\varphi)^{-1} (Y - X\hat{\theta}(\varphi)) \right\} \\
&\quad - \frac{1}{2} \ln |C_1(\varphi)| + C = L_P(\varphi) - \frac{1}{2} \ln |C_1(\varphi)| + C,
\end{aligned}$$

Therefore, $\hat{\varphi}_{REML}$ is given as $L_{REML}(\varphi) = L_P(\varphi) - \frac{1}{2} \ln |C_1(\varphi)| + C$ which maximizes

$$L_{REML}(\varphi) = L_P(\varphi) - \frac{1}{2} \ln |X'V(\varphi)^{-1}X|.$$

In conjunction with ML and REML as method for estimating fixed effect and random effect parameters based on unknown covariance structure. We can be able to claim and make a summary from

$V(\varphi) = DZ(\varphi)D' + U(\varphi)$ that the covariance parameter vector φ can be obtained by either $\hat{\varphi}_{ML}$ which maximizes the profile log likelihood given as

$$L(\theta, \varphi) = -\frac{1}{2} \left\{ \ln |V(\varphi)| + (Y - X\hat{\theta})' V(\varphi)^{-1} (Y - X\hat{\theta}) \right\} \tag{3.39}$$

Where $\hat{\theta}(\varphi) = (X'V(\varphi)^{-1}X)^{-1}X'V(\varphi)^{-1}Y$ or by $\hat{\varphi}_{REML}$ that maximizes Equation (3.39).

Therefore, the random vector effect ($\hat{\delta}$) and fixed vector ($\hat{\theta}$) effect is estimated by

$$\hat{\delta} = ZD'\hat{V}^{-1}(Y - X\hat{\theta}). \quad (3.40)$$

$$\hat{\theta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y. \quad (3.41)$$

where $\hat{V} = V(\varphi_{ML})$ or $\hat{V} = V(\varphi_{REML})$.

3.3.6.3 Estimation Whole Plot Random Effect Using Bayesian Perspective Approach

The random effect δ_j in Equation (3.6) in split plot experiment that is associated with whole plot can also be estimated using Bayesian approach. However, δ_j in this case is regarded to be known with probability distribution referred as prior distribution while θ and subplot error (ε) are fixed and known. This is because θ and ε are not considered as random quantities with suitable prior distribution from the perspective classical Bayesian approach but they are taken as fixed and known. For estimation purposes for whole plot random effect using this approach we let the prior distribution of $\delta_j \sim N(0, \mathbf{Z})$ be $h(\delta, z)$. In addition, we define the likelihood function which is assumed to be conditional normal distribution and basing on Bayes theorem as $h(y_j | x_j, \delta_j, \theta, \varepsilon(U))$ where $y_j | x_j, \delta_j \sim N(\mathbf{X}\theta + \mathbf{D}\delta_j, U)$. Therefore, we can now define the posterior probability density function of δ subject on observing the response variable $Y = y_j$ as

$$h(\delta_j | y_j, \theta, \varepsilon(U), \mathbf{Z}) = \frac{h(y_j | x_j, \delta_j, \theta, \varepsilon(U)) h(\delta_j, z)}{h(y_j | x_j, \theta, \mathbf{Z}, \varepsilon(U))} \quad (3.42)$$

for all $j = 1, 2, \dots, b$ where b represent the number of whole plots in split plot experimental layout structure and

$$h(y_j | x_j, \theta, \mathbf{Z}, \varepsilon(U)) = \int h(y_j | x_j, \delta_j, \theta, \varepsilon(U)) h(\delta_j, z) d\delta_j$$

The Equation 3.324 which is referred to as posterior distribution is usually distributed with mean

$$ZD'\hat{V}(y - X\theta) \quad (3.43)$$

However, the mean and mode are always equivalent under standardized normal distribution and often maximizes the posterior density. Therefore, when the estimates $\hat{\theta}$ and \hat{V} are substitute in 3.43 yields Empirical Bayes estimator for δ as

$$\hat{\delta} = \hat{Z}D'\hat{V}^{-1}(y - X\hat{\theta}) \quad (3.44)$$

this estimate in 3.44 using the Bayesian approach can used to obtain the random effect associated with whole plot whenever there is prior knowledge of δ_j while θ and $\varepsilon(U)$ is known and fixed. However, θ can also be estimated the same way as δ_j by using similar augment by basing on Bayes theorem but in this case θ is treated as random vector independent of δ_j (Davidian and Giltinan, 1995).

3.3.6.4 Estimation of Parameters Using EMA Method

The Expectation Maximization Algorithm (EMA) is often applied whenever missing observation exist. Therefore, most researchers prefer using this method over other method whenever ML and REML estimates are required since it usually follow the analogy to a missing data problem from viewing full data. This is usually done by maximizing ML or REML objective function. Here, we discuss how $\hat{\delta}$ and $\hat{\theta}$ can be estimated in the case of latent unobservable variable using this EMA method. We follow closely the work done by Laird et al. (1987), on the EM algorithm implementation and more so on how to obtain ML or REML estimates of vector of fixed and random effect associated with whole plots within the split plot layout structure experiments. However, for us to comprehend and understand clearly how this method work we consider model in 3.3 where ε_j and δ_j are assumed to be

identical independent and normally distributed with both having a vector of mean zero and variance covariance Z and U_j respectively as

$$\varepsilon_j \sim N(0, U_j), \delta_j \sim N(0, Z_j) \text{ but } U_j = \sigma^2 I_{n_j}, j = 1, 2, \dots, b \quad (3.45)$$

We consider full data set that is expected to observed and measured from split plot design experiment to be $(y_j, X_j, \delta_j), j = 1, 2, \dots, b$ but the actual observed data from whole plots is (y_j, X_j) . Then we can notice that there is data is missing and, in this case, the random effect δ_j associated with whole plot. Therefore, δ_j treated as missing observation can be traced through conditioning on the data X_j already observed. Furthermore, we can also find the sufficient statistics of Z and σ^2 estimates for the case no missing data. For us to achieve this, we employ the concept of joint density of $(y_j, \delta_j | X_j), j = 1, 2, \dots, b$ that is proportional to

$$\prod_{j=1}^b \sigma^{-1} \exp \left\{ \frac{(y_j - X_j \theta - d_j \delta_j)' (y_j - X_j \theta - d_j \delta_j)}{2\sigma^2} \right\} |Z_j|^{-\frac{1}{2}} \exp \left\{ \frac{-\delta_j' Z^{-1} \delta_j}{2} \right\} \quad (3.46)$$

The sufficient statistics σ^2 and Z in (3.46) can be shown to be

$$t_1 = \sum_{j=1}^b \varepsilon_j' \varepsilon_j, \quad t_2 = \sum_{j=1}^b \delta_j' \delta_j \text{ where } \varepsilon_j = y_j - X_j \theta - d_j \delta_j \quad (3.47)$$

The estimators for Z and σ^2 can be estimated using (3.47) if full observed data exist and $\hat{\theta}$ can easily be computed. Therefore, the estimates for \hat{Z} and $\hat{\sigma}^2$ when full data is available is given as

$$\hat{\sigma}^2 = t_1/N, \quad \hat{Z} = t_2/b. \quad (3.48)$$

However, this estimates in Equation (3.48) applies only in the case of no missing observation. But the question that triggers many scholars is how to find those estimates regardless the latent unobservable variable. However, some scholars such as Laird *et al.* (1987) proposed EM algorithm as possible approach of getting the solution apart from other method can help the solve the same problem. In addition, the parameter δ_j can be estimated using the EM algorithm if it is treated as missing observation conditional on observing the data X_j as

$$\begin{pmatrix} y_j \\ \delta_j \\ \varepsilon_j \end{pmatrix} \Big| X_j \sim \left\{ \begin{pmatrix} X_j \theta \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} D_j Z D_j' & D_j Z & \sigma^2 I_n \\ Z D_j' & Z & 0 \\ \sigma^2 I_n & 0 & \sigma^2 I_n \end{pmatrix} \right\} \quad (3.49)$$

Laird *et al.* (1987) showed that the conditional expectation of δ_j on (y_j, X_j) when having the marginal joint distribution of (y_j, δ_j) and (y_j, ε_j) is

$$E(\delta_j | y_j, X_j) = Z D_j' V_j^{-1} (y_j - X_j \theta), \quad (3.50)$$

given X_j are embedded in (3.49). This (3.50) follows the standard calculation of conditional moments define as

$$\text{Var}(\delta_j | y_j, X_j) = Z - Z D_j' V_j^{-1} D_j Z,$$

$$\text{since } E(\delta_j \delta_j' | y_j, X_j) = E(\delta_j | y_j, X_j) E(\delta_j | y_j, X_j)' + \text{Var}(\delta_j | y_j, X_j), \quad (3.51)$$

$$\begin{aligned} \text{and } E(\varepsilon_j | y_j, X_j) &= \sigma^2 V_j^{-1} (y_j - X_j \theta), \\ &= y_j - X_j \theta - D_j Z D_j' V_j^{-1} (y_j - X_j \theta), \end{aligned} \quad (3.52)$$

$$\text{where } \text{Var}(\varepsilon_j | y_j, X_j) = \sigma^2 (I_n - \sigma^2 V_j^{-1}).$$

However, they also noted that for easy implementation EM algorithm, then the standard result for quadratic form similar to (3.52) is required which is given as

$$\begin{aligned}
E(\varepsilon_j' \varepsilon_j | y_j, X_j) &= \text{trace} \{E(\varepsilon_j \varepsilon_j' | y_j, X_j)\}, \\
&= E(\varepsilon_j | y_j, X_j)E(\varepsilon_j | y_j, X_j)' + \text{Var}(\varepsilon_j | y_j, X_j),
\end{aligned} \tag{3.53}$$

Now for implementation EM algorithm we base on Equation (3.52) and (3.53) in conjunction with tart some algebra ideas. Basing Laird *et al.* (1987) the EM algorithm should start values $\sigma^{2(0)}$ and $Z^{(0)}$ at k^{th} alteration, we let

$$V_j^{(k)} = \sigma^{2(k)} I_{n_j} + D_j Z^{(k)} D_j', \tag{3.54}$$

then we implement the EM algorithm using the following two steps:

(i) We first obtain

$$\theta^{(k)} = \sum_{j=1}^b (X_j' V_j^{(k)-1} X_j)^{-1} \sum_{j=1}^b X_j' V_j^{(k)-1} y_j, \tag{3.55}$$

(ii) Compute $\sigma^{2(k)}$ and $Z^{(k)}$ by first redefine

$$\begin{aligned}
\tau_j^{(k)} &= y_j - X\theta^{(k)}, \\
\delta_j^{(k)} &= Z^{(k)} D_j' V_j^{(k)-1}, j = 1, 2, \dots, b
\end{aligned}$$

There after update $\sigma^{2(k)}$ and $Z^{(k)}$ as

$$\begin{aligned}
\sigma^{2(k+1)} &= N^{-1} \sum_{j=1}^b \left\{ (\tau_j^{(k)} - D_j \delta_j^{(k)})' (\tau_j^{(k)} - D_j \delta_j^{(k)}) \right. \\
&\quad \left. + \sigma^{2(k)} \text{trace} (I_{n_j} - \sigma^{2(k)} V_j^{(k)-1}) \right\}, \\
Z^{2(k+1)} &= b^{-1} \sum_{j=1}^b \left\{ (\delta_j^{(k)} \delta_j^{(k)'} + Z^{(k)}) - \delta_j^{(k)} D_j Z^{(k)} \right\}.
\end{aligned} \tag{3.56}$$

The estimate of $\hat{\sigma}^2$ and \hat{Z} are obtain after complete iteration of step one and two until convergence is achieved. However, the REML estimates of $\hat{\sigma}^2$ and \hat{Z} can be obtained without difficulties by using SAS/JMP software. This because REML method is already implemented in JMP software which serves as another alternative way of implementing

direct maximization and optimization the estimates value of $\hat{\sigma}^2$ and \hat{Z} without encountering any challenges.

3.3.7 SPD Analysis Using a Full REML Method

This section outline the specific method used to analyze MPV data within the split-plot structure arrangement.

We did a split-plot design analysis using a full REML based on Equation (3.6). It is a well-accepted estimation method for estimating variance components in mixed models and is based on the maximum likelihood estimation of the residual errors discussed in Section 4.2. The analysis was done using statistical software (JMP division of SAS) since it produces tests of composite hypotheses or contrasts. In addition, it also provides both DFs and significance tests directly (Sitinjak and Syafitri, 2019). However, the REML method implemented in JMP 15 provides both error variances (subplot and whole plot error variance) that can be plugged into the Equation (3.6) and thus be used to compute t values and their p -values for the effect or regression coefficient as evidenced with the results discussed in Section 4.2. The prime objective for using this method is because it can be used in all reasonably designed cases as described in Njoroge *et al.* (2017).

3.4 Employing the Screening Methodology in the Framework of a Cox MPV Model in Modeling and Estimating the Predicted Yield of the Specific Variety of *Glycine Max* Using Simulation Technique.

This section outline how the Cox mixture model is formulated, the anticipated yield of *Glycine max* estimated using screening methodology in context of Cox mixture model.

Formulation of Cox Mixture Model in the Context of Screening Methodology

This section describes how the Cox polynomial mixture model was developed. The Cox mixture model was used in this context in order to explain each parameter and serves as an alternative to the Scheffe model described by Goos *et al.* (2016). Following the inherent barrier control for mixture tests ($\sum_{i=1}^q x_i = 1$), Cox restricts parameter ratings if a parameter is a redundancy in a combination of tests, as Hassan *et al.* (2020) points out. The parameter β_i represents the slopes in the Cox mixture model. Each β_i represents a unit change by the response measured at a point, say x , for the response measured in a fixed combination, c , is now given;

$$c = (c_1, c_2, \dots, c_q). \quad (3.57)$$

Many researchers take c as the centroid of the region, but as Njoroge *et al.* (2017) points out, it is unnecessary. For example, if we add a small amount of Δ_q to one of the q elements in a selected fixed combination, a new point is formed, which is called x . But this small quantity Δ_q is added in such a style that the other i^{th} components in c remain the same, and only the ratio component changes. However, we can read what Cornell (2011) described as a component effect when we apply it in the same way. The new point x is on the ray, representing component q as an endless combination with the starting point c . Cornell (2011) define the x point mathematically as

$$x_q = (c_q + \Delta_q). \quad (3.58)$$

Whereas the remaining $q - 1$ components in the mixture change according to their ratio

$$x_i = \left(c_i - \frac{\Delta_q c_i}{1 - c_q} \right) \quad (3.59)$$

To start to formulate a model for small experiment to be carried, then we first define a standard first polynomial model as

$$\zeta_1(x) = \beta_0 + \sum_{i=1}^q \beta_i x_i, \quad (3.60)$$

We, therefore, expressed the difference between x and c as the relative change in the response due to Δ_q .

$$\Delta\zeta_1(x) = \zeta(x) - \zeta(c), \quad (3.61)$$

Hence, following closely the discussion of Cornell (2011), this implies that

$$\begin{aligned} \Delta\zeta_1(x) = \beta_0 + \sum_{i=1}^q \beta_i \left(\left(c_i - \frac{\Delta_q c_i}{1 - c_q} \right) \right) + \beta_q (c_q + \Delta_q) - \beta_0 \\ - \sum_{i=1}^q \beta_i c_i \end{aligned} \quad (3.62)$$

Following the substitution of 3.61 and 3.62 and further, utilizing the restriction on the parameters given Cornell (2011) as

$$\sum_{i=1}^q \beta_i c_i = 0 \quad (3.63)$$

and therefore, cancelling terms using this restriction, this reduces model 3.62 to

$$\Delta\zeta_1(x) = \beta_q (c_q + \Delta_q) + \sum_{i=1}^{q-1} \beta_i \left(\left(c_i - \frac{\Delta_q c_i}{1 - c_q} \right) \right). \quad (3.64)$$

Further simplification result to

$$\Delta\zeta_1(x) = \beta_q c_q + \beta_q \Delta_q + \sum_{i=1}^{q-1} \beta_i c_i - \sum_{i=1}^{q-1} \frac{\beta_i c_i \Delta_q}{1 - c_q}.$$

Therefore, combining terms using the restriction in 3.63, result to

$$\Delta\zeta_1(x) = \beta_q \Delta_q - \sum_{i=1}^{q-1} \frac{\beta_i c_i \Delta_q}{1 - c_q}. \quad (3.65)$$

However, if addition and subtraction is involved according to Weese, we obtain

$$\Delta\zeta_1(x) = \beta_q \Delta_q - \sum_{i=1}^{q-1} \frac{\beta_i c_i \Delta_q}{1 - c_q} + \beta_q \Delta_q \frac{\Delta_q}{1 - c_q} - \beta_q c_q \frac{\Delta_i}{1 - c_i}$$

and once again using the constraint 3.63 we obtain

$$\begin{aligned} \Delta\zeta_1(x) &= \beta_q \Delta_q - \sum_{i=1}^{q-1} \frac{\beta_i c_i \Delta_q}{1 - c_q} + \beta_q \Delta_q \frac{\Delta_q}{1 - c_q} \\ &= \beta_q \frac{\Delta_q}{(1 - c_q)} - \sum_{i=1}^{q-1} \frac{\beta_i c_i \Delta_q}{(1 - c_q)}. \end{aligned} \quad (3.66)$$

Eventually, we define the change in the expected response as advocated by Njoroje *et al.* (2017) as

$$\Delta\zeta_1(x) = \beta_q \frac{\Delta_q}{(1 - c_q)}. \quad (3.67)$$

In particular, we concluded that the intercept β_0 is the fixed point response at c . Further, we can obtain an estimate of the effect of component q as described in Cornell (2011) as

$$\beta_q = \frac{\Delta\zeta_1(x)(1 - c_q)}{\Delta_q}. \quad (3.68)$$

where $\Delta\zeta_1(x) = E [y(x) - y(c)]$ and $y(x), y(c)$ represents the observed responses at x and c , respectively. Therefore, the difference in the heightened by incremental change Δ_q in the proportion of component factor q relative to the quantity $(1 - c_q)$ according represents each β value. However, a quadratic surface Cox's model containing an additional term represents the response surface better than the Scheffe model. Therefore, we define the second-degree polynomial of the Cox model as.

$$\zeta_2(x) = \beta_0 + \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} x_i x_j \quad (3.69)$$

Where $\beta_{ij} = \beta_{ji}$ since they are symmetric. Further, the change in the anticipated response between point x and c is given as

$$\Delta\zeta_2(x) = \zeta_2(x) - \zeta_2(c). \quad (3.70)$$

Therefore,

$$\begin{aligned} \Delta\zeta_2(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} x_i x_j - \sum_{i=1}^q \beta_i c_i \\ & - \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} c_i c_j \end{aligned} \quad (3.71)$$

For further simplification, the terms in model (3.71) are rearranged as follows

$$\begin{aligned} \Delta\zeta_2(x) = & \sum_{i=1}^q \beta_i x_i - \sum_{i=1}^q \beta_i c_i + \\ & \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} x_i x_j - \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} c_i c_j \end{aligned} \quad (3.72)$$

Subject to the following restrictions

$$(i) \quad \sum_{i=1}^q \beta_i c_i = 0,$$

$$(ii) \quad \sum_{j=1}^q \beta_{ij} c_j = 0,$$

as provided by Cornell (2011), then model (3.72) reduces to

$$\begin{aligned} \Delta \zeta_2(x) = & \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} x_i x_j + 2 \sum_{i=1}^{q-1} \beta_{iq} x_i x_q + \beta_{qq} x_q^2 \\ & - \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} c_i c_j \end{aligned} \quad (3.73)$$

Therefore, replacing the relations for x_q and x_i and letting

$$\theta = \frac{\Delta_q}{1 - c_q},$$

is equivalent to

$$\begin{aligned} \Delta \zeta_2(x) = & \sum_{i=1}^{q-1} \sum_{j=1}^{q-1} \beta_{ij} c_i c_j (1 - \theta)^2 + 2 \sum_{i=1}^{q-1} \beta_{iq} c_i (1 - \theta)(c_q + \Delta_q) \\ & + \beta_{qq} (c_q + \Delta_q)^2 - \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} c_i c_j \\ = & \sum_{i=1}^{q-1} \sum_{j=1}^{q-1} \beta_{ij} c_i c_j + (\theta^2 - 2\theta)^2 \sum_{i=1}^{q-1} \sum_{j=1}^{q-1} \beta_{ij} c_i c_j + 2 \sum_{i=1}^{q-1} \beta_{iq} c_i c_q \\ & + 2(\Delta_q - \theta c_q - \theta \Delta_q) \sum_{i=1}^{q-1} \beta_{iq} c_i + \beta_{qq} c_q^2 \\ & + \beta_{qq} (2\Delta_q c_q + \Delta_q^2) - \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} c_i c_j \end{aligned}$$

$$\begin{aligned}
&= (\theta^2 - 2\theta)^2 \sum_{i=1}^{q-1} c_i \sum_{j=1}^{q-1} \beta_{ij} c_j + 2(\Delta_q - \theta c_q - \theta \Delta_q) \sum_{i=1}^{q-1} \beta_{ij} c_i \\
&\quad + \beta_{qq} (2\Delta_q c_q + \Delta_q^2 + (\theta^2 - 2\theta)^2) \sum_{i=1}^{q-1} \beta_{iq} c_i c_q \\
&\quad + (-\theta^2 + 2\theta)^2 \sum_{i=1}^{q-1} \beta_{iq} c_i c_q + 2(\Delta_q - \theta c_q - \theta \Delta_q) \beta_{qq} c_q \\
&\quad - 2(\Delta_q - \theta c_q - \theta \Delta_q) \beta_{qq} c_q \\
&= \beta_{qq} \{ 2\Delta_q c_q + \Delta_q^2 - 2\theta c_q^2 + \theta^2 c_q^2 - 2\Delta_q c_q + 2\Delta_q \theta c_q^2 + 2\theta \Delta_q \}
\end{aligned}$$

and eventually,

$$\Delta \zeta_2(x) = \beta_{qq} \{ \Delta_q^2 + \theta^2 c_q^2 + 2\Delta_q \theta c_q^2 \} \quad (3.74)$$

Hence, substituting the value of θ in (3.74) and including this term onto the linear Cox model, we obtain the following expression for the anticipated yield of *Glycine max* as

$$\Delta \zeta_2(x) = \beta_q \left(\frac{\Delta_q}{1 - c_q} \right) + \beta_{qq} \left(\frac{\Delta_q^2}{[1 - c_q]^2} \right) \quad (3.75)$$

When the mixture constraints are an array of linear constraints over the parameters, we can fit the Cox polynomial using constrained least-squares as Cornell (2011) suggested. Alternatively, the coefficients can be calculated from the fitted Scheffe model as described in Goos *et al.* (2016). Jones & Sall (2011) developed the JMP statistical software, which allows a researcher to create a linear and double (quadratic) cox model from the Scheffe model to $q = 10$.

For instance if a fixed combination, $c_1 = c_2 = c_q$ not necessarily. A researcher wants to add Δ_q to component i^{th} while removing Δ_1 from the first component. In that style, only

the size of the first and last elements changes, while the others remain the same as c . Therefore, we define the new point as

$$\omega = (c_1 - \Delta_1, c_2, c_3, \dots, c_q + \Delta_q), \quad (3.76)$$

Subject to the condition that $\Delta_1 = \Delta_q$ so that the points in ω sum up to a unity. Therefore, the change in the expected response between c and the new point ω can be derived using the linear Cox model where the expected change in response can be represented as

$$\Delta\zeta_1(\omega) = \zeta_1(\omega) - \zeta_1(c), \quad (3.77)$$

Substituting for $\zeta_1(c)$ and $\zeta_1(\omega)$ we obtain

$$\Delta\zeta_1(\omega) = \beta_0 + \beta_1(c_1 - \Delta_1) + \beta_q(c_q + \Delta_q) + \sum_{i=1}^{q-2} \beta_i c_i - \beta_0 - \sum_{i=1}^q \beta_i c_i.$$

Further simplification result to

$$\begin{aligned} \Delta\zeta_1(\omega) &= \beta_q \Delta_q - \beta_1 \Delta_1, \\ &= (\beta_q - \beta_1) \Delta \quad \text{where } \Delta_1 = \Delta_q = \Delta. \end{aligned} \quad (3.78)$$

The expression in Equation (3.78) demonstrates the simple interpretation of the Cox mixture model regression coefficients. Subtracting the amount Δ from each c_i and c_1 again, the change in response from c to new point ω results from the difference in each component's slope weighed by the change Δ . Typically, if β_i is greater than β_1 , adding more component i , automatically, the response will increase drastically, regardless of the amount Δ subtracted from component one. If the difference is nearly zero, then neither subtracting from one nor adding to the other will change the response much from the fixed point c , and the predicted response surface is flat in that direction. However, when β_q and β_{qq} parameter in the Cox model deviates from the static combination, the effect of the

component q shows the shape and direction of the response. Further, there is no need to experiment over the entire space if the interest lies only in a certain component or a few components, as pointed by Njoroge *et al.* (2017). However, considering the four mixture components and two process variable used in this study we formulated the screening methods for agriculture mixture experimenters with the basic intuitive interpretation of the Cox mixture model with *Glycine max* as

$$\begin{aligned}
\hat{Y}(x, z) = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 a_4 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 \\
& + \beta_{44} x_4^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 \\
& + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 + \gamma_{11} x_1 z_1 + \gamma_{21} x_2 z_1 + \gamma_{31} x_3 z_1 \\
& + \gamma_{41} x_4 z_1 + \gamma_{12} x_1 z_2 + \gamma_{22} x_2 z_2 + \gamma_{32} x_3 z_2 + \gamma_{42} x_4 z_2
\end{aligned} \tag{3.79}$$

where $\hat{Y}(x, z)$ represent the expected response variable, $\beta_0, \beta_i, \beta_{ii}$, and β_{ij} indicates the regression coefficient of variable intercept, linear, quadratic and interaction terms, respectively. On the other hand γ_{ij} denotes the regression coefficient of interaction terms between the process variable and a mixture component. The simulation of predicted response of *Glycine max* was carried out based on model (3.70), (3.76) and (3.79) using the algorithms implemented in the JMP software version in conjunction with the prior knowledge of the anticipated yield of *Glycine max* described in Wanyama (2013).

