Grain Yield Stability, Genetic Gain and Path Coefficient Analyses in Advanced Soybean (*Glycine max (L.) Merr.)* Lines

By

Mwila, CHIBANDA

BSc Agricultural Science (University of Zambia)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science (MSc) in Plant Breeding

School of Agriculture, Earth and Environmental Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

2017

DECLARATION

I, Mwila Chibanda, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   a) their words have been re-written but the general information attributed to them has been referenced;
   b) Where their exact words have been used, their writing has been placed inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References section.

Signed:

……..……  ……..

As the candidate's supervisor, I agree to the submission of this thesis

……  ……

Professor John Derera (Supervisor)

As the candidate's co-supervisor, I agree to the submission of this thesis

……  ……

Dr Julia Sibiya (Co-Supervisor)

ABSTRACT

Zambia is among the top 20 leading global producers of soybean (*Glycine max* L. *Merr.*) but adequate production is still hampered by low productivity. The yields of soybean in Zambia average below 3.0 t/ha against a yield potential of 5.0 t/ha. This is attributed in part to poor availability of well adapted and improved cultivars. Therefore, selection for high yield potential is the prime objective of the breeding programme in the medium altitude and subtropical environments in Africa. Unfortunately, spatial and seasonal variability is large in this ecological zone. Therefore, the objectives of the study were to assess the nature and magnitude of the genotype x environmental (G x E) interactions for grain yield, to identify stable genotypes; to determine the genetic gains achieved in breeding for grain yield over 12 years, and to determine the secondary traits that directly or indirectly affect yield in soybean cultivars. Thirty genotypes that were drawn from the advanced set of lines in the programme were evaluated across 16 locations in Zambia, Malawi and Zimbabwe. The experiments were laid out in a 6 x 5 alpha lattice design, with three replications at each site. The recommended cultural practices were followed at all sites in all countries. The data were subjected to analysis of variance (ANOVA, correlation and path coefficient analysis, cultivar superiority index, Additive main effects and multiplicative interaction (AMMI) and Genotype, Genotype and Environment (GGE) biplot analyses, in GenStat statistical software. There were significant genotypes main effects, environment main effects and their interaction effects. The G x E of cross over type was observed. The genotypes G2, G10 and G15 were ranked among most stable genotypes by all methods, while G2 was the most desirable genotype across locations, followed by G15. Biplot analysis revealed that E6 was the most discriminative test location while the most representative one was E4. The genetic gain study showed a 21% gain in Zambia and Malawi. No significant gain was registered in Zimbabwe. An across site analysis of all test locations resulted in a disappearance of all genetic gain earlier observed. The cross over GXE interaction negatively affected heritability of grain yield and masked the appearance of any gains. Overall, a 6.5% gain over the population mean, showed that selection was successful in increasing yield. However, there was no significant gain observed relative to the current commercial cultivars, indicating limited breeding progress. The results of PATH correlation analysis showed that yield was positively and significantly correlated with all traits except the number of seeds per pod. However, the correlation was weak with the exception of harvest index. The harvest index, biomass and number of pods per plant had significant influence on

yield. Selection for these three traits, Harvest index, biomass and number of pods per plant would be emphasised to improve yield potential in the soybean programme.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors Professor John Derera and Dr Julia Sibiya who provided me with guidance and encouragement throughout the period of my research and thesis writing. Thank you to the Alliance for a Green Revolution in Africa (AGRA) for meeting the costs of my studies and research, without them, this study would have been difficult to implement.

Special appreciation goes to SeedCo Limited Company who provided mentorship through Dr H. Mushoriwa, Mr. J. Tchagwa and Mr. H. Masole and provided the germplasm used in this study. Through their dedicated teams in Zambia, Malawi and Zimbabwe, I was able to monitor the trials and collect the data needed to complete my research.

Lastly but most importantly, I would like to acknowledge my parents for readily stepping in as guardians to my son, making it possible for me to leave Zambia and explore this career opportunity.