3.5 Estimating the Optimal Yield of *Glycine Max* in the Framework of SPD

This section outline how soybean yield is estimated from the actual field. Here, we follow closely the discussion of Chad and Herbek (2005), University of Kentucky College of Agriculture, on how to estimate soybean yield.

Estimating *Glycine max* yield while the crop is still standing in the field can be a daunting task. The estimating procedure can proceed with caution since variability population, seed per pod and seed size (seed per pound) can all drastically affect the final soybean yield. The best estimate can always be attained at the reproductive stage, as shown in Figure 3.12. Usually, this happens when the green pod's cavity is filled with seeds since the yield estimate may be inaccurate when carried out before the source serves is complete. Assumption of the final number of pods, seed per pod, and seed per pound may not accurately reflect those values at maturity. Therefore, when estimating the yield of components, the best approach is when 5 to 10 random locations across the field are used. Typically, this aid in obtaining a better field average yield since sampling from multiple locations in the area is regarded as the best way to improve the overall yield estimate per acre or ha.



Plate 3.1: Soybean plants at reproductive stage

However, Soybean yield estimating Equation with the aim of estimating Glycine max yield

per acre given as

$$\eta(x, z) = \frac{Y_0 \times Y_4 \times Y_7}{\vartheta_1 \times \tau} : \begin{cases} Y_0 = \text{plants per acre} \\ Y_4 = \text{pods per plant} \\ Y_7 = \text{seeds per pod} \\ \vartheta_1 = \text{seeds per pound} \\ \tau = \text{pounds per bushel} \end{cases} \quad (3.80)$$

Where $\eta(x, z)$ is the *Glycine max* yields estimated per acre in bushels (1 bushel = 25.4 Kg). One bushel of *Glycine max* (L.) Merrill weighs 60 pounds as reported by Chad and Herbek (2005).

3.5.1 Plants per Acre

We obtain the total number of plants per acre for each sample area in the field. The sample space, in this case, is nine whole plots. The number of plants in 10 ft of one row is counted and then divided by 10 to find plants per foot for widths of 30, 25, and 20 inches. For the case of the width of 15 and 10 inches, the number of plants in 40 ft of one row (or 10 ft of four separate rows) is counted to determine plants per acre.

3.5.2 Pods per Plant

The number of pods on each plant for 10 consecutive plants in one row is counted regardless of plant size. The average number of the pods is then determined.

3.5.3 Seeds per Pod

Count the number of seeds per pod on each plant for 10 consecutive plants used in Section 3.5.1 above. Then the average of the number of the seeds per pods is determined.

3.5.4 Seeds per Pound (Seed Size)

Seeds per pound refer to a predetermined number of live seeds per square foot to attain a specific plant density. For instance, on average, they can be 22 seed per head and ten head plant stem, or 220 seeds per plant. If they are 100 plants per square foot, then the whole seeds are 2200, described as 2200 seeds per pound. Seed per pound, generally, may be the most daunting task to estimate, although research by Lee and Jim Herbek (2005) indicates that 2500 seed per pound is an average seed weight estimate, sometimes seeds per pound

can be high 3400. However, the initial seed size provides a reasonable indication of soybean seed size, but where the seed tag is not available, then Chad Lee and Jim Herbek (2005) recommend 2500 seed per pound to be used. We therefore, reported the result in Chapter Four using 2500 seed per pound since seed tag was not available.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction

This section outlines the statistical output obtained based on the split-plot designs implemented and mixture experiments that have been employed so as to solve each of the specific objectives.

4.2 The Improved Model for Split-Plot Design in the Context of MPV

This section outlines the results six different design option in their performance in fitting the developed model 3.5 and the proposed design in Figure 3.3.

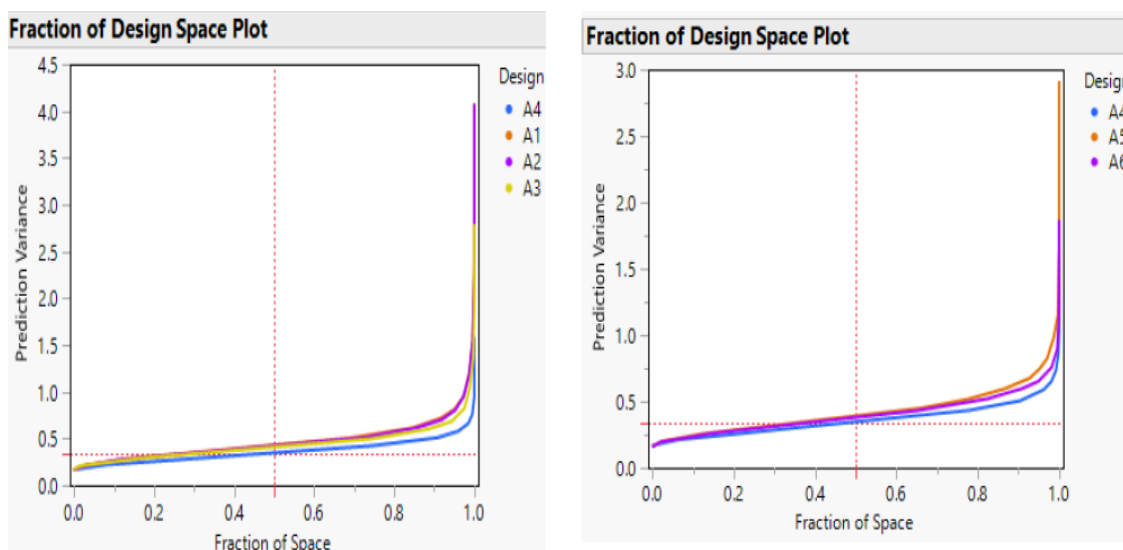


Figure 4.1: Sliced FDS plot of design A4 Relative to design A1, A2, A3, A5 and A6

In this Figure 4.1, Sliced FDS plots shows that Design A4 is better than the rest of designs as it has smaller prediction variation less than 0.5. The D-, A-, I-, and G- efficiency of

design A4 relative to design A1, A2, A3, A4, A5, and A6 in Table 4.1 is above 1.0, which shows again that design A4 in this case is good as compared to others.

Table 4.1: Optimality criterion efficiency of design A4 Relative to design A1, A2, A3, A5 and A6

Optimality Criterion	Efficiency of A4 Relative to A1	Efficiency of A4 Relative to A2	Efficiency of A4 Relative to A3	Efficiency of A4 Relative to A5	Efficiency of A4 Relative to A6
D-efficiency	1.450	1.328	1.067	1.159	1.239
G-efficiency	2.572	2.567	1.758	1.834	1.176
A-efficiency	2.007	1.807	1.507	1.307	1.395
I-efficiency	1.298	1.277	1.207	1.181	1.119

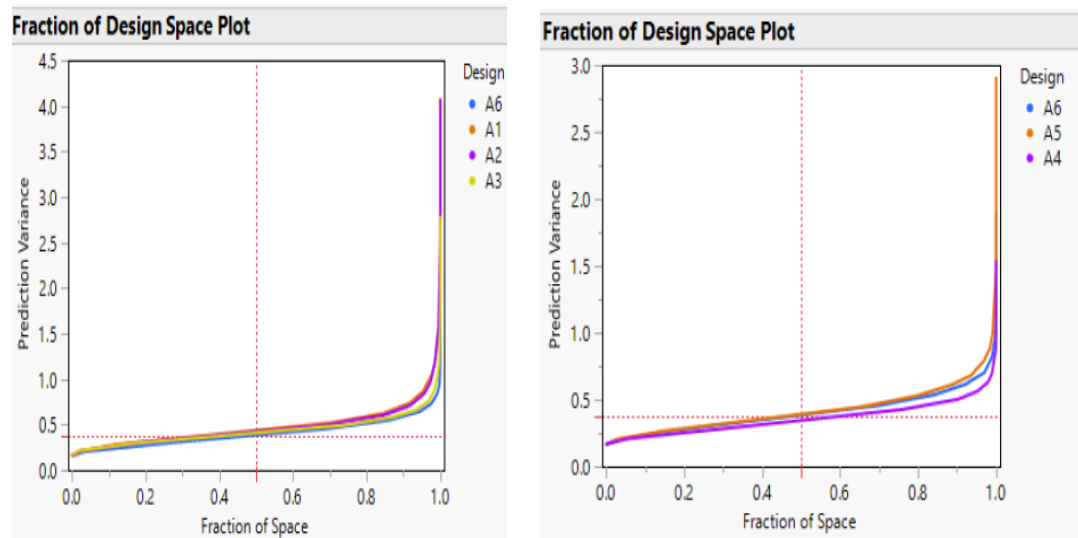


Figure 4.2: Sliced FDS plot of design A6 Relative to design A1, A2, A3, A4 and A5

Table 4.2: The optimality criterion efficiency of design A6 Relative to design A1, A2, A3, A4 and A5

Optimality Criterion	Efficiency of A6 Relative to A1	Efficiency of A6 Relative to A2	Efficiency of A6 Relative to A3	Efficiency of A6 Relative to A4	Efficiency of A6 Relative to A5
D-efficiency	1.171	1.072	0.861	0.935	0.807
G-efficiency	2.104	2.100	1.438	1.534	0.812
A-efficiency	1.439	1.296	1.080	0.937	0.717
I-efficiency	1.178	1.149	1.069	1.044	0.891

Design A6 relative to design A1, A2, A3, A4, and A5 in Figure 4.2 shows that the variance prediction above 0.5. Again, in Table 4.2 shows that not all the D-, A-, I-, and G- efficiency of design A6 relative to design A1, A2, A3, A4, and A5 is above 1.0. Therefore, design A6 is not good comparative the other design.

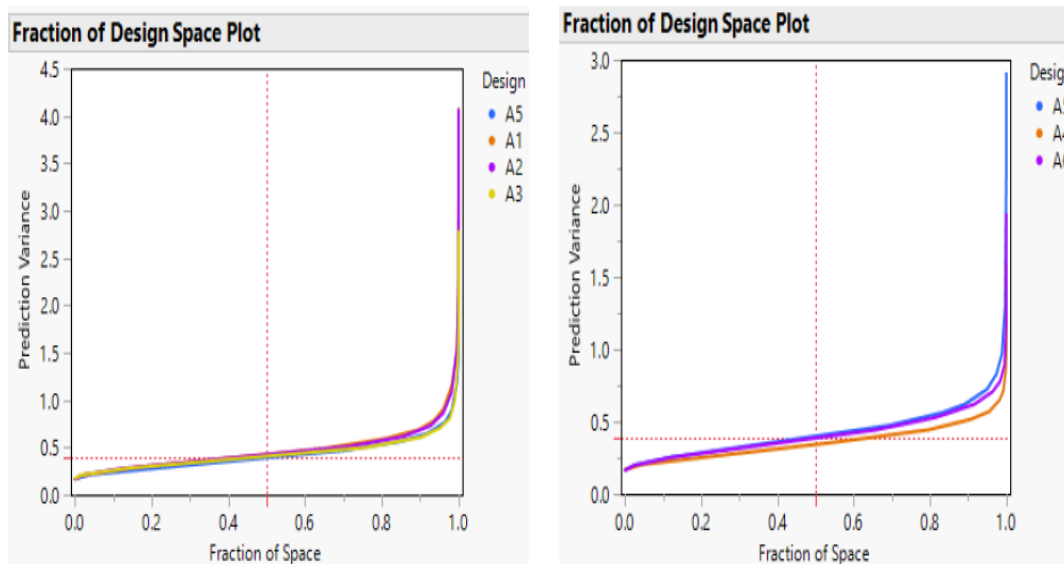


Figure 4.3: Sliced FDS plot of design A5 Relative to design A1, A2, A3, A4 and A6

In this Figure 4.3 shows that design A5 relative to design A1, A2, A3, A4 and A6 has scaled prediction variance above 0.5. Further, Table 4.3 shows that not all the D-, A-, I-, and G- efficiency of this design relative to design A1, A2, A3, A4, and A6 is above 1.0. Therefore, design A5 is not good comparative the other design.

Table 4.3: Optimality criterion efficiency of design A5 Relative to design A1, A2, A3, A4 and A6

Criterion	Efficiency of A5 Relative to A1	Efficiency of A5 Relative to A2	Efficiency of A5 Relative to A3	Efficiency of A5 Relative to A4	Efficiency of A5 Relative to A6
D-efficiency	1.252	1.146	0.921	0.863	1.069
G-efficiency	1.402	1.399	0.958	0.545	0.667
A-efficiency	1.536	1.383	1.153	0.765	1.067
I-efficiency	1.121	1.102	1.024	0.843	0.957

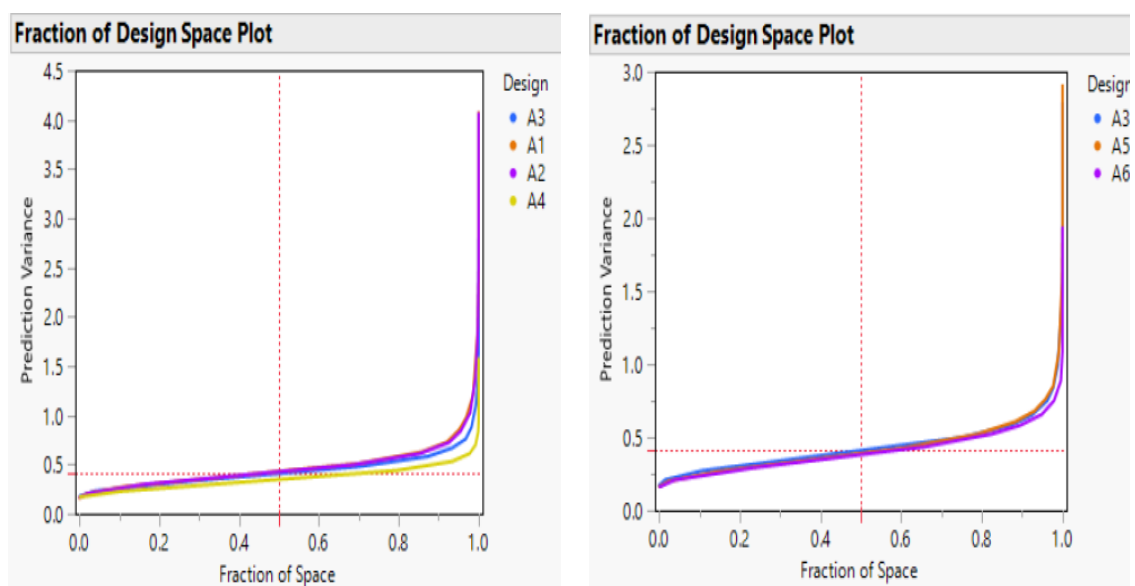


Figure 4.4: Sliced FDS plot of design A3 Relative to design A1, A2, A4, A5 and A6

In this Figure 4.4 shows that design A3 relative to design A1, A2, A4, A5 and A6 has scaled prediction variance above 0.5. Further, Table 4.4 shows that not all the D-, A-, I-, and G- efficiency of this design relative to design A1, A2, A4, A5, and A6 is above 1.0. Therefore, design A3 is not good comparative the other design.

Table 4.4: The optimality criterion efficiency of design A3 Relative to design A1, A2, A4, A5 and A6

Criterion	Efficiency of A3 Relative to A1	Efficiency of A3 Relative to A2	Efficiency of A3 Relative to A4	Efficiency of A3 Relative to A5	Efficiency of A3 Relative to A6
D-efficiency	1.359	1.245	0.937	1.086	1.161
G-efficiency	1.463	1.460	0.569	1.043	0.695
A-efficiency	1.332	1.199	0.664	0.867	0.926
I-efficiency	1.085	1.071	0.828	0.975	0.924

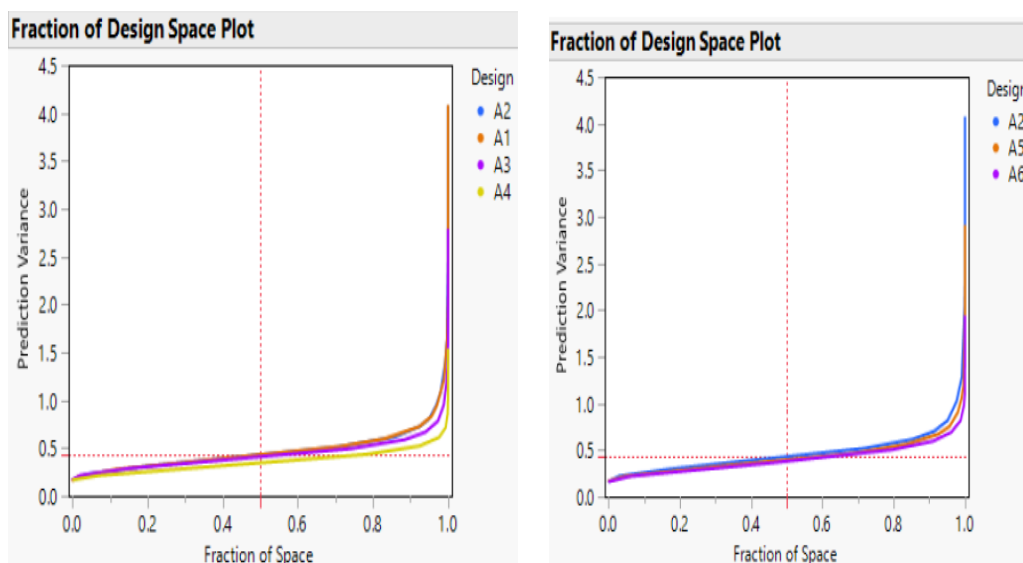


Figure 4.5: Sliced FDS plot of design A2 Relative to design A1, A3, A4, A5 and A6

In this Figure 4.5 shows that design A2 relative to design A1, A3, A4, A5 and A6 has scaled prediction variance above 0.5. Further, Table 4.5 shows that not all the D-, A-, I-, and G- efficiency of this design relative to design A1, A3, A4, A5, and A6 is above 1.0. Therefore, design A2 is not good comparative the other design.

Table 4.5: The Optimality criterion efficiency of design A2 Relative to design A1, A3, A4, A5 and A6

Criterion	Efficiency of A2 Relative to A1	Efficiency of A2 Relative to A3	Efficiency of A2 Relative to A4	Efficiency of A2 Relative to A5	Efficiency of A2 Relative to A6
D-efficiency	1.092	0.803	0.753	0.872	0.933
G-efficiency	1.002	1.685	0.378	0.715	0.476
A-efficiency	1.111	0.834	0.553	0.723	0.772
I-efficiency	1.000	0.923	0.766	0.918	0.867

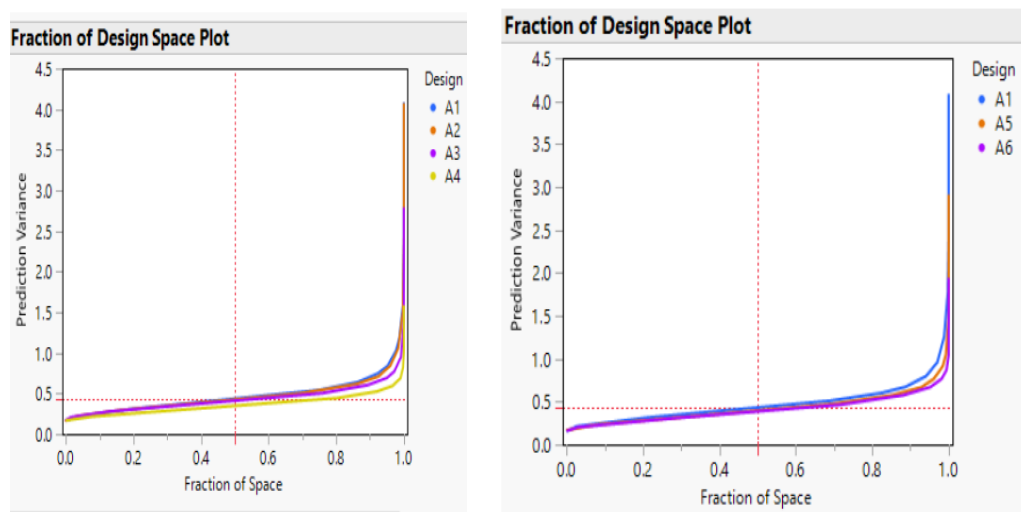


Figure 4.6: Sliced FDS plot of design A1 Relative to design A2, A3, A4, A5 and A6

In this Figure 4.6 shows that design A1 relative to design A2, A3, A4, A5 and A6 has scaled prediction variance above 0.5. Further, Table 4.6 shows none of the D-, A-, I-, and G- efficiency of this design relative to design A2, A3, A4, A5, and A6 is above 1.0. Therefore, design A1 is not good comparative the other design.

Table 4.6: The optimality criterion efficiency of design A1 Relative to design A2, A3, A4, A5 and A6

Criterion	Efficiency of A1	Efficiency of A1	Efficiency of A1	Efficiency of A1	Efficiency of A1
	Relative to A2	Relative to A3	Relative to A4	Relative to A5	Relative to A6
D-efficiency	0.916	0.736	0.690	0.799	0.854
G-efficiency	0.998	0.683	0.389	0.713	0.475
A-efficiency	0.900	0.751	0.498	0.651	0.695
I-efficiency	0.975	0.916	0.762	0.898	0.862

However, we also report the D-efficiency and Average Variance Prediction obtained using JMP software division of SAS for each of the six design as shown in Table 4.7.

Table 4.7: The D-efficiency and average variance prediction for design A1, A2, A3, A4, A5 and A6

Design	D-efficiency	Average Variance Prediction
A1	0.850732	0.154656
A2	0.985671	0.146223
A3	1.17388	0.146866
A4	1.391721	0.089642
A5	1.212415	0.107664
A6	1.049311	0.113173

According to scale provided by Jones and Sall (2011), then design A4 with D-efficiency 1.391721 is best design since it has average variance prediction 0.089642 which is the smallest amongst all the other designs as shown in Table 4.7. Basing on relative efficiency shown in Table 4.1 to 4.6 together with sliced FDS plots in Figure 4.1 to 4.6, we conclude that design A4 is the best desirable design that support and fit combined second order mixture process variable model within the split-plot layout structure. We therefore, employed this design A4 for growing *Glycine max* (L.) Merrill in order to maximize the yield output of Soybean on small scale farm.

4.2 The Modified MPV Model in Predicting the Yield of *Glycine Max*

This section outline the results obtained when modified MPV model in the context of split-plot structure arrangement was employed in modeling the yield of Soybean with minimal split-plot and main plot errors. In addition, the MPV data obtained was categorized into two: simulated data and the actual data collected from the field of experiment for two variety of soybean used (R 184 and Blyvoor) using the same settings of MPV within SPD.

4.2.1 Data Source

Simulated and Experimental Results of *Glycine max* in the Context of SPD

Table 4.8 shows the general simulated results of *Glycine max* crop based on MPV settings within SPD. On the other hand, Table 4.9 and 4.10 shows the experimental results for R 184, and Blyvoor, respectively. In both cases, the MPV data consist of eight response measurements obtained from *Glycine max* (L.) Merrill after the soybean varieties were subjected to different types of MPV treatments subject to SPD layout developed MPV design. These eight responses measured include the number of branches per plant (Y_1), number of pods on branches (Y_2), pods per branch (Y_3), pods on the main stem (Y_4), entire

Pods per plants (Y_5), number of seeds per plant (Y_6), seeds per pod (Y_7), and yield of seeds in grams per plant (Y_8). These results show that there is good agreement between simulated and experimental results. Therefore, this indicates that formulation MPV settings considering SPD layout and simulation using the same settings act as the guiding principle for the actual field MPV experiment utilizing the split-plot structure framework. The results depict that Blyvoor performed slightly better than variety R 184 with the same mixture process variable setting.