DEDICATION

This dissertation is dedicated to my children, Luthanda Chilufya Mkandawire, and Malala Thandiwe Ngoma. May my achievements be a source of inspiration to you.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

Soybean (*Glycine max L.*) is an important source of protein and edible vegetable oil in many parts of the world. In Zambia, it is grown in three agro-ecological regions (Regions I, II and III), primarily to feed into oil and feed processing industries. Studies have shown a change in nutritional lifestyle of the general world population in favour of meat (Wang *et al.*, 2003; Iqbal *et al.*, 2006; Goldsmith, 2008). Zambia is no exception to this trend. Therefore, this justifies studies that aim to enhance productivity of the crop in Zambia.

With a rapid rise in livestock production, demand for soya grain as a source of protein in the manufacture of livestock feed has increased. This presents a greater market opportunity for the crop in Zambia. According to TechnoServe (2011), 89% of all soybean produced in Zambia in 2010 was used as a protein source in livestock feed production, while only 11% went to edible oil and other food productions. In response to the growing demand for feed, soybean production in Zambia rose from 2,350 tonnes in 2001 to over 261,000 tonnes in 2013 (FAOSTAT, 2015). It has continued to grow since then. However, trends show that this increase in production has been largely a result of farmers expanding the hectarage and the increase in production has not been obtained as a result of increase in yields as illustrated in Figure 1.

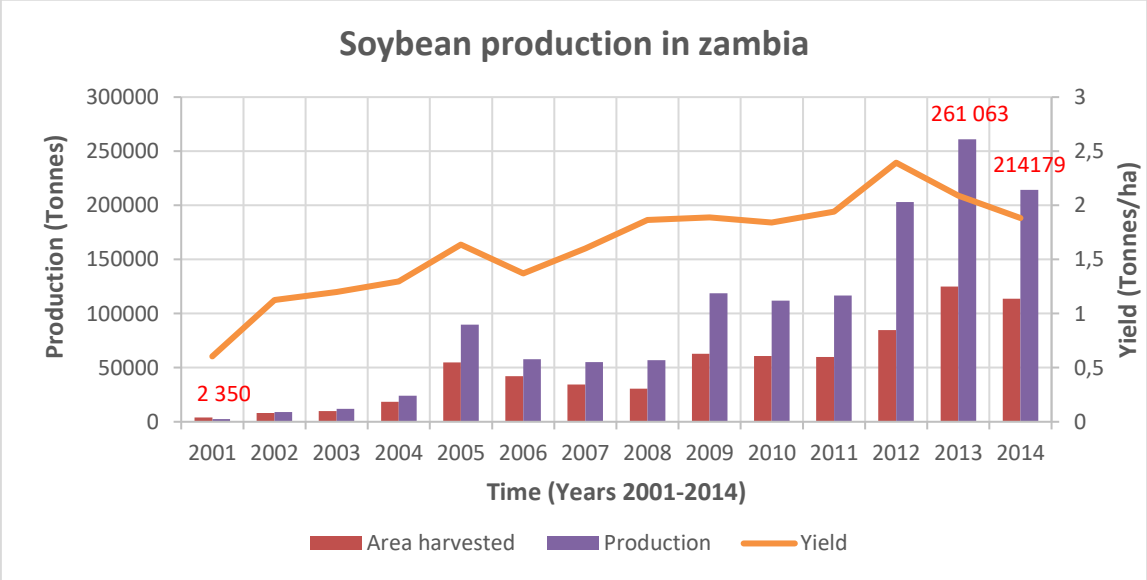

Figure 1. Soybean production and productivity in Zambia between 2001 and 2014

(Source: FAOSTAT 2015)

This has implications on production costs and productivity of the crop. It has generally been viewed as unprofitable to produce soybean, since lower yields translate to high costs of production and low returns on production (Opperman and Varia, 2011).Therefore profitability can be enhanced by increasing the yield to above global averages. The average yield of soybean is 3.2tonnes/ha in the United States and 2.4tonnes/ha in Brazil (Alves *et al.*, 2003) while it remains below 1tonnes/ha in Zambia.