Table 4.8: The simulated response of Glycine max

R	W	X_1	X_2	X_3	X_4	Z_1	Z_2	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
1	1	0.25	0.25	0.25	0.25	-1	1	1.7	9.6	4.6	29.5	39.1	100.5	2.5	13.9
2	1	0	0	0	1	-1	1	0.6	2	2.5	20.8	22.8	51.7	2.1	9.7
3	1	0	1	0	0	-1	1	0.5	2.1	2.1	24.3	26.4	52.3	2.1	9.5
4	1	0.25	0.25	0.25	0.25	-1	1	1.7	9.3	4.6	29.9	39.2	100.3	2.5	13.7
5	1	0	0	1	0	-1	1	0.5	1.9	2.5	25.4	27.3	50.4	2.1	9.8
6	1	1	0	0	0	-1	1	0.6	1.8	2.6	20.7	22.5	50.2	2.1	8.9
7	2	0.25	0.25	0.25	0.25	1	-1	2.2	10.5	5.1	31.0	41.5	98.1	2.6	14.6
8	2	0.25	0.25	0.25	0.25	1	-1	2.2	9.8	4.6	30.5	40.3	97.7	2.6	14.1
9	2	0	0	1	0	1	-1	0.7	3.7	2.5	21.1	24.8	61.5	2.1	10.4
10	2	0	1	0	0	1	-1	0.8	3.4	2.3	20.1	23.5	60.3	2.1	9.7
11	2	0	0	0	1	1	-1	0.6	3.2	2.6	19.9	23.1	58.4	2.1	9.5
12	2	1	0	0	0	1	-1	0.8	3.1	2.6	20.5	23.6	58.9	2.1	9.8
13	3	0.5	0.5	0	0	1	1	2	5.5	4	29.5	35	83.9	2.4	12.3
14	3	0.5	0	0.5	0	1	1	2	5.4	4.3	29.7	35.1	89.1	2.4	12.8
15	3	0.5	0	0	0.5	1	1	1.9	5.3	4.1	30.2	35.5	87.2	2.4	12.4
16	3	0	0.5	0	0.5	1	1	1.8	5.3	4	30	35.3	84.9	2.4	12.3
17	3	0	0.5	0.5	0	1	1	1.9	5.1	4.2	29.1	34.2	89	2.4	12.4
18	3	0	0	0.5	0.5	1	1	1	5.2	4.1	29.2	34.4	80.9	2.4	12.5
19	4	0	0.5	0.5	0	-1	-1	1	5.3	3.4	21.6	26.9	80.6	2.2	11.3
20	4	0.5	0.5	0	0	-1	-1	0.9	4.9	2.8	21	25.9	64.7	2.2	10.6
21	4	0	0	0.5	0.5	-1	-1	0.8	5	2.9	20.9	25.9	67.4	2.2	10.9
22	4	0.5	0	0.5		-1	-1	0.8	4.1	2.9	21	25.1	66.3	2.2	10.6
23	4	0.5	0	0	0.5	-1	-1	0.9	4.8	2.7	21.2	26	64.6	2.1	10.5

24	4	0.25	0.25	0.25	0.25	-1	-1	0.9	3.8	2.9	22.3	26.1	66.5	2.1	10.8
25	5	1	0	0	0	0	0	0.8	2.4	2.3	23.8	26.2	53.1	2.1	9.6
26	5	0	1	0	0	0	0	0.8	2.8	2.5	23.9	26.7	57.8	2.1	9.9
27	5	0	0	1	0	0	0	0.7	2.6	2.3	22.7	25.3	55.1	2.1	9.6
28	5	0	0	0	1	0	0	0.7	2.5	2.5	22.9	25.4	52.5	2.1	9.9
29	5	0.25	0.25	0.25	0.25	0	0	2.2	9.7	4.3	30.7	40.4	94.1	2.5	14.1
30	5	0.25	0.25	0.25	0.25	0	0	2.1	9.6	4.3	31.5	41.1	94.8	2.5	13.8
31	6	0.5	0	0.5	0	0	1	1.5	7.4	3.8	23	30.4	75.5	2.3	11.3
32	6	0.5	0.5	0	0	0	1	1.4	7.1	3.9	24.3	31.4	79.5	2.3	11.5
33	6	0.5	0	0	0.5	0	1	1.5	6.8	3.5	26.3	33.1	78.3	2.3	11.3
34	6	0	0.5	0.5	0	0	1	1.4	6.2	3.6	27.5	33.7	75.1	2.3	11.5
35	6	0	0.5	0	0.5	0	1	1.4	6.4	3.6	26.2	32.6	78.9	2.3	11.3
36	6	0	0	0.5	0.5	0	1	1.3	6	3.6	24	30	70.8	2.3	11.6
37	7	0.5	0	0.5	0	0	-1	1.3	6.7	3.7	20.5	27.2	80.6	2.3	12.1
38	7	0.5	0.5	0	0	0	-1	1.3	5.9	3.5	23	28.9	73.4	2.2	11.3
39	7	0.5	0	0	0.5	0	-1	1.2	5.9	3.5	22.7	28.6	69.6	2.3	11.9
40	7	0	0.5	0.5	0	0	-1	1.2	5.6	3.4	21.1	26.7	67.5	2.3	11.1
41	7	0	0.5	0	0.5	0	-1	1.1	5.6	3.3	20.8	26.4	67.7	2.3	11.2
42	7	0	0	0.5	0.5	0	-1	1	5.7	3.4	21.3	27	68.2	2.2	11.7
43	8	0.25	0.25	0.25	0.25	-1.414	0	1.8	8.1	4.5	27.8	35.9	93.9	2.5	13.5
44	8	0.25	0.25	0.25	0.25	-1.414	0	1.9	8.8	4.4	27.4	36.2	89.9	2.5	13.6
45	8	0.25	0.25	0.25	0.25	-1.414	0	1.7	8.9	4.5	29.7	38.6	93.1	2.5	13.7
46	8	0.25	0.25	0.25	0.25	-1.414	0	1.6	8.6	4.4	29.1	37.7	92.2	2.5	13.3
47	8	0.25	0.25	0.25	0.25	-1.414	0	1.7	8.6	4.4	29.6	38.2	91.7	2.4	13.4
48	8	0.25	0.25	0.25	0.25	-1.414	0	1.6	8.2	4.5	30.6	38.8	90.2	2.4	13.4
49	9	0.25	0.25	0.25	0.25	1.414	0	2.3	7.4	4.8	35.8	43.2	101.7	2.6	14.3
50	9	0.25	0.25	0.25	0.25	1.414	0	2.4	10.5	4.7	33.2	43.7	103.6	2.6	14.6
51	9	0.25	0.25	0.25	0.25	1.414	0	2.3	10.9	4.7	31.2	42.1	104.3	2.7	14.2
52	9	0.25	0.25	0.25	0.25	1.414	0	2.3	10.3	4.8	32.9	43.2	106.8	2.7	14.5
53	9	0.25	0.25	0.25	0.25	1.414	0	2.4	10.2	4.9	31.9	42.1	106	2.6	14.2
54	9	0.25	0.25	0.25	0.25	1.414	0	2.3	11.4	5.1	32.3	43.7	106.8	2.7	14.4

*R=run, W= whole plot

Table 4.9: The experimental results of Glycine max for R 184

R	W	X_1	X_2	X_3	X_4	Z_1	Z_2	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
1	1	0.25	0.25	0.25	0.25	-1	1	2.7	10.6	5.6	30.5	40.1	105	2.5	14.9
2	1	0	0	0	1	-1	1	1.6	4.0	3.5	21.8	24.8	56.7	2.3	10.7
3	1	0	1	0	0	-1	1	1.5	4.1	3.1	25.3	28.4	57.3	2.1	10.5
4	1	0.25	0.25	0.25	0.25	-1	1	2.7	5.6	30.9	40.2	105.3	2.5	14.7	
5	1	0	0	1	0	-1	1	1.5	3.9	3.5	26.4	29.3	55.4	2.4	10.8
6	1	1	0	0	0	-1	1	1.6	3.8	3.6	21.7	24.5	55.2	2.2	9.9
7	2	0.25	0.25	0.25	0.25	1	-1	3.2	11.5	6.1	32	43.5	104.1	2.6	15.6
8	2	0.25	0.25	0.25	0.25	1	-1	3.2	10.8	5.6	31.5	43.3	102.7	2.6	15.1
9	2	0	0	1	0	1	-1	1.7	4.7	3.5	22.1	26.8	66.5	2.3	11.4
10	2	0	1	0	0	1	-1	1.8	5.4	3.3	21.1	25.5	65.3	2.2	10.7
11	2	0	0	0	1	1	-1	1.6	5.2	3.6	20.9	25.1	63.4	2.1	10.5
12	2	1	0	0	0	1	-1	1.8	5.1	3.6	21.5	25.6	63.9	2.2	10.8
13	3	0.5	0.5	0	0	1	1	3	7.5	5.1	30.5	37	88.9	2.5	13.3
14	3	0.5	0	0.5	0	1	1	3	7.4	5.3	30.7	37.1	94.1	2.4	13.8
15	3	0.5	0	0	0.5	1	1	2.9	7.3	5.1	31.2	37.5	93.2	2.5	13.4
16	3	0	0.5	0	0.5	1	1	2.8	7.3	5	31	37.3	89.9	2.4	13.3
17	3	0	0.5	0.5	0	1	1	2.9	7.1	5.2	30.1	36.2	94	2.4	13.4
18	3	0	0	0.5	0.5	1	1	2	7.2	5.1	30.2	36.4	85.9	2.4	13.5
19	4	0	0.5	0.5	0	-1	-1	2	7.3	4.4	22.6	28.9	85.6	2.4	12.3
20	4	0.5	0.5	0	0	-1	-1	1.9	6.9	3.8	22	27.9	69.7	2.2	11.6
21	4	0	0	0.5	0.5	-1	-1	1.8	7	3.9	21.9	27.9	72.4	2.3	11.9
22	4	0.5	0	0.5		-1	-1	1.8	6.1	3.9	22	27.1	71.3	2.3	11.6
23	4	0.5	0	0	0.5	-1	-1	1.9	6.8	3.7	22.2	28	69.6	2.1	11.5
24	4	0.25	0.25	0.25	0.25	-1	-1	1.9	5.8	3.9	23.3	28.1	71.5	2.3	11.8
25	5	1	0	0	0	0	0	1.8	4.4	3.3	24.8	28.2	58.1	2.1	10.6
26	5	0	1	0	0	0	0	1.8	4.8	3.5	24.9	28.7	62.8	2.2	10.9
27	5	0	0	1	0	0	0	1.7	4.6	3.3	23.7	27.3	59.1	2.1	10.6
28	5	0	0	0	1	0	0	1.7	4.5	3.5	23.9	22.4	57.5	2.0	10.9
29	5	0.25	0.25	0.25	0.25	0	0	3.2	10.7	5.3	31.7	43.4	99.1	2.6	15.1
30	5	0.25	0.25	0.25	0.25	0	0	3.1	10.6	5.3	32.5	43.1	99.8	2.5	14.8
31	6	0.5	0	0.5	0	0	1	2.5	8.4	5.8	24	32.4	80.5	2.3	13.3
32	6	0.5	0.5	0	0	0	1	2.4	8.1	4.9	25.3	33.4	84.5	2.4	13.5
33	6	0.5	0	0	0.5	0	1	2.5	7.8	4.5	27.3	35.1	83.3	2.3	13.3
34	6	0	0.5	0.5	0	0	1	2.4	7.2	4.6	28.5	35.7	80.1	2.3	13.5

35	6	0	0.5	0	0.5	0	1	2.4	7.4	4.6	27.2	34.6	83.9	2.3	13.3
36	6	0	0	0.5	0.5	0	1	2.3	7.0	4.6	25	32	75.8	2.4	13.6
37	7	0.5	0	0.5	0	0	-1	2.3	7.7	4.7	21.5	29.2	85.6	2.4	13.1
38	7	0.5	0.5	0	0	0	-1	2.3	6.9	4.5	24	30.9	78.4	2.2	12.3
39	7	0.5	0	0	0.5	0	-1	2.2	6.9	4.5	23.7	30.6	74.6	2.4	12.9
40	7	0	0.5	0.5	0	0	-1	2.2	6.6	4.4	22.1	28.7	72.5	2.3	12.1
41	7	0	0.5	0	0.5	0	-1	2.1	6.6	4.3	21.8	28.4	72.7	2.2	12.2
42	7	0	0	0.5	0.5	0	-1	2.0	6.7	4.4	22.3	29	73.2	2.2	12.7
43	8	0.25	0.25	0.25	0.25	-1.414	0	2.8	9.1	5.5	28.8	37.9	98.9	2.7	14.5
44	8	0.25	0.25	0.25	0.25	-1.414	0	2.9	9.8	5.4	28.4	37.2	94.9	2.5	14.6
45	8	0.25	0.25	0.25	0.25	-1.414	0	2.7	9.9	5.5	30.7	40.6	98.1	2.6	14.7
46	8	0.25	0.25	0.25	0.25	-1.414	0	2.6	9.6	5.4	30.1	39.7	97.2	2.5	14.3
47	8	0.25	0.25	0.25	0.25	-1.414	0	2.7	9.6	5.4	30.6	40.2	96.7	2.4	14.4
48	8	0.25	0.25	0.25	0.25	-1.414	0	2.6	9.2	5.5	31.6	40.8	95.2	2.4	14.4
49	9	0.25	0.25	0.25	0.25	1.414	0	3.3	8.4	5.8	36.8	45.2	106.7	2.6	15.3
50	9	0.25	0.25	0.25	0.25	1.414	0	3.4	11.5	5.7	34.2	45.7	108.6	2.6	15.6
51	9	0.25	0.25	0.25	0.25	1.414	0	3.3	11.9	5.7	32.2	44.1	109.3	2.8	15.2
52	9	0.25	0.25	0.25	0.25	1.414	0	3.3	11.3	5.8	33.9	45.2	111.8	2.7	15.5
53	9	0.25	0.25	0.25	0.25	1.414	0	3.4	11.2	5.9	32.9	44.1	111.5	2.8	15.2
54	9	0.25	0.25	0.25	0.25	1.414	0	3.3	12.4	6.1	33.3	45.7	111.8	2.7	15.4

*R=run, W= whole plot

Table 4.10: The experimental results of Glycine max for Blyvoor

R	W	X_1	X_2	X_3	X_4	Z_1	Z_2	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
1	1	0.25	0.25	0.25	0.25	-1	1	2.8	10.7	5.8	31.5	41.1	107	2.6	15.9
2	1	0	0	0	1	-1	1	1.7	4.1	3.7	22.8	25.8	58.7	2.3	11.7
3	1	0	1	0	0	-1	1	1.6	4.2	3.3	26.3	29.4	59.3	2.2	11.5
4	1	0.25	0.25	0.25	0.25	-1	1	2.8	10.4	5.8	31.9	41.2	107.3	2.5	15.7
5	1	0	0	1	0	-1	1	1.6	5.9	3.7	27.4	30.3	57.4	2.4	11.8
6	1	1	0	0	0	-1	1	1.7	4.8	3.8	22.7	25.5	57.2	2.2	10.9
7	2	0.25	0.25	0.25	0.25	1	-1	3.3	11.6	6.3	33.0	44.5	107.1	2.6	16.6
8	2	0.25	0.25	0.25	0.25	1	-1	3.3	11.8	5.8	32.5	44.3	104.7	2.7	16.1
9	2	0	0	1	0	1	-1	1.8	6.7	3.7	23.1	27.8	68.5	2.3	12.4
10	2	0	1	0	0	1	-1	1.9	6.4	3.5	22.1	26.5	67.3	2.2	11.7
11	2	0	0	0	1	1	-1	1.7	6.2	3.8	21.9	26.1	65.4	2.1	11.5
12	2	1	0	0	0	1	-1	1.9	6.1	3.9	22.5	26.6	65.9	2.4	11.8
13	3	0.5	0.5	0	0	1	1	3.1	8.1	5.3	31.5	37.9	90.9	2.5	14.3
14	3	0.5	0	0.5	0	1	1	3.1	8.4	5.5	31.7	38.5	96.1	2.6	14.8
15	3	0.5	0	0	0.5	1	1	3.0	8.0	5.3	32.2	38.1	95.2	2.5	14.4
16	3	0	0.5	0	0.5	1	1	2.9	7.8	5.2	32.0	38.3	90.9	2.4	14.3
17	3	0	0.5	0.5	0	1	1	3.0	7.9	5.3	31.1	37.2	96.0	2.5	14.4
18	3	0	0	0.5	0.5	1	1	2.2	7.7	5.2	31.2	37.4	87.9	2.4	14.5
19	4	0	0.5	0.5	0	-1	-1	2.2	7.8	4.6	23.6	29.9	87.6	2.5	13.3
20	4	0.5	0.5	0	0	-1	-1	2.0	7.4	4.0	23.0	28.9	70.7	2.2	12.6
21	4	0	0	0.5	0.5	-1	-1	1.9	7.5	4.1	22.9	28.9	74.4	2.3	12.9
22	4	0.5	0	0.5		-1	-1	1.9	7.6	4.1	23.1	28.1	73.3	2.3	12.6
23	4	0.5	0	0	0.5	-1	-1	2.1	7.3	3.9	23.2	29.0	70.6	2.2	12.5
24	4	0.25	0.25	0.25	0.25	-1	-1	2.0	7.8	4.0	24.3	29.2	73.5	2.3	12.8
25	5	1	0	0	0	0	0	1.9	4.9	3.5	25.8	29.2	60.1	2.4	11.6
26	5	0	1	0	0	0	0	1.9	5.3	3.7	25.9	29.7	64.8	2.3	11.9

27	5	0	0	1	0	0	0	1.8	5.1	3.5	24.7	28.3	61.1	2.1	11.6
28	5	0	0	0	1	0	0	1.8	5.0	3.6	24.9	27.4	59.5	2.3	11.9
29	5	0.25	0.25	0.25	0.25	0	0	3.1	10.8	5.5	32.7	44.4	101.1	2.5	16.1
30	5	0.25	0.25	0.25	0.25	0	0	3.0	10.6	5.5	33.5	44.1	101.8	2.6	15.8
31	6	0.5	0	0.5	0	0	1	2.6	8.5	5.0	25.0	33.4	82.5	2.5	14.3
32	6	0.5	0.5	0	0	0	1	2.5	8.2	5.1	26.3	34.4	86.5	2.4	14.5
33	6	0.5	0	0	0.5	0	1	2.6	7.9	4.7	28.3	33.5	85.3	2.3	14.3
34	6	0	0.5	0.5	0	0	1	2.5	7.3	4.8	29.5	37.7	82.1	2.5	14.5
35	6	0	0.5	0	0.5	0	1	2.6	7.5	4.8	28.2	35.6	85.9	2.3	14.3
36	6	0	0	0.5	0.5	0	1	2.4	7.1	5.1	26.0	36.1	77.8	2.4	14.6
37	7	0.5	0	0.5	0	0	-1	2.4	7.8	4.9	22.5	30.2	87.6	2.5	14.1
38	7	0.5	0.5	0	0	0	-1	2.4	7.0	4.7	25.0	31.9	79.4	2.2	13.3
39	7	0.5	0	0	0.5	0	-1	2.3	7.0	4.7	24.7	31.6	77.6	2.5	13.9
40	7	0	0.5	0.5	0	0	-1	2.3	6.7	4.6	23.1	29.7	74.5	2.3	13.1
41	7	0	0.5	0	0.5	0	-1	2.2	6.7	4.5	22.8	29.4	74.7	2.2	13.2
42	7	0	0	0.5	0.5	0	-1	2.1	6.7	4.6	23.3	30.0	75.2	2.2	13.7
43	8	0.25	0.25	0.25	0.25	-1.414	0	2.9	9.6	5.7	29.8	38.9	100.9	2.6	15.5
44	8	0.25	0.25	0.25	0.25	-1.414	0	3.0	10.3	5.6	29.6	38.2	97.9	2.5	15.6
45	8	0.25	0.25	0.25	0.25	-1.414	0	2.8	10.4	5.7	31.7	41.6	100.1	2.5	15.7
46	8	0.25	0.25	0.25	0.25	-1.414	0	2.7	10.1	5.6	31.1	40.7	99.2	2.5	15.3
47	8	0.25	0.25	0.25	0.25	-1.414	0	2.8	10.1	5.6	31.6	41.2	98.7	2.4	15.4
48	8	0.25	0.25	0.25	0.25	-1.414	0	2.7	9.7	5.7	31.9	41.8	97.2	2.4	15.4
49	9	0.25	0.25	0.25	0.25	1.414	0	3.4	11.6	6.0	37.8	46.2	108.7	2.6	16.3
50	9	0.25	0.25	0.25	0.25	1.414	0	3.5	12.4	6.1	35.2	46.7	112.6	2.6	16.6
51	9	0.25	0.25	0.25	0.25	1.414	0	3.4	11.7	5.9	33.2	45.1	111.3	2.8	16.2
52	9	0.25	0.25	0.25	0.25	1.414	0	3.4	11.9	6.2	34.9	46.2	111.8	2.5	16.5
53	9	0.25	0.25	0.25	0.25	1.414	0	3.5	11.4	5.9	33.9	45.1	111.5	2.5	16.2
54	9	0.25	0.25	0.25	0.25	1.414	0	3.4	12.0	6.1	34.3	46.7	111.4	2.7	16.4

*R=run, W= whole plot

4.2.2 Exploration and Estimation of Parameters for MPV within Split Plot Layout

This section outline the estimates of parameters in model (3.6) and evaluated two sources of errors arising from Split-plot design as result of modeling the yield of Glycine max using modified MPV settings. The two sources errors includes Whole plot error (WPE) and split-plot error (SPE)

Evaluated WPE and SPE Using REML Method

Table 4.11: The REML variance component $(\sigma_{\delta}^2; \sigma_{\varepsilon}^2)$ estimates obtained from the model

Yield	Random effect	Variance ratio (d)	Variance component	Standard error	95% Lower	95% Upper	Wald p –value
Y_1	W	1.860	0.02602	0.01644	-0.006205	0.05825	0.1135
	R		0.01399	0.003606	0.008941	0.02497	
Y_2	W	2.712	1.19444	0.736727	-0.249519	2.63840	0.1050
	R		0.440428	0.113656	0.2813097	0.786625	
Y_3	W	2.208	0.0625945	0.0390899	-0.01402	0.139209	0.1093
	R		0.0283492	0.0073119	0.0181109	0.050615	
Y_4	W	3.466	5.397922	3.2690654	-1.009329	11.80517	0.0987
	R		1.5574357	0.4016197	0.9950471	2.780324	
Y_5	W	8.495	9.6251669	5.6705645	-1.488935	20.73927	0.0896
	R		1.1330746	0.2924273	0.7236878	2.023855	
Y_6	W	2.677	0.0048252	0.0029668	-0.00099	0.010640	0.109
	R		0.0018023	0.0004647	0.0011515	0.003217	
	W	1.909	25.881354	16.281081	-6.028979	57.79169	0.1119

Y_7	R		13.55451	3.4925294	8.6627493	24.18453	
Y_8	W	6.4905	0.4111375	0.2439755	-0.067046	0.889321	0.092
	R		0.0633445	0.0163494	0.0404565	0.113149	

*W= Whole plot, R= Residual

The result in Table 4.11 shows that variance ratio for each of the eight responses. The variance ratio was obtained using whole plot and subplot variance error given as

$$d = \frac{\sigma_{\delta}^2}{\sigma_{\varepsilon}^2}$$

where σ_{δ}^2 represents the WPE variance whereas σ_{ε}^2 is the SPE variance. The results shows clearly that random effect as well as restricted randomization was completely solved with split-plot structure arrangement since WPE variance is greater than SPE variance as evidenced with variance ratio (d). In addition, none of variance ratio was significant as indicated with the Wald p – value at 5% significance level.

Table 4.12: The summary fit of the eight responses obtained using MPV setting model structure

Summary of fit	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
Multiple R^2	0.9769	0.9789	0.9660	0.9551	0.9857	0.9711	0.9740	0.9875
Adjusted R^2	0.9660	0.9690	0.9500	0.9339	0.9790	0.9575	0.9617	0.9816
RMSE	0.1183	0.1684	0.6636	1.248	1.064	0.0424	3.681	0.2517
Mean response	1.407	3.659	6.239	26.06	32.30	2.331	78.48	11.94

The averagely adjusted R^2 from Table 4.12 shows that 96.23% of the variation in the response is explained by the model. The result shows clearly that the model fits the data well for the eight responses. Also, the results indicate that the second-order MPV model

formulated adequately represents the growth and pod development of *Glycine max*. We can also observe that the model has a reliability of 96.23% on average which can also provide some vital information regarding germination of *Glycine max* (L) Merrill. RMSE in the Table stands for root mean squared error which is defined as the squared root of the square's mean of all of the error.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - O_i)^2}, \quad (4.1)$$

where O_i , S_i and n are the observations, predicted values of a variable (predicted response) and the number observations available for analysis, respectively. RMSE is commonly used because of its nature to provide a good degree of accuracy and compare the predictive errors of different models or model configurations for a given variable and between variables. It depends on the scale as described by Carlisle (2005). RMSE is reported for each of the eight responses fitted with the MPV setting model structure. The RMSE value shown in Table 4.12 shows that the observation captured on the number of seed per pods was well fitted and explained by the model since the lower the value of RMSE, the better the model performance. Thus the more accurate the mixture data recorded.