There is scope to increase grain yield of soybean in Zambia to match and exceed the global averages. For example it has been reported that the soybean cultivars available in Zambia have yield potential of up to 5 t/ha, but the average yields being realised are low, ranging from 0.9 t/ha to 3.0 t/ha (TechnoServe, 2011). Regrettably, these low yields have remained fairly constant over time creating and allowing the yield gap to persist. The low grain yields being realized do not always merit the cost of harvesting such that many fields remain abandoned at the end of the season. This is a waste of valuable resources given that the environment in Zambia is very suitable for soybean production. The constraints that hamper adequate production include biotic and abiotic factors.

1.1.1 Biotic constraints

Diseases threaten adequate production of soybean in the medium altitude environments. The major diseases affecting soybean production in sub Saharan Africa are Soybean Rust (Miles *et al.*, 2008) , Frog eye leaf spot (*Cercospora sojina. (K.) Hara*) and Red leaf blotch (*Pyrenochaeta glycine (*R.B) Stewart). The most damaging of these being Soybean Rust. By 2000, yield loses of 60-80% were reported in Zimbabwe and Zambia (Levy, 2005). Scientists such as Jacob Tichagwa, have identified a few rust tolerant cultivars with yield potential of at least 4.0 t/ha (*Personal communication*) reducing yield losses by up to 90%. However, the presence of rust tolerant varieties has allowed diseases that initially had little economic value to increase.

1.1.2 Abiotic constraints

Another important constraint that negatively affect yield is drought, among other abiotic constraints. Soybean is greatly affected by drought in the early and later stages of development. This compromises yield because even under non-stress conditions, soybean would abort a fair number of flowers. With added water stress at flowering, the rate of abortion has been found to increase and directly reduce biomass or crop yield (Kokubun, 2011). With the erratic rainfall pattern being experienced in most parts of southern Africa, water became even more limiting for increased crop yields, during the 2015/16 season when the current study was conducted. The ideal genotypes for the prevailing environment in the region would

be drought tolerant soybean varieties and the capability to irrigate the crop during the dry and drought spells in Zambia. However, poor availability of adapted genotypes and lack of irrigation are ranked as the major limiting factors to soybean production. Therefore, development of varieties with a high adaptability is imperative for soybean production in Zambia. This is because the agro-ecological zones that are suited for production of the crop in Zambia are generally variable in temperature, precipitation and other factors that affect the physiological development of the crop.

1.2 Soybean breeding prospects

Through plant breeding efforts, well-adapted genotypes can be developed in order to increase yields in the medium altitude and subtropical production environments. However, this means that breeders must simultaneously select material with good performance as well as the ability to perform consistently and produce mean performance that is above average in all locations (Gurmu *et al.*, 2009). Due the large seasonal and spatial variability that is experienced in the subtropical production environments in Zambia, farmers would ideally require varieties that are productive under both non-stress and stress environments. Farmers do not want to incur a yield penalty when a favourable environment occurs. This has breeding implications. Previous researchers have suggested that the genotypes selected must have genetic potential for superior performance under ideal growing conditions, and must produce acceptable yields under less favourable environments (Yan and Rajcan, 2002; Gurmu *et al.*, 2009). The desired genotype must ideally be responsive to good growing conditions. Therefore, breeding for stability is imperative. However, a clear understanding of how the genotypes would interact with the environment is crucial for the subtropical breeding programme in Zambia.

Environmental variability, as exhibited in Zambia's three agro-ecologies, produces complex interactions between genotypes and environments so that the yield of a given genotype may vary between locations (see further discussion in section 2.11). This phenomenon is called genotype x environment (G x E) interaction (Muthoni *et al.*, 2015). It is of major importance in developing improved varieties when it causes changes in rankings of genotype performance in different environments. This type of G x E is called cross-over interaction. This presents breeders with complications regarding which experimental varieties must be advanced when they obtain data from multi-location and multi-season environments.

It is desirable to develop cultivars that exhibit high yield in all target environments (Munawar *et al.*, 2013). However, a large G x E variance reduces heritability. Previous studies have reported that GEI lowers the correlation between phenotypic and genotypic values, thus, complicating the demonstration of superiority in a genotype (Cucolotto *et al.*, 2007).