4.2.3 ANOVA Tables Based on Data from Table 4.8, Analyzing of Mixture Process Data

Table 4.13: The t student test for the fitted MPV model for the No. of branches

Term	Estimate	Standard Error	t value	p value
X_1	0.6260	0.0885	6.99	0.0001
X_2	0.5927	0.0885	6.62	0.0001
X_3	0.5260	0.0885	5.88	0.0001
X_4	0.5260	0.0885	5.88	0.0001
X_1X_2	3.8264	0.3004	12.74	0.0001
X_1X_3	3.7600	0.3004	12.52	0.0001
X_1X_4	3.7600	0.3004	12.52	0.0001
X_2X_3	3.8264	0.3004	12.74	0.0001
X_2X_4	3.6157	0.3357	10.77	0.0001
X_3X_4	2.6215	0.2920	8.98	0.0001
X_1Z_1	0.4189	0.0867	4.83	0.0001
X_2Z_1	0.3395	0.0891	3.81	0.0009
X_3Z_1	0.1776	0.0869	2.04	0.0534
X_4Z_1	0.07090	0.0877	0.81	0.4277
X_1Z_2	0.2484	0.0875	2.84	0.0128
X_2Z_2	0.15491	0.0897	1.73	0.1039
X_3Z_2	0.0554	0.0874	0.63	0.5361
X_4Z_2	0.0800	0.0886	0.90	0.3805

The estimate, standard errors, t values and p values of the fitted Scheffe model for the number of branches per plant observed. The fitted Scheffe model is therefore,

$$\begin{aligned}
\hat{Y} = & 0.626 X_1 + 0.5927X_2 + 0.5260 X_3 + 0.5260 X_4 + 3.8264 X_1X_2 \\
& + 3.76 X_1X_3 + 3.76 X_1X_4 + 3.8264 X_2X_3 + 3.6157 X_2X_4 \\
& + 2.6215 X_3X_4 + 0.4189 X_1Z_1 + 0.3395 X_2Z_1 \\
& + 0.1776 X_3Z_1 + 0.0709 X_4Z_1 + 0.2484 X_1Z_2 \\
& + 0.1549 X_2Z_2 + 0.0554 X_3Z_2 + 0.08 X_4Z_2
\end{aligned} \tag{4.2}$$

The significant factors

were $X_1, X_2, X_3, X_4, X_1 X_2, X_1 X_3, X_1 X_4, X_2 X_3, X_3 X_4, X_1 Z_1, X_2 Z_1$ and $X_1 Z_2$ as shown in Table (4.13). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 0.626 X_1 + 0.5927X_2 + 0.5260 X_3 + 0.5260 X_4 + 3.8264 X_1X_2 \\
& + 3.76 X_1X_3 + 3.76 X_1X_4 + 3.8264 X_2X_3 + 3.6157 X_2X_4 \\
& + 2.6215 X_3X_4 + 0.4189 X_1Z_1 + 0.3395 X_2Z_1 \\
& + 0.2484 X_1Z_2
\end{aligned} \tag{4.3}$$

Table 4.14: The t student test for the fitted MPV model for the No. of Pods on branches

Term	Estimate	Standard Error	t value	p value
X_1	1.6009	0.5425	2.95	0.0078
X_2	1.9342	0.5425	3.57	0.0019
X_3	1.9009	0.5425	3.50	0.0022
X_4	1.7342	0.5425	3.20	0.0044
X_1X_2	20.8843	1.6868	12.38	0.0001
X_1X_3	18.9509	1.6868	11.23	0.0001
X_1X_4	19.5843	1.6868	11.61	0.0001
X_2X_3	17.9843	1.6868	10.66	0.0001
X_2X_4	18.4069	1.8851	9.76	0.0001
X_3X_4	17.4320	1.6401	10.63	0.0001

X_1Z_1	0.6569	0.5326	1.23	0.2344
X_2Z_1	0.2411	0.5447	0.44	0.6633
X_3Z_1	0.6596	0.5334	1.24	0.2331
X_4Z_1	0.4257	0.5378	0.79	0.4393
X_1Z_2	0.2478	0.5511	0.45	0.6611
X_2Z_2	-0.1548	0.5620	-0.28	0.7875
X_3Z_2	-0.0200	0.5503	-0.04	0.9716
X_4Z_2	0.0229	0.5565	0.04	0.9679

The estimate, standard errors, t values and p values of the fitted Scheffe model for the number of pods on branches of *Glycine max* per plant stem observed. The fitted Scheffe model is therefore,

$$\begin{aligned}
\hat{Y} = & 1.6009 X_1 + 1.9342X_2 + 1.9009 X_3 + 1.7342 X_4 + 20.8843 X_1X_2 \\
& + 18.9509 X_1X_3 + 19.5843 X_1X_4 + 17.9843 X_2X_3 \\
& + 18.4069 X_2X_4 + 17.4320 X_3X_4 + 0.6569 X_1Z_1 \\
& + 0.2411 X_2Z_1 + 0.6596X_3Z_1 + 0.4257 X_4Z_1 \\
& + 0.2478 X_1Z_2 - 0.1548 X_2Z_2 - 0.0200 X_3Z_2 \\
& + 0.0229 X_4Z_2
\end{aligned} \tag{4.4}$$

The significant factors were $X_1, X_2, X_3, X_4, X_1 X_2, X_1 X_3, X_1 X_4, X_2 X_3$ and $X_3 X_4$ as shown in Table (4.14). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 1.6009 X_1 + 1.9342X_2 + 1.9009 X_3 + 1.7342 X_4 + 20.8843 X_1X_2 \\
& + 18.9509 X_1X_3 + 19.5843 X_1X_4 + 17.9843 X_2X_3 \\
& + 18.4069 X_2X_4 + 17.4320 X_3X_4
\end{aligned} \tag{4.5}$$

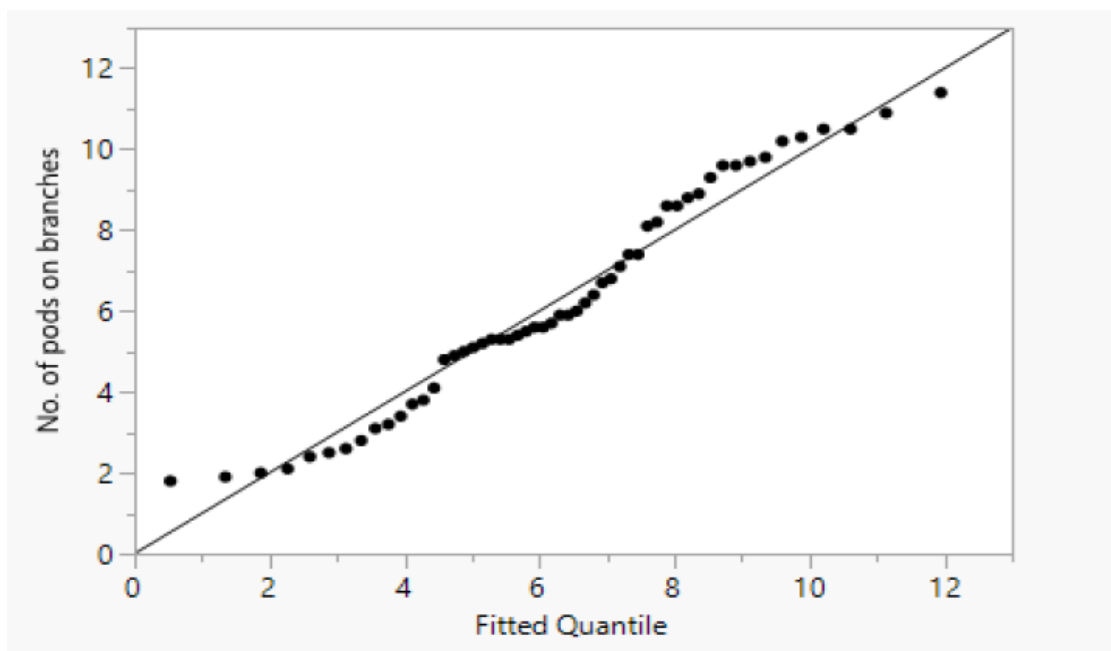


Figure 4.7: The QQ plot for the number of pods on branches verses the fitted quantile

This Figure 4.7 shows that the second order model 3.5 was well fitted since it can be observed that the data obtained for the number of pods on branches satisfied the assumption of normality as most of the points lie on the straight line. The result also indicates that the mixture of experimental data collected from the field is normally distributed. Therefore, the problem of restricted randomization within split-plot design structure arrangements was solved.

Table 4.15: The t student test for the fitted MPV model for Pods on Main stem

Term	Estimate	Standard Error	t value	p value
X_1	20.2988	1.0824	18.75	0.0001
X_2	21.3988	1.0824	19.77	0.0001
X_3	21.6988	1.0824	20.05	0.0001
X_4	19.8321	1.0824	18.32	0.0001
X_1X_2	18.3605	3.1731	5.79	0.0001
X_1X_3	21.7605	3.1731	6.86	0.0001
X_1X_4	27.8938	3.1731	8.79	0.0001
X_2X_3	20.8605	3.1731	6.57	0.0001
X_2X_4	23.3040	3.5459	6.57	0.0001
X_3X_4	21.1135	3.0854	6.84	0.0001
X_1Z_1	2.2803	1.0724	2.13	0.0509
X_2Z_1	1.2280	1.0937	1.12	0.2785
X_3Z_1	0.6804	1.0737	0.63	0.5360
X_4Z_1	2.0735	1.0815	1.92	0.0744
X_1Z_2	1.5415	1.1268	1.37	0.1996
X_2Z_2	2.9897	1.1458	2.61	0.0239
X_3Z_2	2.2016	1.1254	1.96	0.0774
X_4Z_2	2.0861	1.1361	1.84	0.0937

The estimate, standard errors, t values and p values of the fitted Scheffe model for the number of pods on main stem of *Glycine max* per plant stem observed. The fitted Scheffe model is therefore,

$$\begin{aligned}
\hat{Y} = & 20.2988 X_1 + 21.3988X_2 + 21.6988 X_3 + 19.8321 X_4 \\
& + 18.3605 X_1X_2 + 21.7605 X_1X_3 + 27.8938 X_1X_4 \\
& + 20.8605 X_2X_3 + 23.3040 X_2X_4 + 21.1135 X_3X_4 \\
& + 2.2803 X_1Z_1 + 1.2280 X_2Z_1 + 0.6804X_3Z_1 \\
& + 2.0735 X_4Z_1 + 1.5415 X_1Z_2 + 2.9897 X_2Z_2 \\
& + 2.2016 X_3Z_2 + 2.0861 X_4Z_2
\end{aligned} \tag{4.6}$$

The significant factors were X_1 , X_2 , X_3 , X_4 , $X_1 X_2$, $X_1 X_3$, $X_1 X_4$, $X_2 X_3$, $X_3 X_4$ and X_2Z_2 as shown in Table (4.15). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 20.2988 X_1 + 21.3988X_2 + 21.6988 X_3 + 19.8321 X_4 \\
& + 18.3605 X_1X_2 + 21.7605 X_1X_3 + 27.8938 X_1X_4 \\
& + 20.8605 X_2X_3 + 23.3040 X_2X_4 + 21.1135 X_3X_4 \\
& + 2.9897 X_2Z_2
\end{aligned} \tag{4.7}$$

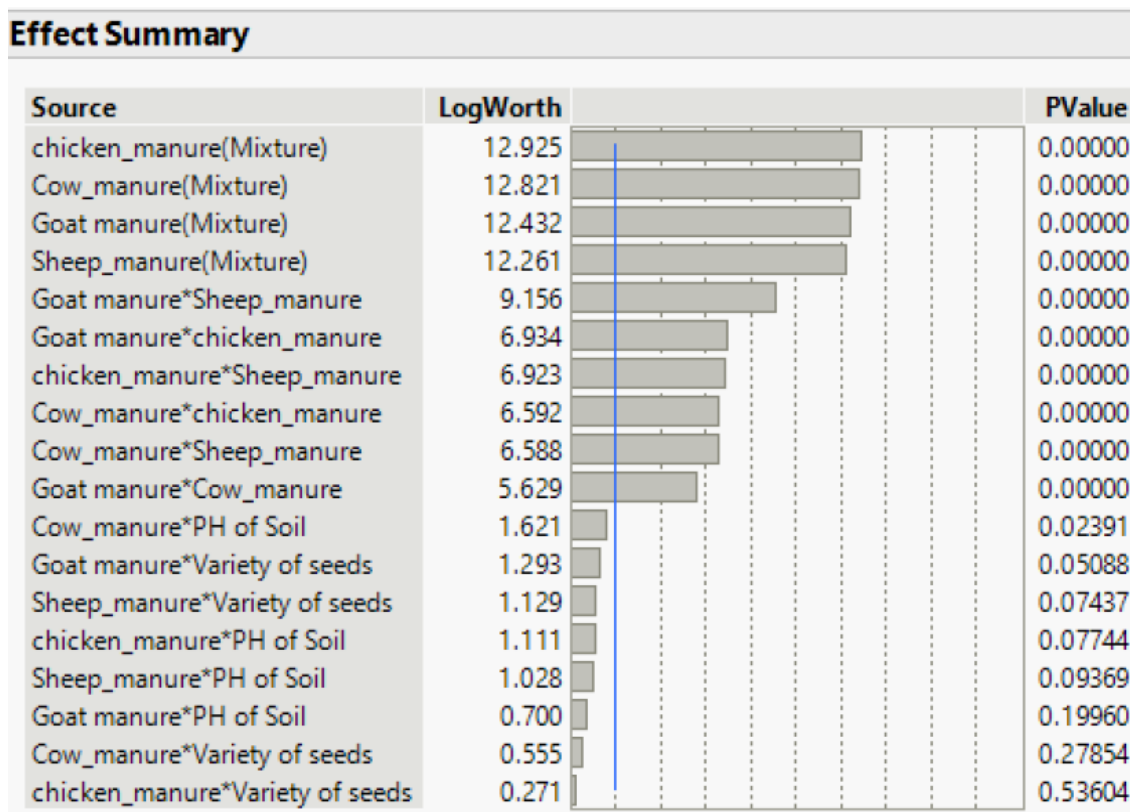


Figure 4.8: The effect summary of MPV settings on Pod development on main stem

The result in Figure 4.8 shows how pure mixture components (Chicken, Cow, Goat, and Sheep manure) have a great influence on soybean growth and pods development on the main stem and plant in general at a 5% significance level. However, this result helps soybean producers to achieve a better yield. We can also observe that the interaction effect resulting from the mixture components greatly impacts *Glycine max* growth and pod development at a 5% significance level. Again, we can observe that the type of variety seed used also has a slight impact on growth and pods development compared to the PH of the soil, which has little significance on Soybean yield's performance, as evidenced from the graph. We can also note that the combination of Goat and sheep manure mixture blend, when set at 1 of various seeds, outweighs the performance chicken and cow manure

mixture setting under the same condition within SPD layout with a significance level close to 5% (0.05088, 0.07437, respectively). We can conclude that chicken manure greatly influences the growth and pods development of *Glycine max*. We, therefore, recommend farmers embrace the use of these organic manures when planting different types of crops on the farm as they have good nutritional value on the growth of plants.

Table 4.16: The least square means (LSM for the number of pods on main stem

Whole plot	LSM of Y_4	Standard error
1	29.98	1.3825
2	30.00	1.3825
3	28.65	1.3406
4	27.90	1.3483
5	31.37	0.6396
6	25.95	1.0726
7	26.67	1.0731
8	31.15	1.2404
9	30.60	1.2400

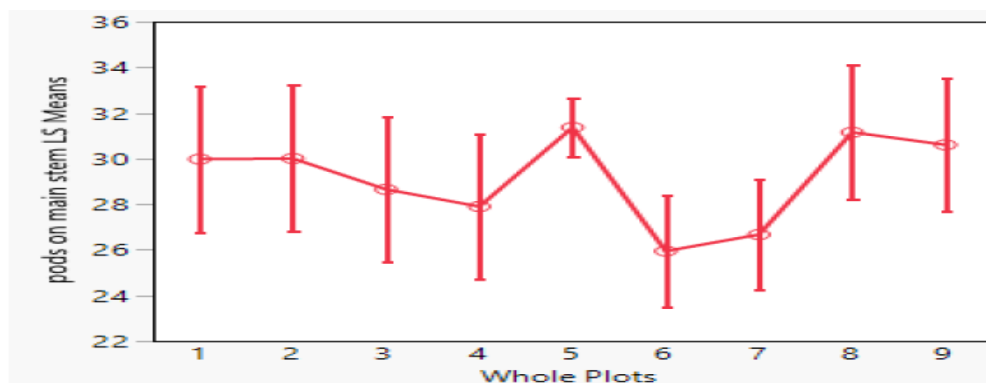


Figure 4.9: The LSM plot for the response Y_4 on main stem within the whole plots

Table 4.16 shows the least square mean (LSM) response on the number of pods development on the main stem. However, it was observed that the performance pods

development on the main stem is averagely the same across the nine whole plots subject to different treatment combinations using mixture process variable settings within each subplot. This also evidenced from Figure 4.9 showing where it was also that whole plots 5 and 8 have posted a higher value with plot 5 being slightly higher with 0.2133 with standard deviation of 0.6396 and 1.0726 respectively with minimal variation in the total yield obtained from whole plot 5. The study therefore, recommend the mixture process variable settings to whole plots 5 and 8 to be used with farmers to achieve a better yield of *Glycine max*.

Table 4.17: The t student test for the fitted MPV model for Pods per branch

Term	Estimate	Standard Error	t value	p value
X_1	2.3704	0.1317	18.00	0.0001
X_2	2.1704	0.1317	16.48	0.0001
X_3	2.3037	0.1317	17.49	0.0001
X_4	2.4037	0.1317	18.25	0.0001
X_1X_2	5.9847	0.4278	13.99	0.0001
X_1X_3	6.0180	0.4278	14.07	0.0001
X_1X_4	5.0180	0.4278	11.73	0.0001
X_2X_3	6.4180	0.4278	15.00	0.0001
X_2X_4	5.4782	0.4781	11.46	0.0001
X_3X_4	5.342	0.4159	12.85	0.0001
X_1Z_1	0.3144	0.1284	2.45	0.0240
X_2Z_1	0.1553	0.1316	1.18	0.2514
X_3Z_1	0.1936	0.1285	1.51	0.1482
X_4Z_1	0.2618	0.1297	2.02	0.0574
X_1Z_2	0.2934	0.1310	2.24	0.0430
X_2Z_2	0.0961	0.1340	0.72	0.4849
X_3Z_2	0.2268	0.1308	1.73	0.1065
X_4Z_2	0.2113	0.1325	1.59	0.1336

The estimate, standard errors, t values and p values of the fitted Scheffe model for the number of pods per branch of *Glycine max* per plant stem observed. The fitted Scheffe model is therefore,

$$\begin{aligned}
 \hat{Y} = & 2.3704 X_1 + 2.1704X_2 + 2.3037X_3 + 2.4037 X_4 + 5.9847 X_1X_2 \\
 & + 6.0180 X_1X_3 + 5.0180 X_1X_4 + 6.4180 X_2X_3 \quad (4.8) \\
 & + 5.4782 X_2X_4 + 5.342 X_3X_4 + 0.3144 X_1Z_1 \\
 & + 0.1553 X_2Z_1 + 0.1936X_3Z_1 + 0.2618X_4Z_1 \\
 & + 0.2934 X_1Z_2 + 0.0961 X_2Z_2 + 0.2268 X_3Z_2 \\
 & + 0.2113 X_4Z_2
 \end{aligned}$$

The significant factors were X_1 , X_2 , X_3 , X_4 , $X_1 X_2$, $X_1 X_3$, $X_1 X_4$, $X_2 X_3$, $X_3 X_4$, X_1Z_1 and X_1Z_2 as shown in Table (4.17). Thus, the final model is

$$\begin{aligned}
 \hat{Y} = & 2.3704 X_1 + 2.1704X_2 + 2.3037X_3 + 2.4037 X_4 + 5.9847 X_1X_2 \\
 & + 6.0180 X_1X_3 + 5.0180 X_1X_4 + 6.4180 X_2X_3 \\
 & + 5.4782 X_2X_4 + 5.342 X_3X_4 + 0.3144 X_1Z_1 \quad (4.9) \\
 & + 0.2934 X_1Z_2
 \end{aligned}$$

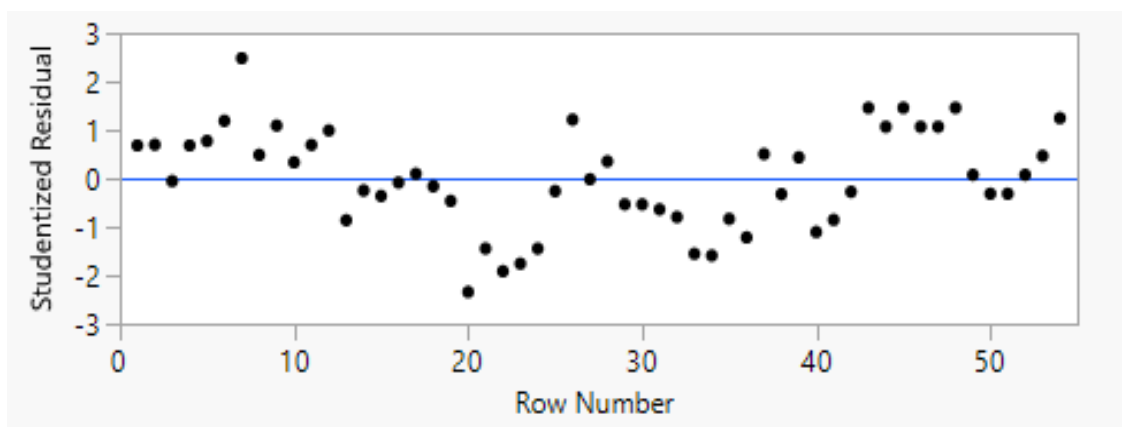


Figure 4.10: The graph of Studentized Residuals for No. of Pods per branch fitted by Scheffe Model

From Figure 4.10, it can be observed that the residuals from the plot are randomly distributed. The studentized residual plot was used to check the independence assumption. As evidenced from this Figure 4.10 without any obvious trend to correlate them, there is no reason to suspect that the independence assumption was not valid. This Figure was also used to verify the linear model-related sample's inclusion and from the plot and it was concluded that there was no reason to suspect that the additivity assumption should not have been accepted since the plot's residuals are randomly distributed around zero. In addition, this plot does not indicate the growth of variance with an increase in the fitted value. Therefore, this enabled to check the adequacy SPD used it, and therefore, it is also worth pointing out the same diagnosis. And again, this reflects that the data was well fitted.

Table 4.18: The t student test for the fitted MPV model for the total No. of Pods

Term	Estimate	Standard Error	t value	p value
X_1	21.8741	1.2191	17.94	0.0001
X_2	23.3074	1.2191	19.12	0.0001
X_3	23.5741	1.2191	19.34	0.0001
X_4	21.54074	1.2191	17.67	0.0001
X_1X_2	39.36272	2.7087	14.53	0.0001
X_1X_3	40.8294	2.7087	15.07	0.0001
X_1X_4	47.5961	2.7087	17.57	0.0001
X_2X_3	38.9627	2.7087	14.38	0.0001
X_2X_4	41.6515	3.0267	13.76	0.0001
X_3X_4	38.7299	2.6341	14.70	0.0001
X_1Z_1	2.9268	1.2446	2.35	0.0420
X_2Z_1	1.4991	1.2580	1.19	0.2615
X_3Z_1	1.3146	1.2454	1.06	0.3175
X_4Z_1	2.5141	1.2503	2.01	0.0433
X_1Z_2	1.7823	1.3687	1.30	0.2299
X_2Z_2	2.8601	1.3801	2.07	0.0716
X_3Z_2	2.1625	1.3679	1.58	0.1535
X_4Z_2	2.1221	1.3743	1.54	0.1614

The estimate, standard errors, t values and p values of the fitted Scheffe model for the total number of pods of *Glycine max* per plant stem observed. The fitted Scheffe model is therefore,

$$\begin{aligned}
\hat{Y} = & 21.8741 X_1 + 23.3074X_2 + 23.5741X_3 + 21.54074 X_4 \\
& + 39.36272 X_1X_2 + 40.8294 X_1X_3 + 47.5961 X_1X_4 \\
& + 38.9627 X_2X_3 + 41.6515 X_2X_4 + 38.7299 X_3X_4 \quad (4.10) \\
& + 2.9268 X_1Z_1 + 1.4991 X_2Z_1 + 1.3146X_3Z_1 \\
& + 2.5141X_4Z_1 + 1.7823 X_1Z_2 + 2.8601 X_2Z_2 \\
& + 2.1625 X_3Z_2 + 2.1221 X_4Z_2
\end{aligned}$$

The significant factors were X_1 , X_2 , X_3 , X_4 , $X_1 X_2$, $X_1 X_3$, $X_1 X_4$, $X_2 X_3$, $X_3 X_4$ and X_1Z_1 as shown in Table (4.18). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 21.8741 X_1 + 23.3074X_2 + 23.5741X_3 + 21.54074 X_4 \\
& + 39.36272 X_1X_2 + 40.8294 X_1X_3 + 47.5961 X_1X_4 \quad (4.11) \\
& + 38.9627 X_2X_3 + 41.6515 X_2X_4 + 38.7299 X_3X_4 \\
& + 2.9268 X_1Z_1
\end{aligned}$$

Table 4.19: The t student test for the fitted MPV model for the No. of seeds per pod

Term	Estimate	Standard Error	t value	p value
X_1	2.0597	0.03460	59.52	0.0001
X_2	2.0597	0.03460	59.52	0.0001
X_3	2.0597	0.03460	59.52	0.0001
X_4	2.0597	0.03460	59.52	0.0001
X_1X_2	1.2119	0.10790	11.23	0.0001
X_1X_3	1.1119	0.10790	10.30	0.0001
X_1X_4	1.1119	0.10790	10.30	0.0001
X_2X_3	1.2119	0.10790	11.23	0.0001
X_2X_4	1.2207	0.12058	10.12	0.0001
X_3X_4	1.0695	0.10491	10.19	0.0001
X_1Z_1	0.0719	0.03396	2.12	0.0491

X_2Z_1	0.0426	0.03473	1.23	0.2352
X_3Z_1	0.0634	0.03400	1.86	0.0796
X_4Z_1	0.0795	0.03429	2.32	0.0326
X_1Z_2	0.0431	0.03511	1.23	0.2434
X_2Z_2	0.0097	0.03581	0.27	0.7916
X_3Z_2	0.0462	0.03505	1.32	0.2120
X_4Z_2	0.0492	0.03545	1.39	0.1898

The estimate, standard errors, t values and p –values of the fitted Scheffe model for the total number of seeds per pod of *Glycine max* per plant stem observed. The fitted Scheffe model is therefore,

$$\begin{aligned}
\hat{Y} = & 2.0597 X_1 + 2.0597X_2 + 2.0597X_3 + 2.0597 X_4 + 1.2119 X_1X_2 \\
& + 1.1119 X_1X_3 + 1.1119 X_1X_4 + 1.2119 X_2X_3 \\
& + 1.2207 X_2X_4 + 1.0695 X_3X_4 + 0.0719 X_1Z_1 \quad (4.12) \\
& + 0.0426 X_2Z_1 + 0.0634X_3Z_1 + 0.0795X_4Z_1 \\
& + 0.0431 X_1Z_2 + 0.0097 X_2Z_2 + 0.0462 X_3Z_2 \\
& + 0.0492 X_4Z_2
\end{aligned}$$

The significant factors were $X_1, X_2, X_3, X_4, X_1 X_2, X_1 X_3, X_1 X_4, X_2 X_3,$

$X_3 X_4, X_1Z_1$ and X_4Z_1 as shown in Table (4.19). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 2.0597 X_1 + 2.0597X_2 + 2.0597X_3 + 2.0597 X_4 + 1.2119 X_1X_2 \\
& + 1.1119 X_1X_3 + 1.1119 X_1X_4 + 1.2119 X_2X_3 \quad (4.13) \\
& + 1.2207 X_2X_4 + 1.0695 X_3X_4 + 0.0719 X_1Z_1 \\
& + 0.0795X_4Z_1
\end{aligned}$$

Table 4.18 and 4.19 represents the analysis of variance results for the determination of the fit of the MPV model. They were both obtained using a REML method. The results in both Tables clearly shows that x_1, x_2, x_3, x_4 , and all the interaction x_1, x_2, x_3 and x_4 are all significant and have a great impact on the number of pods on the main stem per plant. Also, the interaction between the process variable and mixture component factor x_1z_1 and x_4z_1 are significant at 5%. Further, this suggest a possible effects of the mixture process variable interaction resulting from soil pH and the number of seeds used per acre.

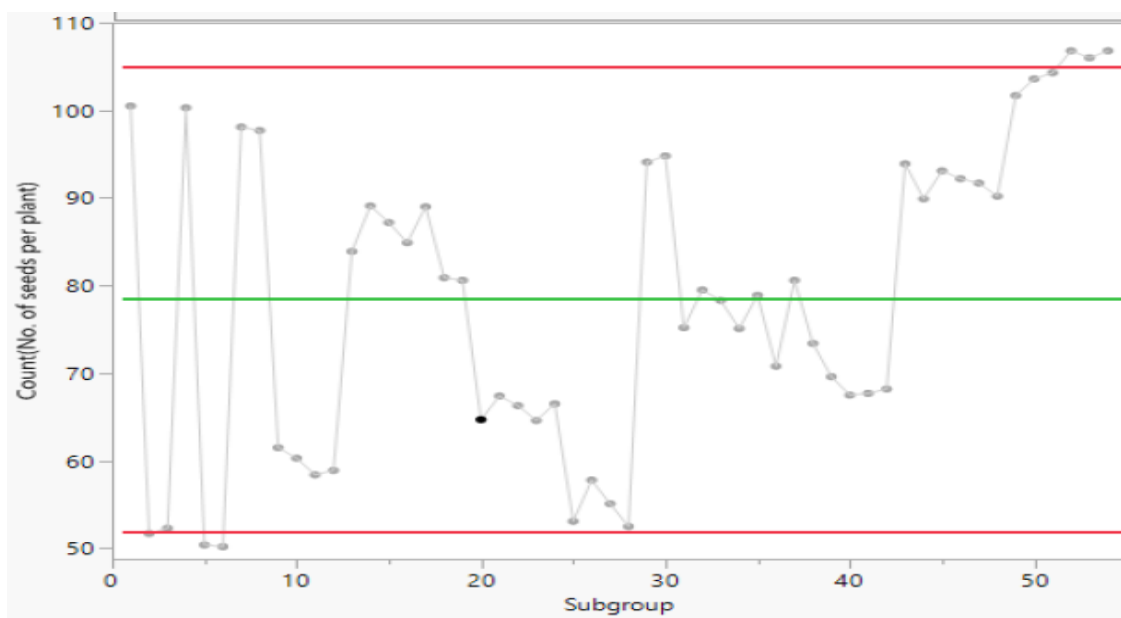


Figure 4.11: The number of seeds per plant observed in each subplot within a split plot layout arrangement

The count chart as shown in Figure 4.11 for the number of seeds per plant indicates which subgroup treatment performed better than the other. This count chart was fitted using the Poisson distribution model. The subgroup between 0 to 5 and 5 to 54 indicates the high number of seeds harvested per plant within a whole plot. The process also at subgroup

indicates there is more likely high fertility at that particular point than the other points. The fitted Poisson model was used to find the maximum number of seeds to be attained with MPV settings. The Count chart also shows the maximum value obtained.

Table 4.20: The t student test for the fitted MPV model for the total No. of seeds per plant

Term	Estimate	Standard Error	t value	p value
X_1	50.7922	2.8001	18.14	0.0001
X_2	53.5255	2.8001	19.12	0.0001
X_3	52.3922	2.8001	18.71	0.0001
X_4	50.9255	2.8001	18.19	0.0001
X_1X_2	113.8722	9.3495	12.18	0.0001
X_1X_3	120.0388	9.3495	12.84	0.0001
X_1X_4	114.3722	9.3495	12.23	0.0001
X_2X_3	118.4722	9.3495	12.67	0.0001
X_2X_4	111.4577	10.4494	10.67	0.0001
X_3X_4	99.1119	9.0894	10.90	0.0001
X_1Z_1	8.0396	2.7149	2.96	0.0074
X_2Z_1	2.8369	2.7877	1.02	0.3197
X_3Z_1	4.4079	2.7193	1.62	0.1199
X_4Z_1	4.7505	2.7462	1.73	0.0979
X_1Z_2	4.6172	2.7449	1.68	0.1142
X_2Z_2	1.1349	2.8121	0.40	0.6920
X_3Z_2	1.4418	2.7397	0.53	0.6068
X_4Z_2	4.0959	2.7780	1.47	0.1612

The estimate, standard errors, t values and p values of the fitted Scheffe model for the total number of seeds of *Glycine max* per plant stem observed. The fitted Scheffe model is therefore,

$$\begin{aligned}
\hat{Y} = & 50.7922 X_1 + 53.5255X_2 + 52.3922X_3 + 50.9255 X_4 \\
& + 113.8722 X_1X_2 + 120.0388 X_1X_3 + 114.3722 X_1X_4 \\
& + 118.4722 X_2X_3 + 111.4577 X_2X_4 + 99.1119 X_3X_4 \quad (4.14) \\
& + 8.0396 X_1Z_1 + 2.8369 X_2Z_1 + 4.4079X_3Z_1 + 4.7505X_4Z_1 \\
& + 4.6172 X_1Z_2 + 1.1349 X_2Z_2 + 1.4418 X_3Z_2 + 4.0959X_4Z_2
\end{aligned}$$

The significant factors were X_1 , X_2 , X_3 , X_4 , $X_1 X_2$, $X_1 X_3$, $X_1 X_4$, $X_2 X_3$, $X_3 X_4$ and X_1Z_1 as shown in Table (4.20). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 50.7922 X_1 + 53.5255X_2 + 52.3922X_3 + 50.9255 X_4 \\
& + 113.8722 X_1X_2 + 120.0388 X_1X_3 + 114.3722 X_1X_4 \quad (4.15) \\
& + 118.4722 X_2X_3 + 111.4577 X_2X_4 + 99.1119 X_3X_4 \\
& + 8.0396 X_1Z_1
\end{aligned}$$

Table 4.21: The t student test for the fitted MPV model for the total yield of seeds in grams

Term	Estimate	Standard Error	t value	p value
X_1	8.9459	0.26262	34.06	0.0001
X_2	9.2126	0.26262	35.08	0.0001
X_3	9.4459	0.26262	35.97	0.0001
X_4	9.2126	0.26262	35.08	0.0001
X_1X_2	12.1977	0.64035	19.05	0.0001
X_1X_3	11.6311	0.64035	18.16	0.0001
X_1X_4	11.9977	0.64035	18.74	0.0001
X_2X_3	11.1977	0.64035	17.49	0.0001
X_2X_4	11.0337	0.71553	15.42	0.0001
X_3X_4	11.5295	0.62271	18.52	0.0001
X_1Z_1	0.7074	0.26594	2.66	0.0230
X_2Z_1	0.2008	0.26946	0.75	0.4718

X_3Z_1	0.5138	0.26616	1.93	0.0809
X_4Z_1	0.2526	0.26744	0.94	0.3657
X_1Z_2	0.1243	0.28910	0.43	0.6780
X_2Z_2	0.0890	0.29211	0.30	0.7678
X_3Z_2	0.2594	0.28887	0.90	0.3942
X_4Z_2	0.2605	0.29057	0.90	0.3946

The estimate, standard errors, t values and p values of the fitted Scheffe model for the total yield of seeds in grams of *Glycine max* per plant stem observed. The fitted Scheffe model is therefore,

$$\begin{aligned}
\hat{Y} = & 8.9459 X_1 + 9.2126X_2 + 9.4459X_3 + 9.2126 X_4 \\
& + 12.1977 X_1X_2 + 11.6311 X_1X_3 + 11.9977 X_1X_4 \quad (4.16) \\
& + 11.9977 X_2X_3 + 11.0337 X_2X_4 + 11.5295 X_3X_4 \\
& + 0.7074 X_1Z_1 + 0.2008 X_2Z_1 + 0.5138X_3Z_1 \\
& + 0.2526X_4Z_1 + 0.1243 X_1Z_2 + 0.0890 X_2Z_2 \\
& + 0.2594 X_3Z_2 + 0.2605X_4Z_2.
\end{aligned}$$

The significant factors were X_1 , X_2 , X_3 , X_4 , $X_1 X_2$, $X_1 X_3$, $X_1 X_4$, $X_2 X_3$, $X_3 X_4$ and X_1Z_1 as shown in Table (4.21). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 8.9459 X_1 + 9.2126X_2 + 9.4459X_3 + 9.2126 X_4 \\
& + 12.1977 X_1X_2 + 11.6311 X_1X_3 \quad (4.17) \\
& + 11.9977 X_1X_4 + 11.9977 X_2X_3 \\
& + 11.0337 X_2X_4 + 11.5295 X_3X_4 + 0.7074 X_1Z_1.
\end{aligned}$$

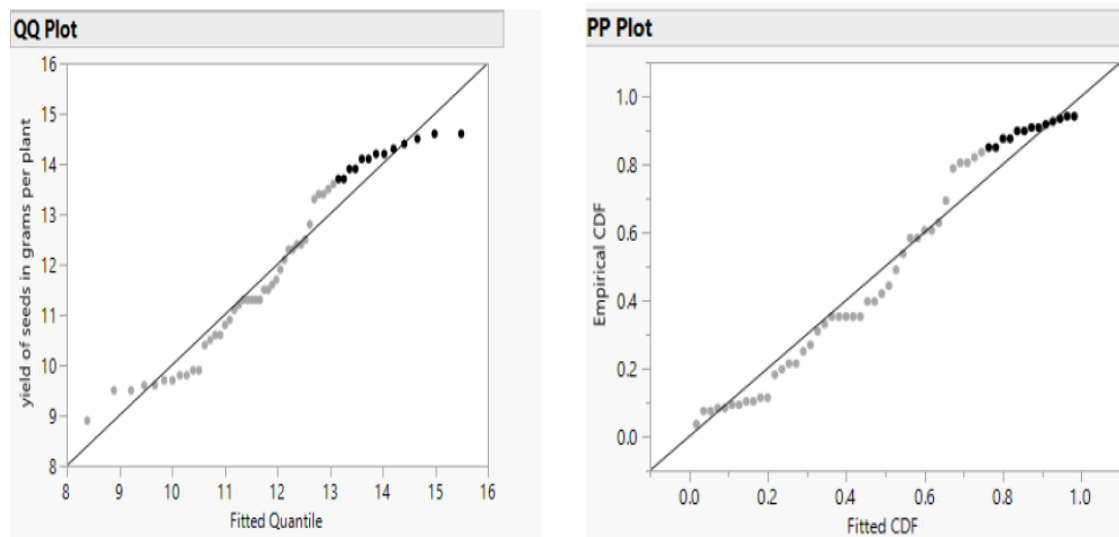


Figure 4.12: The QQ plot and PP plot for the yield of seeds in grams per plants

The QQ plot and PP plot in Figure 4.12 present diagnosis plots used to check the fitted model's adequacy. In the normal probability plot for QQ and PP plot, we may observe that there is no indication that normality assumption should not be accepted as most of the number of point lies on straight line. Therefore, we have no doubt to believe the data collected is normal distributed. However this also indicates that the second order modified MPV model 3.5 fits the data well.

4.3 The Screening Methodology for Estimating the Predicted Yield Using Simulation Technique in the Cox MPV Model.

This section outline the results of the predicted yield of Glycine max in a mixed environment in the presences of process variable in the framework of Cox Mixture Model using simulation technique.

4.3.1 Formulated of Cox MPV Model

This section outline the ANOVA Table based on the results in Table 4.1, analyzing MPV data using REML method in the context of Cox mixture model.

Table 4.22: The t student test for the fitted Cox MPV model for the No. of branches

Term	Estimate	Standard Error	t value	p value
Intercept	1.9024	0.0602	31.596	0.0001
X_1	0.2242	0.0741	3.024	0.0051
X_2	0.2076	0.0782	2.653	0.0126
X_3	-0.1952	0.0726	-2.687	0.0116
X_4	-0.2216	0.0741	-2.990	0.0055
X_1^2	-1.5005	0.1116	-13.447	0.0001
X_2^2	-1.5172	0.1151	-13.183	0.0001
X_3^2	-1.1812	0.1096	-10.780	0.0001
X_4^2	-1.1547	0.1114	4.426	0.0001
X_1X_2	0.8086	0.1827	6.093	0.0001
X_1X_3	1.0780	0.1769	6.265	0.0001
X_1X_4	1.1045	0.1797	6.145	0.0001
X_2X_3	1.1280	0.1800	6.265	0.0001
X_2X_4	0.9438	0.2097	4.501	0.0001
X_3X_4	0.2856	0.1724	1.657	0.1079
X_1Z_1	0.4189	0.08673	4.830	0.0001
X_2Z_1	0.3395	0.08908	3.811	0.0009
X_3Z_1	0.1776	0.08687	2.044	0.0534
X_4Z_1	0.0709	0.08774	0.808	0.4277
X_1Z_2	0.2484	0.0875	2.838	0.0128
X_2Z_2	0.1549	0.0897	1.727	0.1039
X_3Z_2	0.0554	0.0874	0.634	0.5361
X_4Z_2	0.0800	0.0886	0.903	0.3805

The estimate, standard errors, t values and p values of the fitted Cox polynomial model for the number of branches per plant observed. The fitted Cox mixture model is therefore,

$$\begin{aligned}
 \hat{Y} = & 1.9024 + 0.2242 X_1 + 0.2076X_2 - 0.1952X_3 - 0.2216 X_4 \\
 & - 1.5005X_1^2 - 1.517X_2^2 - 1.1812X_3^2 - 1.1547X_4^2 \\
 & + 0.8086 X_1X_2 + 1.0780 X_1X_3 + 1.1045 X_1X_4 \\
 & + 1.1280 X_2X_3 + 0.9438 X_2X_4 + 0.2856 X_3X_4 \\
 & + 0.4189 X_1Z_1 + 0.3395 X_2Z_1 + 0.1776 X_3Z_1 \\
 & + 0.0709 X_4Z_1 + 0.2484 X_1Z_2 + 0.1549 X_2Z_2 \\
 & + 0.0554 X_3Z_2 + 0.08 X_4Z_2
 \end{aligned} \tag{4.18}$$

The significant factors were $X_1, X_2, X_3, X_4, X_1^2, X_2^2, X_3^2, X_4^2, X_1 X_4, X_2 X_3, X_3 X_4$

, $X_1 Z_1, X_2 Z_1, X_1 X_2, X_1 X_3$ and $X_1 Z_2$ as shown in Table (4.22). Thus, the final model is

$$\begin{aligned}
 \hat{Y} = & 1.9024 + 0.2242 X_1 + 0.2076X_2 - 0.1952X_3 - 0.2216 X_4 \\
 & - 1.501X_1^2 - 1.517X_2^2 - 1.1812X_3^2 - 1.1547X_4^2 \\
 & + 0.8086 X_1X_2 + 1.0780 X_1X_3 + 1.1045 X_1X_4 \\
 & + 1.1280 X_2X_3 + 0.9438 X_2X_4 + 0.4189 X_1Z_1 \\
 & + 0.3395 X_2Z_1 + 0.2484 X_1Z_2
 \end{aligned} \tag{4.19}$$

Table 4.23: The t student test for the fitted Cox MPV model for Pods on branches

Term	Estimate	Standard Error	t value	p value
Intercept	8.8702	0.3950	22.458	0.0001
X_1	0.5079	0.4176	1.216	0.2333
X_2	0.3052	0.42351	0.7210	0.4767
X_3	-0.4552	0.4235	-1.075	0.2910
X_4	-0.3579	0.4176	-0.8570	0.3982
X_1^2	-7.7772	0.62709	-12.402	0.0001
X_2^2	-7.2412	0.63112	-11.474	0.0001
X_3^2	-6.5142	0.63105	-10.323	0.0001
X_4^2	-6.7781	0.62712	-10.808	0.0001
X_1X_2	5.8659	1.0155	5.776	0.0001
X_1X_3	4.6596	1.0022	4.649	0.0001
X_1X_4	5.0289	1.0094	4.982	0.0001
X_2X_3	4.2289	1.0094	4.189	0.0002
X_2X_4	4.3876	1.1654	3.765	0.0007
X_3X_4	4.1398	0.9766	4.239	0.0002
X_1Z_1	0.6569	0.5326	1.233	0.2344
X_2Z_1	0.2411	0.54474	0.4430	0.6633
X_3Z_1	0.6596	0.53337	1.237	0.2331
X_4Z_1	0.42571	0.5378	0.7920	0.4393
X_1Z_2	0.2478	0.5511	0.4500	0.6611
X_2Z_2	-0.1548	0.5620	-0.2750	0.7875
X_3Z_2	-0.01998	0.5503	-0.0360	0.9716
X_4Z_2	0.02288	0.5565	0.0410	0.9679

The estimate, standard errors, t values and p –values of the fitted Cox polynomial model for the number of pods on branches of Glycine max per plant stem observed. The fitted Cox mixture model is therefore,

$$\begin{aligned}
\hat{Y} = & 8.8702 + 0.5079 X_1 + 0.3052X_2 - 0.4552X_3 - 0.3579 X_4 \\
& - 7.7772X_1^2 - 7.2412X_2^2 - 6.5142X_3^2 - 6.7781X_4^2 \\
& + 5.8659 X_1X_2 + 4.6596 X_1X_3 + 5.0289 X_1X_4 \\
& + 4.2289 X_2X_3 + 4.3876 X_2X_4 + 4.1398 X_3X_4 \\
& + 0.6569 X_1Z_1 + 0.2411 X_2Z_1 + 0.6596 X_3Z_1 \\
& + 0.42571X_4Z_1 + 0.2478 X_1Z_2 - 0.1548 X_2Z_2 \\
& - 0.01998 X_3Z_2 + 0.02288 X_4Z_2
\end{aligned} \tag{4.20}$$

The significant factors were X_1^2 , X_2^2 , X_3^2 , $X_1 X_4$, $X_2 X_3$, $X_3 X_4$, $X_1 X_2$, $X_1 X_3$, and X_4^2 as shown in Table (4.23). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 8.8702 - 7.7772X_1^2 - 7.2412X_2^2 - 6.5142X_3^2 - 6.7781X_4^2 \\
& + 5.8659 X_1X_2 + 4.6596 X_1X_3 + 5.0289 X_1X_4 \\
& + 4.2289 X_2X_3 + 4.3876 X_2X_4 + 4.1398 X_3X_4
\end{aligned} \tag{4.21}$$

Table 4.24: The t student test for the fitted Cox MPV model for the No. of Pods on Main stem

Term	Estimate	Standard Error	t value	p value
Intercept	29.1417	0.8258	35.289	0.0001
X_1	-0.1395	0.7821	-0.178	0.8596
X_2	-0.2503	0.8254	-0.303	0.7637
X_3	-0.0343	0.7664	-0.045	0.9646
X_4	0.4161	0.7821	0.532	0.5986
X_1^2	-8.7034	1.1788	-7.383	0.0001
X_2^2	-7.4926	1.2158	-6.163	0.0001
X_3^2	-7.4087	1.1576	-6.400	0.0001
X_4^2	-9.7257	1.1767	-8.265	0.0001
X_1X_2	2.1645	1.9278	1.123	0.2704
X_1X_3	5.6485	1.8671	3.025	0.0050
X_1X_4	9.4647	1.8967	4.990	0.0001
X_2X_3	5.9593	1.9000	3.137	0.0038
X_2X_4	6.0857	2.2128	2.750	0.0100
X_3X_4	3.9791	1.8200	2.186	0.0366
X_1Z_1	2.2803	1.0724	2.126	0.0509
X_2Z_1	1.2280	1.0937	1.123	0.2785
X_3Z_1	0.6803	1.0737	0.634	0.5360
X_4Z_1	2.0735	1.0815	1.917	0.0744
X_1Z_2	1.5415	1.1254	1.368	0.1996
X_2Z_2	2.9897	1.1458	2.609	0.0239
X_3Z_2	2.2016	1.1254	1.956	0.0774
X_4Z_2	2.0861	1.1361	1.836	0.0937

The estimate, standard errors, t values and p –values of the fitted Cox MPV model for the number of pods on main stem of Glycine max per plant stem observed. The fitted Cox mixture model is therefore,

$$\begin{aligned}
 \hat{Y} = & 29.1417 - 0.1395 X_1 - 0.2503X_2 - 0.0343X_3 + 0.4161 X_4 \\
 & - 8.7034X_1^2 - 7.4926X_2^2 - 7.40872X_3^2 - 9.72571X_4^2 \\
 & + 2.1645 X_1X_2 + 5.6485 X_1X_3 + 9.4647 X_1X_4 \\
 & + 5.9593 X_2X_3 + 6.0857 X_2X_4 + 3.9791 X_3X_4 \\
 & + 2.2803 X_1Z_1 + 1.2280 X_2Z_1 + 0.6803 X_3Z_1 \\
 & + 2.07351X_4Z_1 + 1.5415 X_1Z_2 + 2.9897 X_2Z_2 \\
 & + 2.2016 X_3Z_2 + 2.0861 X_4Z_2
 \end{aligned} \tag{4.22}$$

The significant factors were X_1^2 , X_2^2 , X_3^2 , $X_1 X_4$, $X_2 X_3$, $X_3 X_4$, $X_1 X_3$, X_2Z_2 , X_4^2 and intercept as shown in Table (4.24). Thus, the final model is

$$\begin{aligned}
 \hat{Y} = & 29.1417 - 8.7034X_1^2 - 7.4926X_2^2 - 7.40872X_3^2 - 9.72571X_4^2 \\
 & + 5.6485 X_1X_3 + 9.4647 X_1X_4 + 5.9593 X_2X_3 \\
 & + 6.0857 X_2X_4 + 3.9791 X_3X_4 + 2.9897 X_2Z_2
 \end{aligned} \tag{4.23}$$

Table 4.25: The t student test for the fitted Cox MPV model for the No. of Pods per branch

Term	Estimate	Standard Error	t value	p value
Intercept	4.4537	0.0919	48.464	0.0001
X_1	0.03169	0.1055	0.300	0.7660
X_2	0.1059	0.1113	0.951	0.3494
X_3	0.0948	0.1034	0.917	0.3663
X_4	-0.2350	0.1055	-2.198	0.0358
X_1^2	-2.1150	0.1589	-13.309	0.0001
X_2^2	-2.3892	0.1639	-14.577	0.0001
X_3^2	-2.2448	0.1560	-14.386	0.0001
X_4^2	-1.8181	0.1586	-11.461	0.0001
X_1X_2	1.4804	0.2601	5.693	0.0001
X_1X_3	1.6581	0.2519	6.583	0.0001
X_1X_4	1.0849	0.2559	4.240	0.0002
X_2X_3	1.7840	0.2563	6.961	0.0001
X_2X_4	1.2709	0.2985	4.258	0.0002
X_3X_4	1.2799	0.2455	5.214	0.0001
X_1Z_1	0.3144	0.1283	2.450	0.0240
X_2Z_1	0.1553	0.1316	1.180	0.2514
X_3Z_1	0.1936	0.1285	1.506	0.1482
X_4Z_1	0.2618	0.1297	2.018	0.0574
X_1Z_2	0.2934	0.1310	2.239	0.0430
X_2Z_2	0.0961	0.1340	0.717	0.4849
X_3Z_2	0.2268	0.1308	1.734	0.1065
X_4Z_2	0.2113	0.1325	1.595	0.1336

The estimate, standard errors, t values and p values of the fitted Cox polynomial model for the number of pods per branch of *Glycine max* per plant stem observed. The fitted Cox mixture model is therefore,

$$\begin{aligned}
\hat{Y} = & 4.4537 + 0.03169 X_1 + 0.1059X_2 + 0.0948X_3 - 0.2350 X_4 \\
& - 2.1150X_1^2 - 2.3892X_2^2 - 2.2448X_3^2 - 1.8181X_4^2 \\
& + 1.4804 X_1X_2 + 1.6581 X_1X_3 + 1.0849 X_1X_4 \\
& + 1.7840 X_2X_3 + 1.2709 X_2X_4 + 1.2799 X_3X_4 \\
& + 0.3144 X_1Z_1 + 0.1553 X_2Z_1 + 0.1936 X_3Z_1 \\
& + 0.2618X_4Z_1 + 0.2934 X_1Z_2 + 0.0961 X_2Z_2 \\
& + 0.2268 X_3Z_2 + 0.2113 X_4Z_2
\end{aligned} \tag{4.24}$$

The significant factors were $X_4, X_1^2, X_2^2, X_3^2, X_1 X_4, X_2 X_3, X_3 X_4, X_1 X_3, X_1 Z_1, X_4^2, X_1 Z_2$ and intercept as shown in Table (4.25). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 4.4537 - 0.2350 X_4 - 2.1150X_1^2 - 2.3892X_2^2 - 2.2448X_3^2 \\
& - 1.8181X_4^2 + 1.4804 X_1X_2 + 1.6581 X_1X_3 + 1.0849 X_1X_4 \\
& + 1.7840 X_2X_3 + 1.2709 X_2X_4 + 1.2799 X_3X_4 \\
& + 0.3144 X_1Z_1 + 0.2934 X_1Z_2
\end{aligned} \tag{4.25}$$

Table 4.26: The t student test for the fitted Cox MPV model for the total Pods per plant

Term	Estimate	Standard Error	t value	p value
Intercept	38.0156	1.0628	35.771	0.0001
X_1	0.3800	0.6671	0.570	0.5731
X_2	0.2077	0.7041	0.295	0.7700
X_3	-0.6108	0.6538	-0.934	0.3576
X_4	0.0534	0.6671	0.080	0.9367
X_1^2	-16.5216	1.006	-16.417	0.0001
X_2^2	-14.9159	1.03788	-14.372	0.0001
X_3^2	-13.8307	0.9882	-13.995	0.0001
X_4^2	-16.5283	1.0045	-16.455	0.0001
X_1X_2	7.9252	1.6446	4.819	0.0001
X_1X_3	10.4771	1.5928	6.578	0.0001
X_1X_4	14.5462	1.6180	8.990	0.0001
X_2X_3	10.216	1.6208	6.303	0.0001
X_2X_4	10.2073	1.8879	5.407	0.0001
X_3X_4	8.3710	1.5527	5.391	0.0001
X_1Z_1	2.9268	1.2446	2.352	0.0420
X_2Z_1	1.4991	1.2580	1.192	0.2615
X_3Z_1	1.3146	1.2454	1.056	0.3175
X_4Z_1	2.5141	1.2503	2.011	0.0733
X_1Z_2	1.7823	1.3687	1.302	0.2299
X_2Z_2	2.8601	1.3801	2.072	0.0716
X_3Z_2	2.1625	1.3679	1.581	0.1535
X_4Z_2	2.1221	1.3743	1.544	0.1614

The estimate, standard errors, t values and p values of the fitted Cox polynomial model for the total number of pods of *Glycine max* per plant stem observed. The fitted Cox mixture model is therefore,

$$\begin{aligned}
\hat{Y} = & 38.0156 + 0.3800 X_1 + 0.2077X_2 - 0.6108X_3 + 0.0534 X_4 \\
& - 16.5216X_1^2 - 14.9159X_2^2 - 13.8307X_3^2 - 16.5283X_4^2 \\
& + 7.9252 X_1X_2 + 10.4771 X_1X_3 + 14.5462 X_1X_4 \\
& + 10.216 X_2X_3 + 10.2073 X_2X_4 + 8.3710 X_3X_4 \\
& + 2.9268 X_1Z_1 + 1.4991 X_2Z_1 + 1.3146 X_3Z_1 \\
& + 2.5141 X_4Z_1 + 1.7823 X_1Z_2 + 2.8601 X_2Z_2 \\
& + 2.1625 X_3Z_2 + 2.1221X_4Z_2
\end{aligned} \tag{4.26}$$

The significant factors were X_1^2 , X_2^2 , X_3^2 , $X_1 X_4$, $X_2 X_3$, $X_3 X_4$, $X_1 X_3$, X_4^2 , X_1Z_1 and intercept as shown in Table (4.26). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 38.0156 - 16.5216X_1^2 - 14.9159X_2^2 - 13.8307X_3^2 - 16.5283X_4^2 \\
& + 7.9252 X_1X_2 + 10.4771 X_1X_3 + 14.5462 X_1X_4 \\
& + 10.216 X_2X_3 + 10.2073 X_2X_4 + 8.3710 X_3X_4 \\
& + 2.9268 X_1Z_1
\end{aligned} \tag{4.27}$$

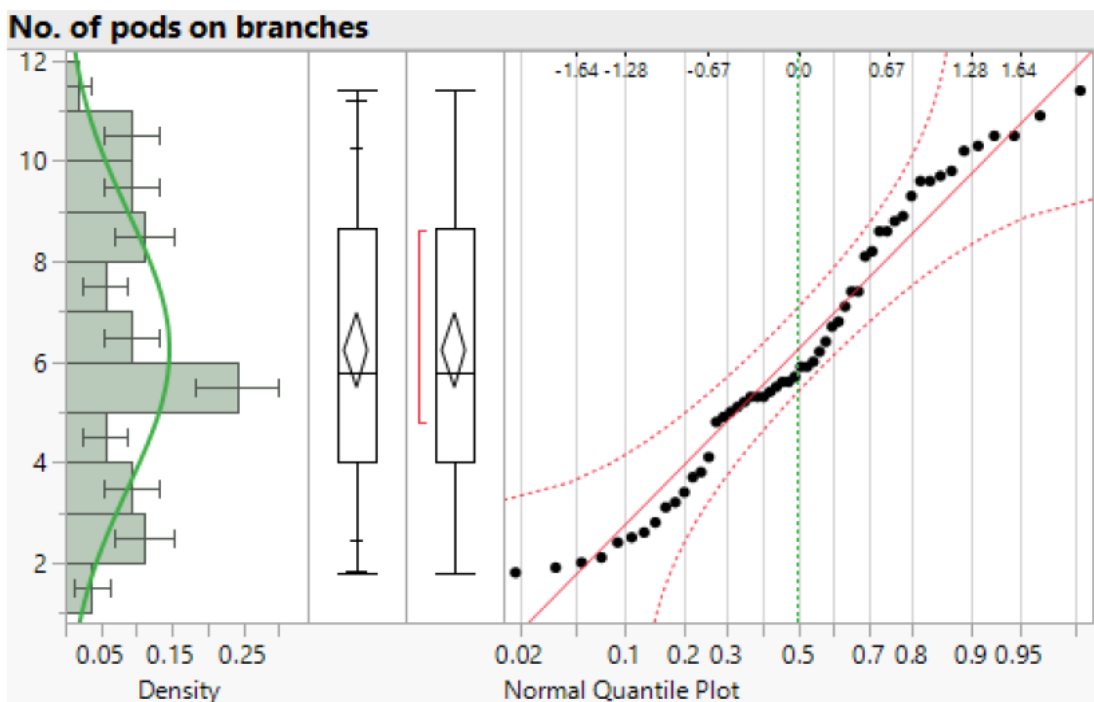


Figure 4.13: Histogram, Box plot and Normal Quantile plot fitted for the number of pods on branches

The Histogram, Box plot, and Normal Quantile plot as shown in Figure 4.13 was to check the adequacy SPD in fitting MPV data in the context of Cox mixture model. It was found out that the number pods development on branches observed satisfied the normality property as most of the points lie on a straight line, as evidenced from the Normal Quantile plot. These data on the number of pod development on branches per plant shows that it is normally distributed when fitted using a normal distribution with a 95 % confidence interval, as shown in Table 4.27. The fitted normal distribution data was also represented plot shown in Figure 4.14. This Figure 4.14 shows that the maximum yield for *Glycine max* achieved was 16.2 with probability 0.981775 MPV settings having well been applied within SPD. However, this Figure 4.14 was also was used to analyze the reliability property

of design where the same result was checked using the goodness of fit test, in which the data collected was found to be fitted with a p-value of 0.0459 using the Shapiro-Wilk (SW) test with test statistics equal to $W = 0.956052$. The SW test tests the null hypothesis that sample X_1, X_2, \dots, X_n come from a normal distributed data with test statistics being defined as

$$W = \frac{\sum_{i=1}^n \delta_i x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (4.28)$$

where x_i and \bar{x} represents the i^{th} order statistics and sample mean respectively. The coefficients δ_i are given by:

$$(\delta_1, \delta_2, \dots, \delta_n) = \frac{\eta' V^{-1}}{\phi}, \quad (4.29)$$

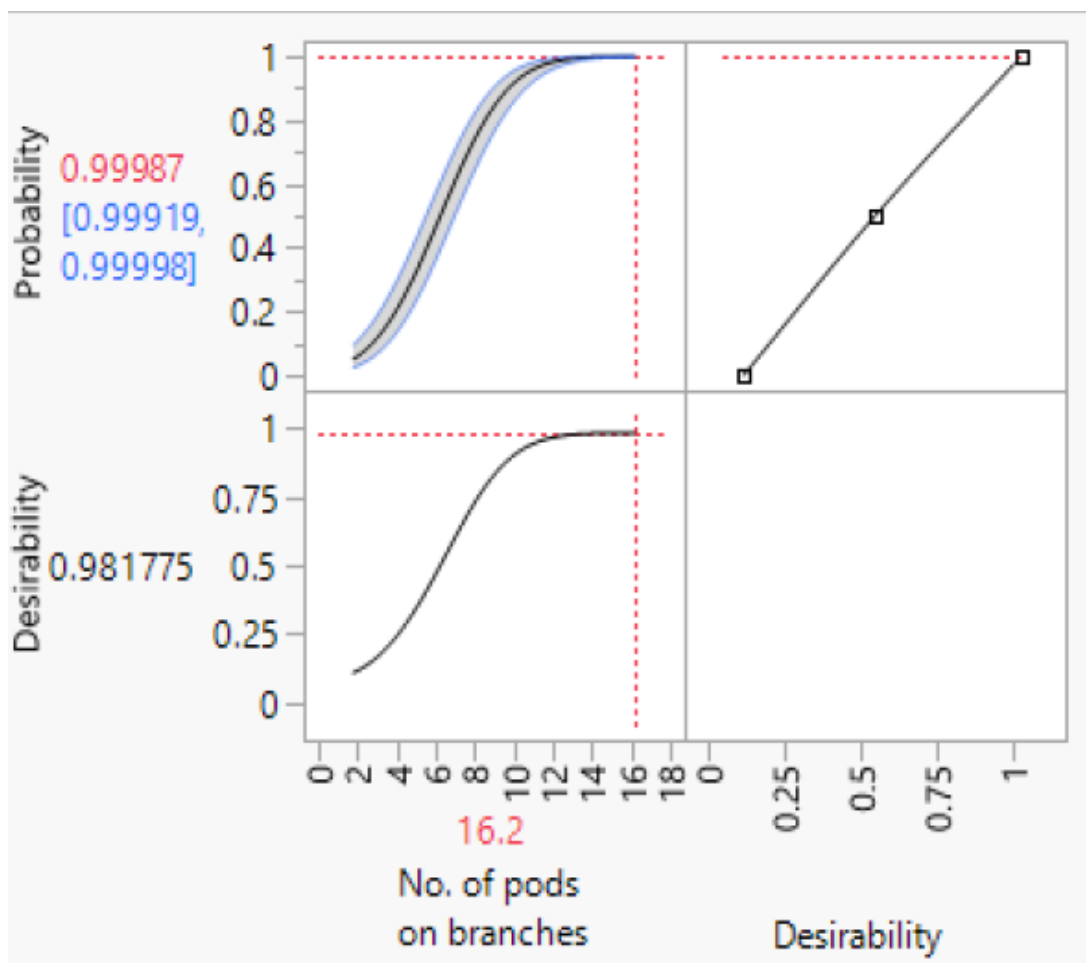
where ϕ is the vector norm:

$$\phi = \|V^{-1}\eta'\| = (\eta' V^{-1} V^{-1} \eta)^{\frac{1}{2}}, \quad (4.30)$$

and vector $\eta = (\eta_1, \eta_2, \dots, \eta_n)'$ constitute the expected values of the order statistics of independent and identically distributed random variables ($\eta = E(a_i)$) sampled from standard normal distribution as described by Shapiro (1965); finally, V represents the covariance matrix of those normal order statistics. This SW test statistics (W) was computed through Monte Carlo simulations in cooperated in JMP 15 software.

Table 4.27: The fitted Normal distribution for total pods on branches

Parameter	Estimate	Standard error	Lower 95%	Upper 95%
Location μ	6.2389	0.3711	5.512	6.966
Dispersion σ	2.727	0.1611	2.429	2.826
Measures				
	-2*Loglikelihood	260.596		
	AICc	264.831		
	BIC	268.574		

**Figure 4.14: The distribution profiler for the normal distribution for the number of pods on branches**

The maximum predicted number of branches was 16.2 as shown in Figure 4.14. The horizontal dotted line represents the desirable probability required to achieve the optimum yield with MPV settings within SPD. On the other hand the vertical dotted line shows the desirable number of pod to be attained with the corresponding probability of 99.987% with confidence interval.

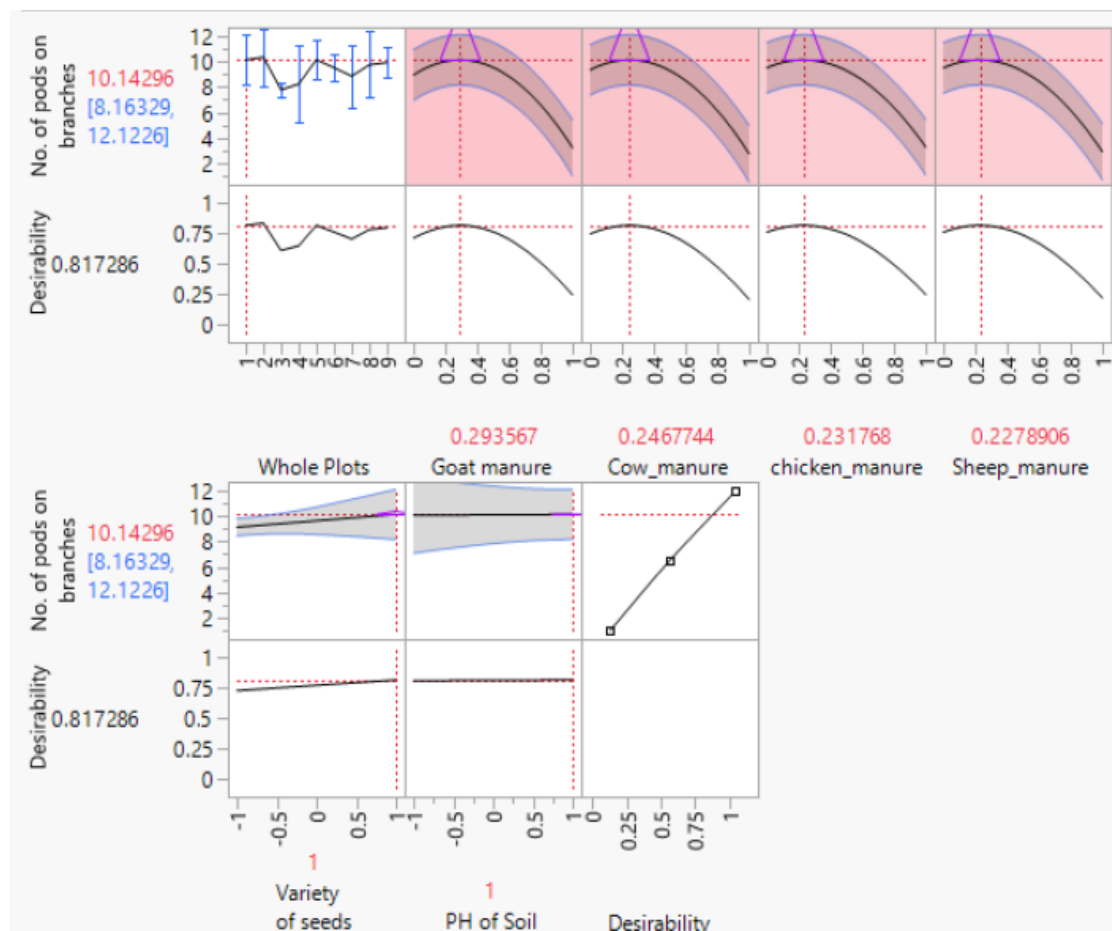


Figure 4.15: The response prediction for the number of pods on a branches subject to mixture setting in the presnces process variable.

Figure 4.15 shows the comparison between the experimental and predicted values. From these Figure, it can be observed that the maximum desirable predicted number of pods on

branches is an averagely of 10.14296 with a 95 % confidence level. The predicted mixture process variable setting is indicated with a vertical dotted line, while the horizontal dotted line indicates the predicted response value. The blue line at maximum response predicted value shows the adjusted MPV setting within SPD can be done to the maximum or minimum value of the response at 95% confidence interval level as shown in Figure 4.15. The whole plot graph is also represented on the response prediction profiler showing the whole plot with the highest predicted response as evidenced from Figure 4.15.

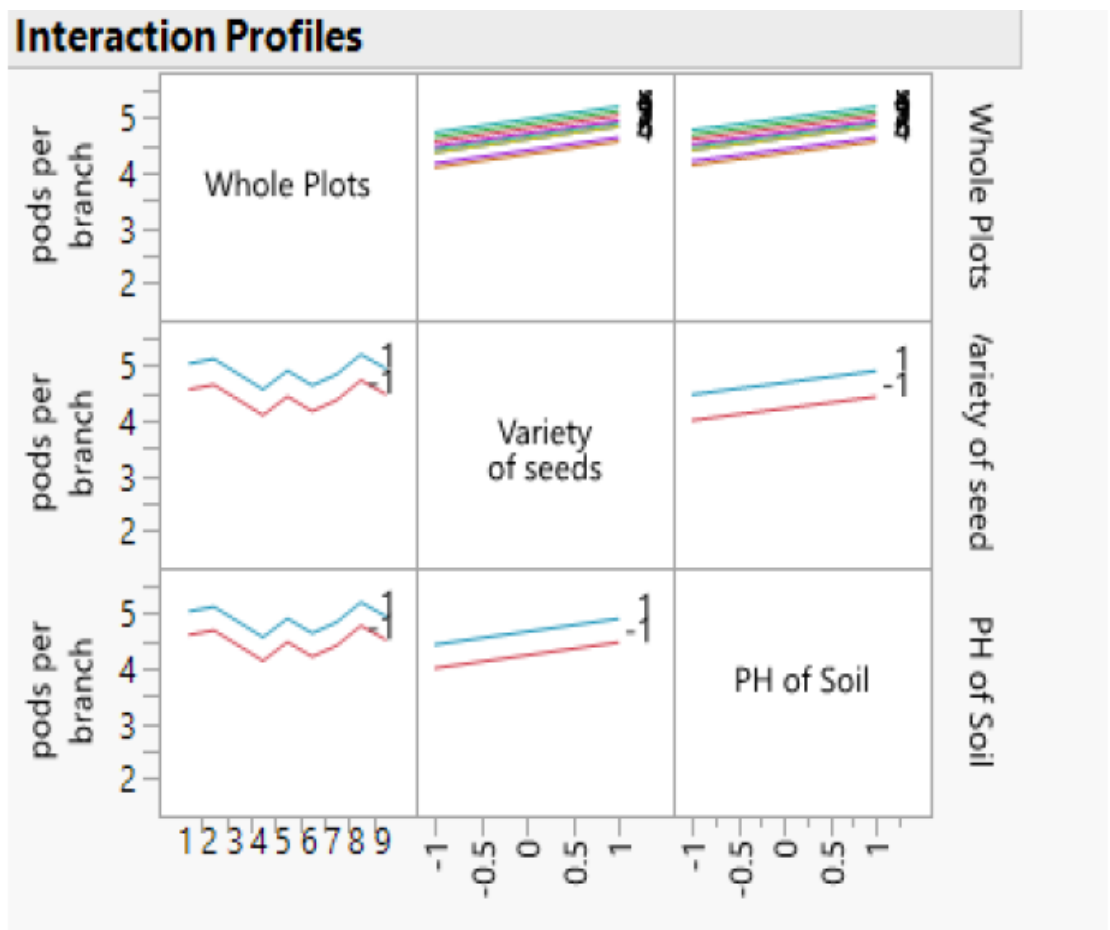


Figure 4.16: The interaction plot resulting from MPV setting in the framework of SPD

From Figure 4.16 it was observed that either plot shows the effects of mixture variable on the predicted response (Number of pods on branches) is fairly constant at a low level of the PH of soil,

whether the variety of seeds is set at a high or low level. This similar scenario is also illustrated in response prediction profiler subject to MPV settings within SPD, as shown in Figures 4.15. However, at a wide or low-level variety of seeds (in terms of row spacing of the plant), the mixture factor component's effect on Y differs based on the mixture setting applied. Variety of +1 leads to a higher predicted $Y(x, z)$ than a variety of seed at -1. In addition, it can be observed the interaction plot illustrates the interaction of the row effect with column effects. However, A line segment shown in Figure 4.15 and 4.16 joins the response value predicted by the model whereas non-parallel

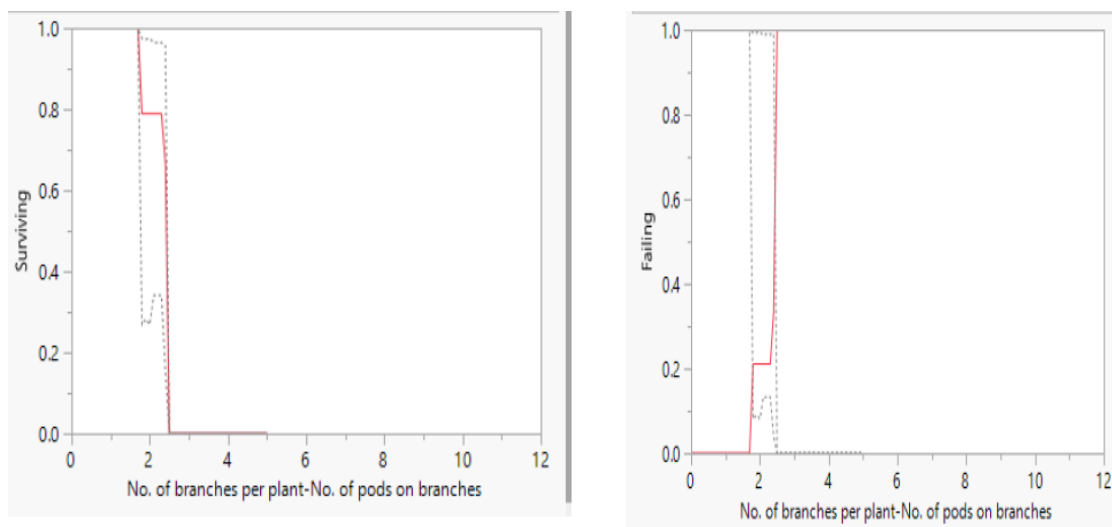


Figure 4.17: The overlaid survival and failure plots for each group (number of branches and pods development per plant stem)

Figure 4.17 shows two graphs; the overlaid survival plots for each group (number of branches and pods development per plant stem). The graphs indicate at least 2.5 branches and pods development per plant stem will survive when MPV settings are well applied. The second graph shows each group's overlaid failure plot (proportion failing over time), which is a similar survival plot. The graph indicates that at most, two branches or pods per plant stem might not well be captured when plotting the recorded experimental field data. However, Figure 4.17 also shows that the second order Cox MPV model was well

fitted with minimal partial information missing as evidenced from failure plots although necessary measures need to be taken by experimenter.

Table 4.28: The t student test for the fitted Cox MPV model for seeds per pod

Term	Estimate	Standard Error	t value	p value
Intercept	2.49262	0.02512	99.228	0.0001
X_1	-0.00786	0.02660	-0.295	0.7697
X_2	0.05649	0.02808	2.012	0.0532
X_3	-0.02875	0.02607	-1.103	0.2789
X_4	-0.01672	0.02660	-0.629	0.5344
X_1^2	-0.42506	0.04008	-10.604	0.0001
X_2^2	-0.48941	0.04134	-11.838	0.0001
X_3^2	-0.40417	0.03936	-10.268	0.0001
X_4^2	-0.4162	0.04001	-10.402	0.0001
X_1X_2	0.29741	0.06558	4.535	0.0001
X_1X_3	0.28265	0.06351	4.450	0.0001
X_1X_4	0.27062	0.06452	4.195	0.0002
X_2X_3	0.31829	0.06463	4.925	0.0001
X_2X_4	0.31509	0.07527	4.186	0.0002
X_3X_4	0.24914	0.06190	4.025	0.0004
X_1Z_1	0.07191	0.03396	2.118	0.0491
X_2Z_1	0.04264	0.03473	1.228	0.2352
X_3Z_1	0.06336	0.03400	1.863	0.0796
X_4Z_1	0.07954	0.03430	2.320	0.0326
X_1Z_2	0.04308	0.03511	1.227	0.2434
X_2Z_2	0.00966	0.03581	0.270	0.7916
X_3Z_2	0.04624	0.03505	1.319	0.2120
X_4Z_2	0.04918	0.03545	1.387	0.1898

The estimate, standard errors, t values and p values of the fitted Cox polynomial model for the number of seeds per pod of *Glycine max* per plant stem observed. The fitted Cox mixture model is therefore,

$$\begin{aligned}
 \hat{Y} = & 2.49262 - 0.00786 X_1 + 0.05649X_2 - 0.02875X_3 - 0.01672 X_4 \\
 & - 0.42506X_1^2 - 0.48941X_2^2 - 0.40417X_3^2 - 0.4162X_4^2 \\
 & + 0.29741 X_1X_2 + 0.28265 X_1X_3 + 0.27062 X_1X_4 \quad (4.31) \\
 & + 0.31829 X_2X_3 + 0.31509 X_2X_4 + 0.24914 X_3X_4 \\
 & + 0.07191 X_1Z_1 + 0.04264 X_2Z_1 + 0.06336 X_3Z_1 \\
 & + 2.5141 X_4Z_1 + 0.04308 X_1Z_2 + 0.00966 X_2Z_2 \\
 & + 0.04624 X_3Z_2 + 0.04918X_4Z_2
 \end{aligned}$$

The significant factors were X_1^2 , X_2^2 , X_3^2 , $X_1 X_4$, $X_2 X_3$, $X_3 X_4$, $X_1 X_3$, X_4^2 , X_1Z_1 , X_4Z_1 and intercept as shown in Table (4.28). Thus, the final model is

$$\begin{aligned}
 \hat{Y} = & 2.49262 - 0.42506X_1^2 - 0.48941X_2^2 - 0.40417X_3^2 \\
 & - 0.4162X_4^2 + 0.29741 X_1X_2 + 0.28265 X_1X_3 \quad (4.32) \\
 & + 0.27062 X_1X_4 + 0.31829 X_2X_3 + 0.31509 X_2X_4 \\
 & + 0.24914 X_3X_4 + 0.07191 X_1Z_1 + 2.5141 X_4Z_1
 \end{aligned}$$

Table 4.29: The t student test for the fitted Cox MPV model for the No. of seeds per plant

Term	Estimate	Standard Error	t value	p value
Intercept	94.2067	1.8939	49.743	0.0001
X_1	1.4050	2.3072	0.609	0.5471
X_2	4.0579	2.4347	1.667	0.1059
X_3	-0.8141	2.2608	-0.360	0.7213
X_4	-4.4684	2.3072	-1.937	0.0622
X_1^2	-4.4.8195	3.4733	-12.904	0.0001
X_2^2	-44.7391	3.5823	-12.489	0.0001
X_3^2	-41.004	3.4105	-12.022	0.0001
X_4^2	-38.8128	3.4671	-11.195	0.0001
X_1X_2	24.3135	5.6859	4.276	0.0002
X_1X_3	34.2189	5.5068	6.214	0.0001
X_1X_4	30.7398	5.5942	5.495	0.0001
X_2X_3	32.7326	5.6037	5.841	0.0001
X_2X_4	27.9058	6.5256	4.276	0.0002
X_3X_4	19.2987	5.3667	3.596	0.0011
X_1Z_1	8.0396	2.7149	2.961	0.0074
X_2Z_1	2.8369	2.7877	1.018	0.3197
X_3Z_1	4.4079	2.7193	1.621	0.1199
X_4Z_1	4.7505	2.7462	1.730	0.0979
X_1Z_2	4.6172	2.7449	1.682	0.1142
X_2Z_2	1.1349	2.8121	0.404	0.6920
X_3Z_2	1.4418	2.7397	0.526	0.6068
X_4Z_2	4.0959	2.7780	1.474	0.1612

The estimate, standard errors, t values and p values of the fitted Cox polynomial model for the number of seeds per pod of *Glycine max* per plant stem observed. The fitted Cox mixture model is therefore,

$$\begin{aligned}
\hat{Y} = & 94.2067 + 1.4050 X_1 + 4.0579X_2 - 0.8141X_3 - 4.4684 X_4 \\
& - 4.48195X_1^2 - 44.7391X_2^2 - 41.004X_3^2 - 38.8128X_4^2 \\
& + 24.3135 X_1X_2 + 34.2189 X_1X_3 + 30.7398 X_1X_4 \\
& + 32.7326 X_2X_3 + 27.9058 X_2X_4 + 19.2987 X_3X_4 \\
& + 8.0396 X_1Z_1 + 2.8369 X_2Z_1 + 4.4079 X_3Z_1 \\
& + 4.7505 X_4Z_1 + 4.6172 X_1Z_2 + 1.1349 X_2Z_2 \\
& + 1.4418 X_3Z_2 + 4.0959X_4Z_2
\end{aligned} \tag{4.33}$$

The significant factors were X_1^2 , X_2^2 , X_3^2 , $X_1 X_4$, $X_2 X_3$, $X_3 X_4$, $X_1 X_3$, X_4^2 , X_1Z_1 , and intercept as shown in Table (4.29). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 94.2067 - 4.48195X_1^2 - 44.7391X_2^2 - 41.004X_3^2 - 38.8128X_4^2 \\
& + 24.3135 X_1X_2 + 34.2189 X_1X_3 + 30.7398 X_1X_4 \\
& + 32.7326 X_2X_3 + 27.9058 X_2X_4 + 19.2987 X_3X_4 \\
& + 8.0396 X_1Z_1
\end{aligned} \tag{4.34}$$

Table 4.30: The t student test for the fitted Cox MPV model for the total yield of seeds in grams

Term	Estimate	Standard Error	t value	p value
Intercept	13.5545	0.2214	61.212	0.0001
X_1	-0.0053	0.1577	-0.034	0.9734
X_2	0.0210	0.1665	0.126	0.9006
X_3	1.0292	0.1546	0.189	0.0414
X_4	-0.0452	0.1577	-0.286	0.7765
X_1^2	-4.6033	0.2379	-19.349	0.0001
X_2^2	-4.3629	0.2454	-17.782	0.0001
X_3^2	-4.1378	0.2336	-17.711	0.0001
X_4^2	-4.2967	0.2375	-18.094	0.0001
X_1X_2	3.2316	0.3888	8.311	0.0001
X_1X_3	2.8900	0.3766	7.674	0.0001
X_1X_4	3.0977	0.3826	8.097	0.0001
X_2X_3	2.6970	0.3832	7.038	0.0001
X_2X_4	2.3741	0.4463	5.319	0.0001
X_3X_4	3.0950	0.3671	8.431	0.0001
X_1Z_1	0.7074	0.26590	2.660	0.0230
X_2Z_1	0.2008	0.2695	0.745	0.4718
X_3Z_1	0.5138	0.2662	1.931	0.0809
X_4Z_1	0.2526	0.2674	0.945	0.3657
X_1Z_2	0.1243	0.2891	0.430	0.6780
X_2Z_2	0.0890	0.2921	0.305	0.7678
X_3Z_2	0.2594	0.2889	0.898	0.3942
X_4Z_2	0.2605	0.2906	0.896	0.3946

The estimate, standard errors, t values and p values of the fitted Cox polynomial model for the total number of seeds in grams of *Glycine max* per plant stem observed. The fitted Cox mixture model is therefore,

$$\begin{aligned}
\hat{Y} = & 13.5545 - 0.0053 X_1 + 0.0210X_2 + 1.0292X_3 - 0.0452 X_4 \\
& - 4.6033X_1^2 - 4.3629X_2^2 - 41.004X_3^2 - 4.2967X_4^2 \\
& + 3.2316 X_1X_2 + 2.8900 X_1X_3 + 3.0977 X_1X_4 \\
& + 2.6970 X_2X_3 + 2.3741 X_2X_4 + 3.0950 X_3X_4 \\
& + 0.7074 X_1Z_1 + 0.2008 X_2Z_1 + 0.5138 X_3Z_1 \\
& + 0.2526 X_4Z_1 + 0.1243 X_1Z_2 + 0.0890 X_2Z_2 \\
& + 0.2594 X_3Z_2 + 0.2605X_4Z_2
\end{aligned} \tag{4.35}$$

The significant factors were X_1^2 , X_2^2 , X_3^2 , $X_1 X_4$, $X_2 X_3$, $X_3 X_4$, $X_1 X_3$, X_4^2 , X_3 , X_1Z_1 , and intercept as shown in Table (4.30). Thus, the final model is

$$\begin{aligned}
\hat{Y} = & 13.5545 + 1.0292X_3 - 4.6033X_1^2 - 4.3629X_2^2 - 4.1004X_3^2 \\
& - 4.2967X_4^2 + 3.2316 X_1X_2 + 2.8900 X_1X_3 \\
& + 3.0977 X_1X_4 + 2.6970 X_2X_3 + 2.3741 X_2X_4 \\
& + 3.0950 X_3X_4 + 0.7074 X_1Z_1
\end{aligned} \tag{4.36}$$

4.3.2 Screening in a MPV Settings

In this section, we discuss how screening in mixture settings was done in the presence of a process variable. The screening methodology was done using the framework of the Cox mixture model approach. The Cox mixture model technique was applied because it provides the experimenter with detailed information about what will happen to the response when an incremental change is made in any direction from some mixture, as demonstrated in the response prediction profiler shown in Figures 47. This type of insight cannot be had when the Scheffe model approach is applied since there is no direct way to cooperate with

the current mixture. Also, the model's parameter does not describe the change in the predicted response, as pointed out by Cornell (2011).

However, we can illustrate this insight using one of the presented results of eight responses of *Glycine max* measured and described in ANOVA Table based on data from Table 20, analyzing mixture process data. We estimated for Scheffe response for the total yield of seeds in grams of *Glycine max* per plant using REML method as

$$\begin{aligned}\hat{Y}(a) = & 8.9459 X_1 + 9.2126X_2 + 9.4459X_3 + 9.2126 X_4 + 12.1977 X_1X_2 \\ & + 11.6311 X_1X_3 + 11.9977 X_1X_4 + 11.9977 X_2X_3 \\ & + 11.0337 X_2X_4 + 11.5295 X_3X_4 + 0.7074 X_1Z_1\end{aligned}\quad (4.37)$$

The estimate for the Cox mixture polynomial model with $c = (0.25, 0.25, 0.25, 0.25)$ for the same response are

$$\begin{aligned}\hat{Y}(x, z) = & 94.2067 + 1.0292X_3 - 4.6033X_1^2 - 4.3629X_2^2 - 4.1004X_3^2 \\ & - 4.2967X_4^2 + 3.2316 X_1X_2 + 2.8900 X_1X_3 + 3.0977 X_1X_4 \\ & + 2.6970 X_2X_3 + 2.3741 X_2X_4 + 3.0950 X_3X_4 \\ & + 0.7074 X_1Z_1\end{aligned}\quad (4.38)$$

From (4.37), we see that the highest response will be an interaction binary mixture blend of components X_1 and X_2 , which is also readily apparent from the fitted Cox mixture model approach (4.38). However, we know that at c , the response of the total yield of seeds of *Glycine max* in grams to our MPV within a split-plot design is 94.2067g. Therefore, adding X_3 , binary mixture interaction ($X_1X_3, X_1X_4, X_2X_3, X_2X_4, X_3X_4$) and taking away quadratic mixture component vertices of component X_1, X_2, X_3 , and X_4 will increase the response. We can also observe that using the Scheffe model estimates, this type of insight

is not easily had as described in Hassan *et al.* (2020). To find out that the Cox compound polymorphism model referred to the deficiency of a stable compound, we directly compared the maximum growth of the current branch and the changes made with the pod growth on the main stem and branches, as shown in Figures 47 and 52. We can note that it is the most efficient way to experiment in an industrial and agricultural setting. Additionally, the Cox model aids the experimenter in quantifying the effect of individual components, either one or a few, but not necessarily all of the ingredients in the mixture.

4.3.3 The Joint Factor Tests of MPV Using the Cox Reference Mixture Model

This section outline the effect of MPV used in the study on the growth and pod development of *Glycine max*. The results are shown in Table 44 using the joint factor test described in Goos and Jones (2007) using the framework of Cox mixture model. The impact of each mixture components together with process variable are indicated with *p*-value at 5% significance level.

Table 4.31: The ANOVA for the joint Factor Tests of MPV

Term	Nparm	DF	DFDen	F Ratio	Prop > F
Whole plot	0
Goat manure	6	6	22.5	147.66	< .0001*
Cow manure	6	6	22.8	148.53	< .0001*
Chicken manure	6	6	22.4	154.51	< .0001*
Sheep manure	6	6	22.5	148.42	< .0001*
Variety of seeds	4	4	22.4	88.18	0.06309
pH of soil	4	4	22.4	78.45	0.07855

* Significantly difference with *p* – value < 5 %

From Table 4.31, using the joint factor test in the context of Cox reference mixture model it was observed that Goat, Cow, Chicken, and Sheep manure have great influence on

soybean growth and pod development as evidenced with p – value $< 5\%$ significance level. However, it also observed that the effect of variety of seed and pH of soil on the growth of *Glycine max* are almost significant at the same level with $p = 0.06309$ and 0.07855 , respectively. Further, this indicates a possible effects of the process variables which cannot be ruled out resulting from soil pH and the variety of seeds in terms of seeding rate. Therefore, with this findings, we recommends small scale farmers to take into account in order to achieve optimized yield in the framework of SPD with MPV settings.

4.4 Estimated Optimal Yield of *Glycine max* in the Framework of Split-plot

Structure Arrangement

This section outline the result of optimal yield of *Glycine max* within a whole plot. In addition, this section also highlight significant MPV settings within SPD required to achieve maximum desirable output of predicted response of soybean yield.

4.4.1 Response optimization in SPD in the presence of MPV

This section presents the response optimization within the split-plot structure, taking model 3.79 and design A4 into account. It can be observed that models 3.79 using design A4 provide the lowest variance of a future response at the point of interest. Several formulations using screening methodologies through the Cox mixture model framework often result in the prediction of a new response equal to the maximum yield expected to be obtained (Guo *et al.*, 2020; Sekaran *et al.*, 2020). Consequently, a desirable prime objective is to minimize the variance of a new response among the combinations of formulation and process variables that result in predicted response of the eight measurements of variety seed of *Glycine max*. For instance Figure 4.18 illustrates how the yield of *Glycine max* can be optimized number of seeds per plants can be obtained through SPD in the context of

Cox MPV settings. However, the Profiling as shown in Figure 4.18 is an approach to visualizing response surface by seeing what would happen if one or two factors at a time is changed following the MPV settings within split plot arrangement layout. Essentially, a prediction profiler is cross-section view as described by Goos *et al.* (2016). The interactive profilers again shown in the same Figure promote exploring opportunity spaces. This prediction profiler enables one to interpret the fitted Equation to MPV data desirable to optimize the response. This prediction profiler was used to recomputed the profiles and predict responses when the mixture variable component value were varied. The vertical dotted line for MPV shows its current value. As shown in Figure 4.18, for each MPV, the value above the factor name (Goat, cow, chicken and sheep manure) is its current value. The horizontal dotted line shows the current predicted value of Y response (yield of seeds in grams per plant) for the current values of MPV settings. The black lines shown shows how the predicted value changes when the current value of an individual mixture variables setting is changed. In fitting the Equation data, 95% confidence interval for the predicted response value was found to be 14.65455g with MPV settings values shown in red color. However, this prediction profile was useful because of multiple response models in order to help judge which factor values can optimize a complex set of criteria.

Figure 4.18 shows the relationship between the actual and predicted values for the yield of *Glycine max* crop.

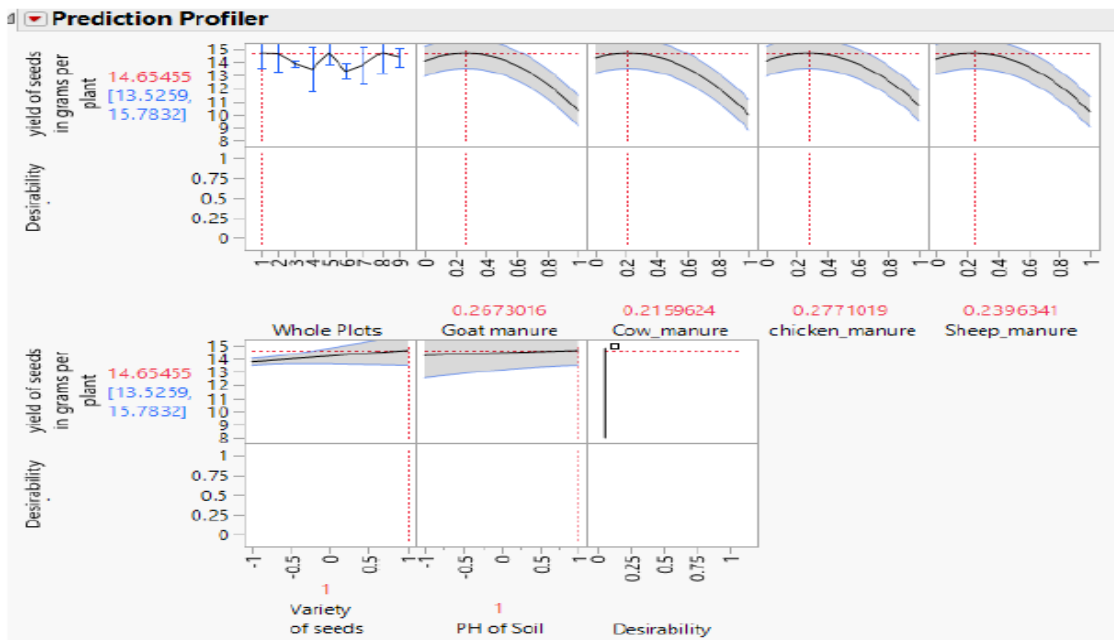


Figure 4.18: The response prediction profiler plus interaction effect

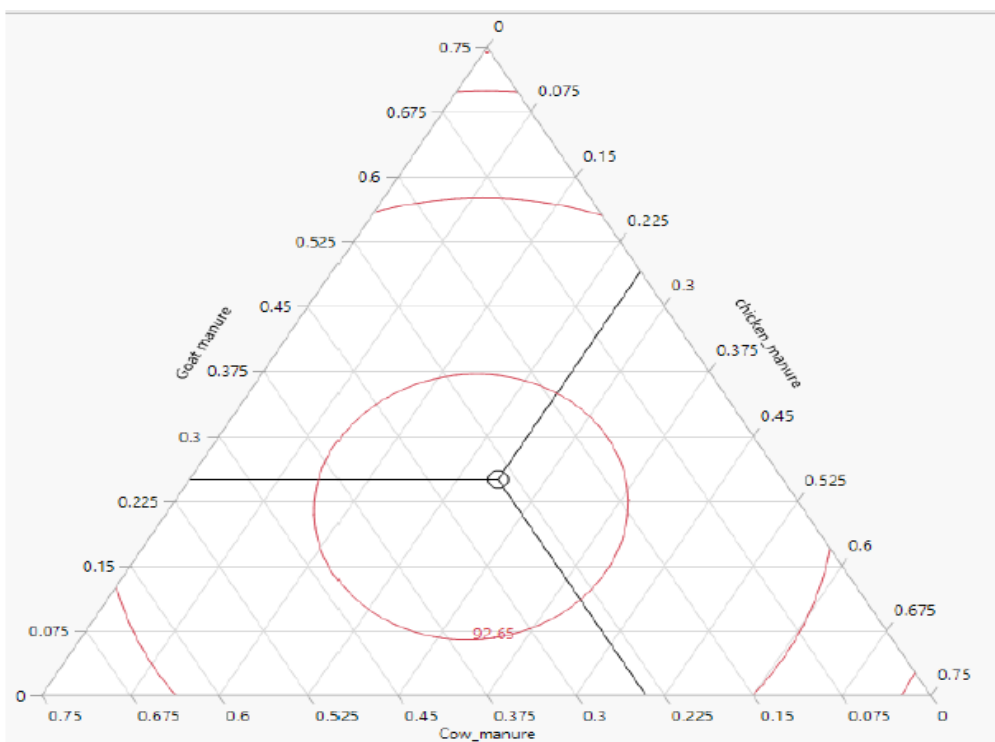


Figure 4.19: The response contours of mixture experiment models on a ternary plot in the presences of process variable.

The response contours of MPV experiment models in Figure 4.19 was produced subject to restriction of Goat and Sheep manure = 0.25. It can be observed that response contours can be optimized to the response surface of the experiment. The predicted response for the average number of seeds per plant was found to be 92.65. If the MPV settings is applied as shown in Table 4.32 the maximum number of seeds can be harvested from each plant is averagely 102.05524 with the assumption that fertility of the land is uniform across.

Table 4.32: The MPV setting that can lead to optimum yield of Glycine max

MPV settings	Whole plot	Goat manure	Cow manure	Chicken manure	Sheep manure	Z_1	Z_2	No. of seeds per plant
Setting 1	1	0.25	0.2407	0.2593	0.25	0	0	94.21
Setting 2	1	0.25	0.2407	0.2593	0.25	1	1	102.06

The optimum yield as shown in Table 45 can be achieved through the mixture setting 2 in whole plot 1 run at both high seeding rate (variety of seeds) and pH of soil. The optimized yield is illustrated on Figure 4.20

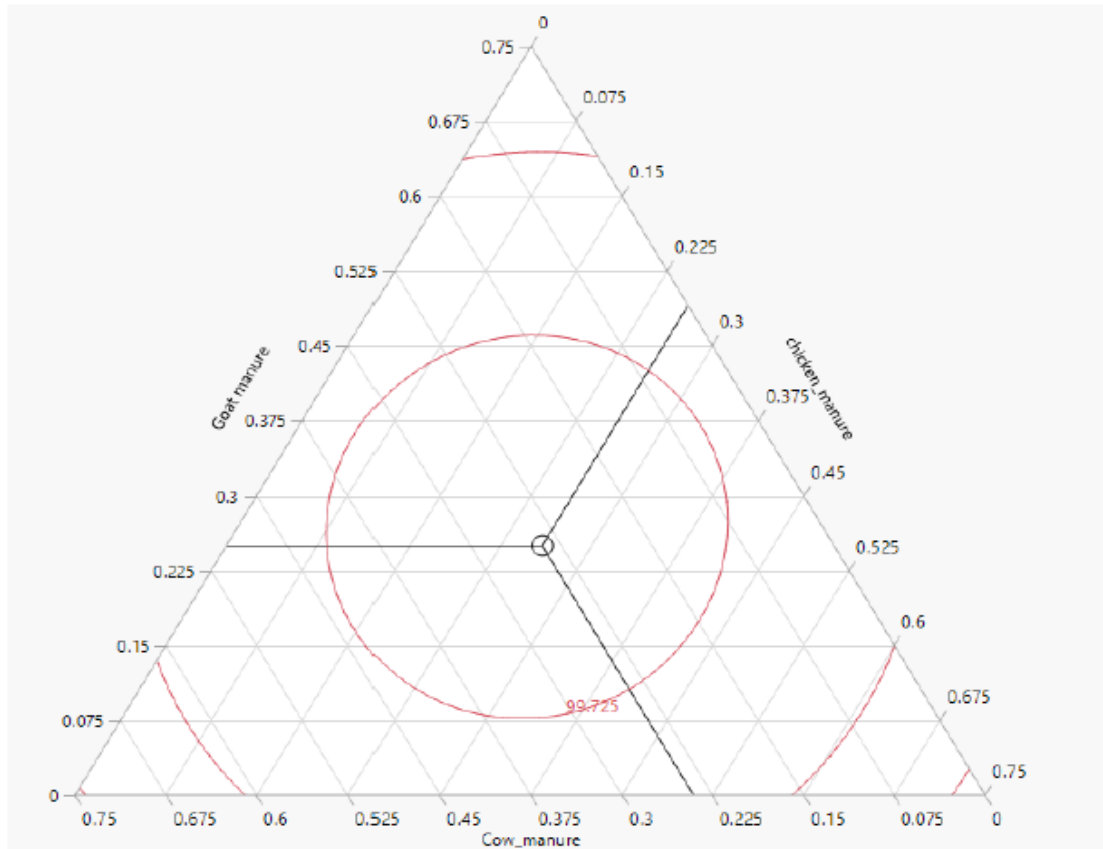


Figure 4.20: The MPV setting with optimum predicted response of *Glycine max*.

The contour plots of response prediction with factors setting of mixture ingredient are presented in Figures 4.19 and 4.20. It can be observed that to attain maximum prediction of response; the process variable settings must be $(z_1 = z_2 = 0, z_1 = z_2 = 1)$ by considering design A4 discussed in Section 3.3. Therefore, analyzing the contour plots, it can be observed that when the goat and sheep manure are fixed at the proportion of 0.25, then the mean response is near 99.65 and 99.75 for the number of seeds plant stem as shown in Figure 4.19 and 4.20, respectively. The result also shows that the optimum yield for the number of seeds is 102.05524, which can be achieved through the following formulation settings shown in Table 45.

4.4.2 Optimal Yield of *Glycine max* within SPD

Table 4.33: Estimation of *Glycine max* yield of variety two (Blyvoor and R 184) in Bushel per acre max

Whole plot	Y_0	Y_{4B}	Y_{7B}	Y_{4R}	Y_{7B}	Bushels per acre for Blyvoor	Bushels per acre for R 184
1	125000	32.2	2.4	31.2	2.3	64.4	59.8
2	225000	32.6	2.4	31.6	2.3	117.36	109.02
3	225000	37.9	2.5	36.9	2.4	142.13	132.84
4	125000	29	2.3	28	2.3	55.58	53.67
5	175000	33.2	2.4	32.2	2.3	92.96	86.4
6	175000	35.1	2.4	33.9	2.3	98.28	90.97
7	175000	30.5	2.3	29.5	2.3	81.84	79.16
8	100000	40.4	2.5	39.4	2.5	67.33	65.67
9	275000	46	2.6	45	2.7	219.27	180.53

*1 bushel = 25.4 Kg

The Table 4.33 shows the predicted *Glycine max* yield per acre in Bushels for each variety basing on the yield obtained from each whole plot in terms Y_0 , Y_4 and Y_7 that represents plant per acre, Pods per plant and seed per pod, respectively based on the MPV data shown Table 20. On the other hand, the subscript B and R denotes the variety Blyvoor and R 184. We estimated optimal yield per acre in Bushel within SPD using the formula described in section 3.5 where one bushel of *Glycine max* (L.) Merrill weighs 60 pounds. The result shown in Table 46 indicates the *Glycine max* growth and pod development increase with the application of MPV settings used. Averagely, the variety Blyvoor does well as compared to variety R 184. The result also shows that the maximum *Glycine max* yield of the two variety is directly proportional to the number of plant per acre. In addition, the

result also indicates that the variety of the seed used has also the impact on the optimum yield obtained.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

We developed a parsimonious model for analyzing mixture process variable tests with control and hard changeable factor within a split-plot structure by expanding initial model produced by Njoroge *et al.* (2017) which considered only three mixture components in the presences of two process variable. The model was developed by introducing SCD of mixture components in the presence of process variables. The SPD, therefore, constituted a simplex SCD of four mixture blends and a 2^2 factorial design with a CCD of the process variable. JMP software version 15 was used to construct D-optimal SPD. The optimality criteria A, D, I, and G was employed in study to compare the constructed designs' relative efficiency. Also, the graphical technique (fraction of design space plot) was used to display, elucidate, and evaluate experimental designs' performance in terms of precision of variance prediction properties of the six designs. The arrangement, where the subplots composed of more SCD points than pure mixture design points or binary mixture design points within a whole plot with presences of two processes both being high, was found to be more efficient and give more precise parameter estimates. However, design A4 with D-efficiency 1.391721 was found desirable and optimal in fitting second order MPV within SPD since its average variance prediction was 0.089642 which was the smallest amongst all the other designs (A1, A2, A3, A4 and A5)

The modified MPV model was employed in modeling and predicting the yield of *Glycine max* with minimal/ reasonable split-plot and main plot errors. The restricted maximum

likelihood method was used to estimate values for P parameter models within the SPD. The experimental results was compared with simulated results using the same MPV setting within SPD found out that there was slight deviation between the simulated and analytical results variety one (R 184) and two (Blyvoor). It was observed that the effect of mixture component at vertices of component X_1, X_2, X_3 and X_4 have the highest impact on the growth and pod development of Soybean together with permutation interaction of these mixture components at 5 % significance level with Chicken manure leading the effect with $\text{LogWorth} = 12.925$ and p –value = 0.0000. On the other hand it was also observed that the effect of variety of seed and pH of soil on the growth of *Glycine max* crop are almost significant at the same level with $p = 0.06309$ and 0.07855 , respectively. Further, this indicates a possible effects of the process variables which cannot be ruled out resulting from soil pH and the variety of seeds in terms of seeding rate. The two source errors that result from main treatments and a subplot treatment for each of the eight responses measured and fitted using the same model formulated were analyzed using REML methods. It found that the average whole plot error variance and average subplot (split-plot) error variance to be 5.325 and 2.099, respectively with corresponding variance ratio of 2.537 since whole plot error variance is often larger than the split-plot error variance. Therefore, this implies that the model adequately represented the mixture data collected from the field, and also, restricted randomization was completely solved with the SPD layout.

The screening methodology was done using the framework of the Cox mixture model approach. The Cox mixture model technique was applied because it provides the experimenter with detailed information about what will happen to the response when an

incremental change is made in any direction from some mixture, as demonstrated in the response prediction profiler since this type of insight cannot be had when the Scheffe model approach is applied since there is no direct way to cooperate with the current mixture. Also, the model's parameter does not describe the change in the predicted response.

The variety Blyvoor was found to perform better than variety R 184 in terms of the yield of seeds harvested and the same condition mixture setting and pH of soil as evidenced. The optimum total yield of *Glycine max* for variety R184 and Blyvoor in Bushel per acre was 180.53 and 219. 217, respectively on the 9th Whole Plot with a pH of soil being 5.4. The predicted maximum optimum yield for the total number of seeds per plant stem of *Glycine max* was 102 and 15.7832 in grams using screening methodology formulated through the framework of the Cox mixture process variable settings.

5.2 Recommendation

We recommend using SPDs in experiments involving mixture settings formulations to measure the interaction effects of both the mixture components and the processing conditions like a pH of the soil and seeding rate. Further, we should also set up the mixture experiment at each of the factorial design points. We also recommend farmers embrace the use of these organic manures when planting different types of crops on the farm as they have good nutritional value for plants' growth. We also advocate for farmers to use soybean variety Blyvoor as it performs better than variety R 184. Again, we recommend screening mixture settings to see which factor components perform better than others in the long run leading to high optimum yield.

In this research, we considered hard-to-change variables (soil pH) as complete plot factors. The researcher can extend the split-plot structure arrangement to a situation where the mixture's components are considered noise variables (hard-change factor). They should develop the Cox mixture model structure's screening methodology to the same problem where the mix materials are considered noise variables.

REFERENCE

- Agromisa, Inckel, M., & Tersmette, T. (2005). *The preparation and use of compost* (Vol. 27). Agromisa.
- Aitchison, J., & Bacon-Shone, J. (1984). Log contrast models for experiments with mixtures. *Biometrika*, 71(2), 323-330.
- Althof, A. J. (2005). Human impact on flora and vegetation of Kakamega Forest, Kenya.
- Anderson Cook, C. M., Goldfarb, H. B., Borror, C. M., Montgomery, D. C., Canter, K. G., & Twist, J. N. (2004). Mixture and mixture-process variable experiments for pharmaceutical applications. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, 3(4), 247-260.
- Atkinson, A. C. (2008). DT-optimum designs for model discrimination and parameter estimation. *Journal of Statistical planning and Inference*, 138(1), 56-64.
- Ayalew, A. (2011). The influence of applying lime and NPK fertilizers on yield of maize and soil properties on acid soil of Areka, southern region of Ethiopia. *Innovative Systems Design and Engineering*, 2(7), 33-42.
- Bekunda, M. A., Bationo, A., & Ssali, H. (1997). Soil fertility management in Africa: A review of selected research trials. *Replenishing soil fertility in Africa*, 51, 63-79.
- Bingham, D. R., & Sitter, R. R. (2001). Design issues in fractional factorial split-plot experiments. *Journal of Quality Technology*, 33(1), 2-15.
- Box, G. E., & Hunter, J. S. (1957). Multi-factor experimental designs for exploring response surfaces. *The Annals of Mathematical Statistics*, 28(1), 195-241.
- Box, G., & Jones*, S. (1992). Split-plot designs for robust product experimentation. *Journal of Applied Statistics*, 19(1), 3-26.
- Burton, J. W. (1997). Soyabean (*Glycine max* (L.) Merr.). *Field Crops Research*, 53(1-3), 171-186.
- Calero, E., West, S. H., & Hinson, K. (1981). Water absorption of soybean seeds and associated causal factors 1. *Crop Science*, 21(6), 926-933.
- Carlisle, B. H. (2005). Modelling the spatial distribution of DEM error. *Transactions in GIS*, 9(4), 521-540.

- Chad Lee and Jim Herbek. (2005). *Kentucky Estimating Soybean Yield, Grain Crops Extension, Plant and Soil Sciences*. University of Kentucky, Kentucky Agricultural Experiment Station.
- Chianu, J. N., Vanlauwe, B., Mahasi, J. M., Katungi, E., Akech, C., Mairura, F. S., & Sanginga, N. (2008). Soybean situation and outlook analysis: the case of Kenya. *TSBF-CIAT. AFNET*, 68.
- Cho, T. Y. (2010). *Mixture-process variable design experiments with control and noise variables within a split-plot structure*. Arizona State University.
- Cho, T. Y., Borrer, C. M., & Montgomery, D. C. (2009). Graphical evaluation of mixture-process variable designs within a split-plot structure. *International Journal of Quality Engineering and Technology*, 1(1), 2-26.
- Chung, P. J., Goldfarb, H. B., Montgomery, D. C., & Borrer, C. M. (2009). Optimal designs for mixture-process experiments involving continuous and categorical noise variables. *Quality Technology & Quantitative Management*, 6(4), 451-470.
- Claringbold, P. J. (1955). Use of the simplex design in the study of joint action of related hormones. *Biometrics*, 11(2), 174-185.
- Cornell, J. A. (1975). Some comments on designs for Cox's mixture polynomial. *Technometrics*, 17(1), 25-35.
- Cornell, J. A. (1988). Analyzing data from mixture experiments containing process variables: a split-plot approach. *Journal of Quality Technology*, 20(1), 2-23.
- Cornell, J. A. (1990). How to apply response surface methodology, Vol. 8. *ASQC, Wisconsin*, 3.
- Cornell, J. A. (2011). *Experiments with mixtures: designs, models, and the analysis of mixture data* (Vol. 403). John Wiley & Sons.
- Crary, S. B., Hoo, L., & Tennenhouse, M. (1992). I-optimality algorithm and implementation. In *Computational Statistics* (pp. 209-214). Physica-Verlag HD.
- Czitrom, V. (1988). Mixture experiments with process variables: D-optimal orthogonal experimental designs. *Communications in Statistics-Theory and Methods*, 17(1), 105-121.
- Czitrom, V. (1992). *Experiments with Mixtures*.

- Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data* (Vol. 62). CRC press.
- De Bruin, J. L., & Pedersen, P. (2008). Effect of row spacing and seeding rate on soybean yield. *Agronomy journal*, 100(3), 704-710.
- De Ridder, N., Breman, H., van Keulen, H., & Stomph, T. J. (2004). Revisiting a 'cure against land hunger': soil fertility management and farming systems dynamics in the West African Sahel. *Agricultural systems*, 80(2), 109-131.
- Demirel, M., & Kayan, B. (2012). Application of response surface methodology and central composite design for the optimization of textile dye degradation by wet air oxidation. *International Journal of Industrial Chemistry*, 3(1), 24.
- Dev, S. P., & Tilak, K. V. B. R. (1976). Effect of organic amendments on the nodulation and nitrogen fixation by soybean [India]. *Indian Journal of Agricultural Sciences (India)*.
- Fabiyyi, E. F. (2006). Soyabean processing, utilization and health benefits. *Pakistan Journal of Nutrition*, 5(5), 453-457.
- Fageria, N. K., Baligar, V. C., & Clark, R. B. (2002). Micronutrients in crop production. In *Advances in agronomy* (Vol. 77, pp. 185-268). Academic Press.
- FAOSTAT. <http://faostat.fao.org> (accessed 6 2011).
- Fedorov, V. V. (1972). Theory of optimal experiments., Translated from the Russian and edited by WJ Studden and EM Klimko. *Probability and mathematical statistics*, (12).
- Gentili, R., Ambrosini, R., Montagnani, C., Caronni, S., & Citterio, S. (2018). Effect of soil pH on the growth, reproductive investment and pollen allergenicity of *Ambrosia artemisiifolia* L. *Frontiers in plant science*, 9, 1335.
- Giovannitti-Giovannitti-Jensen, A., & Myers, R. H. (1989). Graphical assessment of the prediction capability of response surface designs. *Technometrics*, 31(2), 159-171.
- Goldfarb, H. B., & Montgomery, D. C. (2006). Graphical methods for comparing response surface designs for experiments with mixture components. In *Response Surface Methodology and Related Topics* (pp. 329-348).

- Goldfarb, H. B., Anderson Cook, C. M., Borror, C. M., & Montgomery, D. C. (2004a). Fraction of design space plots for assessing mixture and mixture-process designs. *Journal of Quality Technology*, 36(2), 169-179.
- Goldfarb, H. B., Borror, C. M., & Montgomery, D. C. (2003). Mixture-process variable experiments with noise variables. *Journal of Quality Technology*, 35(4), 393-405
- Goldfarb, H. B., Borror, C. M., Montgomery, D. C., & Anderson Cook, C. M. (2004b). Three-dimensional variance dispersion graphs for mixture-process experiments. *Journal of Quality Technology*, 36(1), 109-124.
- Goos, P., Jones, B., & Syafitri, U. (2016). I-optimal design of mixture experiments. *Journal of the American Statistical Association*, 111(514), 899-911.
- Goos, P., & Syafitri, U. (2014). V-optimal mixture designs for the qth degree model. *Chemo metrics and Intelligent Laboratory Systems*, 136, 173-178.
- Goos, P., & Jones, B. (2011). *Optimal design of experiments: a case study approach*. John Wiley & Sons.
- Goos, P., & Donev, A. N. (2007). Tailor-made split-plot designs for mixture and process variables. *Journal of Quality Technology*, 39(4), 326-339
- Goos, P., & Vanderbroek, M. (2003). D-optimal split-plot designs with given numbers and sizes of whole plots. *Technometrics*, 45(3), 235-245.
- Goos, P., & Vandebroek, M. (2001). Optimal split-plot designs. *Journal of Quality Technology*, 33(4), 436-450.
- Gorman, J. W., & Hinman, J. E. (1962). Simplex lattice designs for multicomponent systems. *Technometrics*, 4(4), 463-487.
- Gunaraj, V., & Murugan, N. (1999). Application of response surface methodology for predicting weld bead quality in submerged arc welding of pipes. *Journal of materials processing technology*, 88(1-3), 266-275.
- Guo, L., Xiong, Y., & Joan Hu, X. (2020). Estimation in the Cox cure model with covariates missing not at random, with application to disease screening/prediction. *Canadian Journal of Statistics*, 48(4), 608-632.

- Han, S. H., An, J. Y., Hwang, J., Kim, S. B., & Park, B. B. (2016). The effects of organic manure and chemical fertilizer on the growth and nutrient concentrations of yellow poplar (*Liriodendron tulipifera* Lin.) in a nursery system. *Forest science and technology*, *12*(3), 137-143.
- Hassan, W. N. F. W., Ismail, M. A., Lee, H. S., Meddah, M. S., Singh, J. K., Hussin, M. W., & Ismail, M. (2020). Mixture optimization of high-strength blended concrete using central composite design. *Construction and Building Materials*, *243*, 118251.
- Hossner, L. R., & Juo, A. S. (1999). *Soil nutrient management for sustained food crop production in upland farming systems in the tropics* (p. 18). Food and Fertilizer Technology Center
- Hymowitz, T. (1970). On the domestication of the soybean. *Economic botany*, *24*(4), 408-421.
- Isaev S.Kh., Kadirov Z.Z., Khamraev K.Sh., Atamuradov B.N., Sanoev Kh.A.. SCIENTIFIC BASIS FOR SOYBEAN PLANTING IN THE CONDITION OF GRASSY ALLUVIAL SOIL PRONE TO SALINIZATION. JCR. 2020; 7(4): 354-360. [doi:10.31838/jcr.07.04.68](https://doi.org/10.31838/jcr.07.04.68)
- Iwundu MP. Missing observations: The loss in relative A-, D-and G-efficiency; 2017.
- Jackson, A. S. (2016). A brief history of soybean production in Kenya. *Research Journal of Agriculture and Environmental Management*, *5*(2), 58-64.
- Jaetzold, R., Schmidt, H., Hornetz, B., & Shisanya, C. (2006). Farm management handbook of Kenya Vol. II: Natural conditions and farm management information Part C East Kenya Subpart C1 Eastern Province. *Cooperation with the German Agency for Technical Cooperation (GTZ)*.
- Jagwe, J., & Nyapendi, R. (2004). Evaluating the marketing opportunities for soybean and its products in the East African countries of ASARECA. *Kenya Report. International Institute of Tropical Agriculture (IITA). Foodnet*.
- Jeavons, J. C. (2001). Biointensive sustainable mini-farming: II. Perspective, principles, techniques and history. *Journal of Sustainable Agriculture*, *19*(2), 65-76.
- Jones, B., & Goos, P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *56*(3), 347-364.

- Jones, B., & Sall, J. (2011). JMP statistical discovery software. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(3), 188-194.
- Kamble, B. M., Rathod, S. D., & Phalke, D. H. (2009). Effect of organic & inorganic fertilizers on yield of soybean, wheat and turmeric and available nutrient status of soil. *Advances in Plant Sciences*, 22(2), 511-515.
- Karuga, S., & Gachanja, J. (2004). Soybeans development study: Final report prepared by market economies development for Japan International Cooperation Agency (JICA). *Nairobi, Kenya*.
- Kendall, M. G., & Stuart, A. (1968). *The Advanced Theory of Statistics: By Maurice Kendall and Alan Stuart* (Vol. 1). Hafner Publishing Company.
- Khuri, A. I., Harrison, J. M., & Cornell, J. A. (1999). Using quantile plots of the prediction variance for comparing designs for a constrained mixture region: an application involving a fertilizer experiment. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(4), 521-532.
- Kiefer, J. (1961). Optimum designs in regression problems, II. *The Annals of Mathematical Statistics*, 298-325.
- Kimetu, J. M., Lehmann, J., Ngoze, S. O., Mugendi, D. N., Kinyangi, J. M., Riha, S., & Pell, A. N. (2008). Reversibility of soil productivity decline with organic matter of differing quality along a degradation gradient. *Ecosystems*, 11(5), 726.
- Kinyanjui, M. J., & Wanjala, W. R. (2016). Impact of human activities on land degradation IN Lugari sub county, Kakamega County, Kenya. *Journal of Environmental Research and Management*, Vol. 7(2). pp. 0038-0044.
- Kowalski, S. M., Cornell, J. A., & Vining, G. G. (2002). Split-plot designs and estimation methods for mixture experiments with process variables. *Technometrics*, 44(1), 72-79.
- Kowalski, S., Cornell, J. A., & Geoffrey Vining, G. (2000). A new model and class of designs for mixture experiments with process variables. *Communications in Statistics-Theory and Methods*, 29(9-10), 2255-2280
- Kussmaul, K. (1969). Protection against assuming the wrong degree in polynomial regression. *Technometrics*, 11(4), 677-682.

- Laake, P. (1975). On the optimal allocation of observations in experiments with mixtures. *Scandinavian Journal of Statistics*, 153-157.
- Laird, N., Lange, N., & Stram, D. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association*, 82(397), 97-105.
- Lambrakis, D. P. (1968). Experiments with mixtures: A generalization of the simplex-lattice design. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(1), 123-136.
- Lawson, J., & Willden, C. (2016). Mixture experiments in R using mixexp. *Journal of Statistical Software*, 72 (Code Snippet, 2), 1-20.
- Lee, C. D., Egli, D. B., & TeKrony, D. M. (2008). Soybean response to plant population at early and late planting dates in the Mid-South. *Agronomy Journal*, 100(4), 971-976.
- Liang, L., Anderson Cook, C. M., & Robinson, T. J. (2006). Fraction of design space plots for split-plot designs. *Quality and Reliability Engineering International*, 22(3), 275-289.
- Liu, S., & Neudecker, H. (1995). A V-optimal design for Scheffé's polynomial model. *Statistics & probability letters*, 23(3), 253-258.
- Lohr, S. L. (1995). Hasse diagrams in statistical consulting and teaching. *The American Statistician*, 49(4), 376-381.
- Mahasi, J. M., Mukalama, J., Mursoy, R. C., Mbehero, P., & Vanlauwe, B. (2011). A sustainable approach to increased soybean production in western Kenya. In *10th African Crop Science Conference Proceedings, Maputo, Mozambique, 10-13 October 2011*. African Crop Science Society.
- Mbau, S. K., Karanja, N., & Ayuke, F. (2015). Short-term influence of compost application on maize yield, soil macrofauna diversity and abundance in nutrient deficient soils of Kakamega County, Kenya. *Plant and soil*, 387(1-2), 379-394.
- Mbembe, E. A. (2020). *Determinants of Market Participation by Smallholder Soybean Farmers in Kakamega County, Kenya* (Doctoral dissertation, University of Nairobi).
- McLean, R. A., & Anderson, V. L. (1966). Extreme vertices design of mixture experiments. *Technometrics*, 8(3), 447-454.

- McLean, R. A., & Anderson, V. L. (1966). Extreme vertices design of mixture experiments. *Technometrics*, 8(3), 447-454.
- Meyer, R. K., & Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1), 60-69.
- Mikaeili, F. (1993). D-optimum design for full cubic on q-simplex. *Journal of statistical planning and inference*, 35(1), 121-130.
- MOA, (2017). Ministry of Agriculture, Annual report.
- Murage, E. W., Karanja, N. K., Smithson, P. C., & Woomer, P. L. (2000). Diagnostic indicators of soil quality in productive and non-productive smallholders' fields of Kenya's Central Highlands. *Agriculture, Ecosystems & Environment*, 79(1), 1-8.
- Mwangi, W. P., Anapapa, A., & Otieno, A. (2019). Selection of Second Order Models' Design Using D-, A-, E-, T-Optimality Criteria. *Asian Journal of Probability and Statistics*, 1-15.
- Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2009). Other mixture design and analysis techniques. *Response surface methodology, 3rd ed., Hoboken, NJ, USA: Wiley*, 589-643.
- Neina, D. (2019). The role of soil pH in plant nutrition and soil remediation. *Applied and Environmental Soil Science*, 2019.
- Njoroge, G. G., Simbauni, J. A., & Koske, J. A. (2017). An Optimal Split-Plot Design for Performing a Mixture-Process Experiment. *Science Journal of Applied Mathematics and Statistics*, 5(1), 15.
- Oehlert, G. W. (2010). *A first course in design and analysis of experiments*.
- Ozol-Godfrey, A., Anderson-Cook, C. M., & Montgomery, D. C. (2005). Fraction of design space plots for examining model robustness. *Journal of quality technology*, 37(3), 223-235.
- Park, S. H. (1978). Selecting contrasts among parameters in Scheffe's mixture models: Screening components and model reduction. *Technometrics*, 20(3), 273-279.
- Parker, P. A., Kowalski, S. M., & Vining, G. G. (2007). Construction of balanced equivalent estimation second-order split-plot designs. *Technometrics*, 49(1), 56-65.
- Pedersen, P. (2004). Soil pH and Plant Population Effects on Soybean Yield. *Iowa State University Research and Demonstration Farms Progress Reports*, 2003(1).

- Pender, J., Place, F., & Ehui, S. (Eds.). (2006). *Strategies for sustainable land management in the East African highlands*. Intl Food Policy Res Inst.
- Piepel, G. F. (1990). Screening designs for constrained mixture experiments derived from classical screening designs. *Journal of Quality Technology*, 22(1), 23-33.
- Piepel, G., Anderson, C., & Redgate, P. E. (1993). Variance dispersion graphs for designs on polyhedral regions-revisited. In *Proceedings of the Section on Physical and Engineering Sciences* (pp. 102-107). Alexandria, Virginia: American Statistical Association.
- Prescott, P. (2004). Modelling in mixture experiments including interactions with process variables. *Quality Technology & Quantitative Management*, 1(1), 87-103.
- Probst, A. H., & Judd, R. W. (1973). Origin, US history and development, and world distribution. *Agronomy*.
- Rady, E. A., Abd El-Monsef, M. M. E., & Seyam, M. M. (2009). Relationships among several optimality criteria. *Interstat*, 15(6), 1-11.
- Rodriguez, M., Jones, B., Borrer, C. M., & Montgomery, D. C. (2010). Generating and assessing exact G-optimal designs. *Journal of quality technology*, 42(1), 3-20.
- Rozum, M. A., & Myers, R. H. (1991). Adaptation of Variance Dispersion Graphs to Cuboidal Regions of Interest. In *Joint Statistical Meetings, American Statistical Association, Atlanta, GA*.
- Salehi, M. B., Sefti, M. V., Moghadam, A. M., & Koohi, A. D. (2012). Study of salinity and pH effects on gelation time of a polymer gel using central composite design method. *Journal of Macromolecular Science, Part B*, 51(3), 438-451.
- Scheffe, H. (1958). Experiments with mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 344-360.
- Scheffe, H. (1963). The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25(2), 235-251.
- Sekaran, K., Chandana, P., Krishna, N. M., & Kadry, S. (2020). Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer. *Multimedia Tools and Applications*, 79(15), 10233-10247.

- Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika*. 52 (3-4): 591-611. [doi:10.1093/biomet/52.3-4.591](https://doi.org/10.1093/biomet/52.3-4.591). [JSTOR 2333709](https://www.jstor.org/stable/2333709). [MR 0205384](https://www.jstor.org/stable/0205384). p. 593
- Singh, S. K., Hoyos-Villegas, V., Houx III, J. H., & Fritschi, F. B. (2012). Influence of artificially restricted rooting depth on soybean yield and seed quality. *Agricultural water management*, 105, 38-47.
- Sinha, B. K., Mandal, N. K., Pal, M., and Das, P. (2014), *Optimal Mixture Experiments*, New Delhi: Springer. [900,901,902]
- Sitinjak, M.A., & Syafitri, U. D. (2019, December). A Split Plot Design for an Optimal Mixture Process Variable Design of a Baking Experiment. In *Journal of Physics: Conference Series* (Vol. 1417, No. 1, p. 012018). IOP Publishing.
- Smith, W. F. (2005). *Experimental Design for Formulation*, vol. 15 of ASA-SIAM Series on Statistics and Applied Probability. *Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA*.
- Snee, R. D. (1985). Computer-aided design of experiments—some practical experiences. *Journal of Quality Technology*, 17(4), 222-236.
- Snee, R. D., & Marquardt, D. W. (1976). Screening concepts and designs for experiments with mixtures. *Technometrics*, 18(1), 19-29.
- Song, B., Marchant, M. A., Reed, M. R., & Xu, S. (2009). Competitive analysis and market power of China's soybean import market. *International Food and Agribusiness Management Review*, 12(1030-2016-82749), 21-42.
- Steiner, S. H., & Hamada, M. (1997). Making mixtures robust to noise and mixing measurement errors. *Journal of Quality Technology*, 29(4), 441-450.
- Swain, S. C., Rath, S., & Ray, D. P. (2006). Effect of NPK levels and mulching on yield of turmeric in rainfed uplands. *JOURNAL OF RESEARCH-BIRSA AGRICULTURAL UNIVERSITY*, 18(2), 247.
- Tittonell, P., Vanlauwe, B., Corbeels, M., & Giller, K. E. (2008). Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of western Kenya. *Plant and Soil*, 313(1-2), 19-37.

- Tittonell, P., Zingore, S., Van Wijk, M. T., Corbeels, M., & Giller, K. E. (2007). Nutrient use efficiencies and crop responses to N, P and manure applications in Zimbabwean soils: Exploring management strategies across soil fertility gradients. *Field crops research*, 100(2-3), 348-368.
- Tsikhungu, P. W. (2016). *Analysis of micronutrients and heavy metals of indigenous reed salts and soils from selected areas in Western Kenya* (Doctoral dissertation, Egerton University).
- Uranisi, H. (1964), "Optimal Design for the Special Cubic Regression Model on the q -Simplex," Mathematical Report 1, Kyushu University, General Education Department. [902,909]
- Vining, G. G., Cornell, J. A., & Myers, R. H. (1993). A graphical approach for evaluating mixture designs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 42(1), 127-138.
- Vining, G. G., Kowalski, S. M., & Montgomery, D. C. (2005). Response surface designs within a split-plot structure. *Journal of Quality Technology*, 37(2), 115-129.
- Wang, X., Yang, G., Li, F., Feng, Y., Ren, G., & Han, X. (2013). Evaluation of two statistical methods for optimizing the feeding composition in anaerobic co-digestion: Mixture design and central composite design. *Bioresource technology*, 131, 172-178.
- Wanyama, W. E. (2013). *Response of soybean (Glycine max l.) to application of inorganic fertilizers, cattle manure and lime in western Kenya* (Doctoral dissertation). University of Nairobi
- Wanyonyi, S. W., Mbete, D. A., & Chimusa, E. R. (2018). Computational Generalization of Mixed Models on Large-Scale Data with Applications to Genetic Studies. *Asian Journal of Probability and Statistics*, 1-31.
- Weese, M. (2010). A new screening methodology for mixture experiments (Doctoral Dissertation) University of Tennessee, Knoxville.
- Wong, W. K. (1994). Comparing robust properties of A, D, E and G-optimal designs. *Computational statistics & data analysis*, 18(4), 441-448.

- Zahran, A., Anderson Cook, C. M., & Myers, R. H. (2003). Fraction of design space to assess prediction capability of response surface designs. *Journal of Quality Technology*, 35(4), 377-386.
- Zhang, Z., Dong, X., Wang, S., & Pu, X. (2020). Benefits of organic manure combined with biochar amendments to cotton root growth and yield under continuous cropping systems in Xinjiang, China. *Scientific Reports*, 10(1), 1-10

APPENDICES

APPENDIX I: R Code

```
> library(lattice) ## loading package of lattice

> library(daewr)

> library(mixexp)

> des1<-SLD(3,2) ##design points for simplex lattice design

> des2<-SCD(3)##design points for simplex centroid design

> DesignPoints(des1)## Figure 2

> DesignPoints(des2)## Figure 6
```

Alternatively, load AlgDesign package and apply the r codes shown below.

```
> library(AlgDesign)

> DesignPoints(des = NULL, nmxcmp=3, x = NULL, y = NULL, z = NULL,x1lower=0,x1upper=0, x2lower=0.05, x2upper=0, x3lower=0, x3upper=0, cornerlabs = c("x3","x2","x1"),
+       axislabs=c("x1","x2","x3"),pseudo=FALSE)
```

R Codes for Pseudo Components

```
Actual_data <- Xvert(nfac = 3, lc = c(0.35, 0.2, 0.15), uc = c(1, 1, 1), ndm = 1, plot = FALSE)##
```

nfac= No. of component factors

```
Response1 <- c(15.3, 20.0, 28.6, 12.5, 32.7, 42.4)
```

```
Actual_data <- cbind(Actual_data[1:6, ], Response1)
```

```
Fitted_quadm <- lm(y ~ -1 + x1 + x2 + x3 + x1:x2 + x1:x3 + x2:x3, data = Actual_data)
```

```
title <- c("Actual Component Space", "L Pseudo Component Space")

option <- c(FALSE, TRUE)


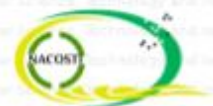



for (i in 1:2) {ModelPlot(model =Fitted_quadm, dimensions = list(x1 = "x1", x2 = "x2", x3 =
"x3"),

      main = title[i], lims = c(0.35, 1, 0.20, 1, 0.15, 1),

      constraints = TRUE, contour = TRUE, cuts = 6, fill = TRUE,

      axislabs = c("x1", "x2", "x3"), cornerlabs = c("x1", "x2", "x3"), pseudo = option[i]}
```

APPENDIX II: Research Licenses

 REPUBLIC OF KENYA	 NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION
RafNo: 770537	Date of issue: 25/January/2021
RESEARCH LICENSE	
	
<p>This is to Certify that Mr.. Samson Wangila Wanyonyi of University of Eldoret, has been licensed to conduct research in Kakamega on the topic: APPLICATION OF A FOUR COMPONENT MIXTURE VARIABLES IN THE PRESENCES OF TWO PROCESS VARIABLES WITHIN A SPLIT-PLOT DESIGN MODELING THE YIELD OF Glycine max (L.) Merrill for the period ending : 25/January/2022.</p>	
License No: NACOSTI/P/21/8540	
770537 Applicant Identification Number	 Director General NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION
	Verification QR Code 
<p>NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.</p>	


APPENDIX III: Similarity Report

Document Viewer

Turnitin Originality Report

Processed on: 09-Jun-2021 11:40 EAT
ID: 1603340281
Word Count: 51231
Submitted: 1

SSCI/MAT/P/003/18 By
Samson Wanyonyi Wangila



Similarity Index	Similarity by Source
14%	Internet Sources: 12% Publications: 6% Student Papers: 2%

include quoted include bibliography excluding matches < 5 words mode:
quickview (classic) report print refresh download

1% match (Internet from 16-Jul-2020) https://pdfs.semanticscholar.org/eb8c/b06b5f7518784c347e6968a46c720892fa3e.pdf	☒
1% match (Internet from 25-Jan-2008) http://www.qualityprogress.org	☒
1% match (Internet from 31-Jul-2014) http://www.scielo.br	☒