Consequently, breeding progress and selection is compromised. It is, therefore, the desire of the breeding programme in Zambia to conduct multi-location trials and account for the G x E. This is because in theory, a successful breeding programme must either decrease the magnitude of G x E or exploit it by identifying genotypes that are specifically adapted in certain regions. Where cross-over interaction is significant among superior genotypes, locations with similar patterns of G x E can be identified and treated as mega environments or recommendation domains for release of the related genotypes (Fox *et al.*, 1997). Yield performance of a genotype is expected to increase with genetic improvement.

1.3 Problem statement

The yields of soybean in Zambia are low, averaging below 3.0 t/ha against a potential of 5.0 t/ha. This is attributed in part to poor availability of well-adapted and improved varieties. The spread of soybean production to new environments such as agro-ecological region III, calls for a continuous and rigorous investigation of genetic gain and G x E pattern before varieties can be released and recommended to growers. There has been a change in focus from breeding for rust resistance to breeding primarily higher yielding cultivars, resulting in a new set of advanced lines, which are intended for the Zambian environment. However, genetic gains have not been measured. Thus it is not known how much more productive and stable the new lines are when compared to earlier lines. This negatively impacts on variety release and strategy review. It is therefore imperative that the strategies used in the soybean-breeding programme in Zambia are evaluated for their effectiveness in producing higher yielding and stable varieties.

1.4 Importance of the study and summing up the research focus

Yield stability data is useful in determining the best performing genotypes for a given environment. With this information, the correct recommendation of varieties for specific or broad adaptation can be made. It serves as an effective way of managing cross over G x E so that farmers obtain higher yields when they grow the most productive genotypes for their given environments. Genetic gain provides a basis for estimating this increase in performance. It is expected to be high and positive, an indication of the success of a breeding programme in developing higher yielding and more adapted genotypes (Lange and Federizzi, 2009).Genetic gains show the efficiency of the strategies employed in a breeding programme so that corrective methods can be made where necessary. In addition to genetic gain, path coefficient analysis helps the breeder understanding the relationship between yield and associated traits to ensure effective and efficient exploitation of the given traits in selection for grain yield. Collectively, this information will result in well-adapted varieties with high yields and contribute

to greater soybean productivity in the country (for a superior package of traits, particularly yield, oil and protein content). Without this data, the contribution of poor adaptability to low yields would persist. The current trends of raising hectarage to increase production as illustrated in **Error! Reference source not found.** are unsustainable and costly to the farmer. As a developing country, Zambia has a finite availability of land competing for multiple developmental uses besides agriculture (Laurance *et al.*, 2014). Future increase in production will be near impossible without expansion into less productive non-traditional areas. The most sustainable option for higher production is therefore, improvement of crop yields at a rate sufficient to keep food prices low and prevent significant expansion of cropping area. This can be achieved by investing in the improvement of yield to close the gap between actual and potential yields realised (Schroeder *et al.*, 2013). The current study pursued identification of highly productive and stable genotypes among the advanced lines in the programme, and established whether there has been real progress towards identifying lines, which combine high productivity with high stability. The study also aimed to investigate whether indirect selection strategies that would improve yield by considering important secondary traits would be effective.

1.5 Research objectives

The main objective of the study was to improve soybean productivity in Zambia through identification of high yielding genotypes that are consistently well ranked and adapted to the Zambian environments.

1.5.1 Specific objectives

The specific objectives of the study were as follows:

a) To assess the nature of genotype x environmental interactions of soybean grain yield.

b) To identify consistently well ranked advanced soybean lines in medium altitude and subtropical environments of Zambia, Zimbabwe and Malawi.

c) To determine the genetic gains achieved in breeding for high yield and stability of 30 advanced soybean lines in the Zambian breeding programme between 1996 and 2007.

d) To determine secondary traits that made direct and indirect contributions to increase in yield potential realised from the soybean breeding program in Zambia between 1996 and 2007.

e) To identify the most ideal test environment for the genotypes under evaluation.

1.6 Hypotheses

The following research hypotheses were tested in the study:

a) The yield of soybean lines under investigation are affected by cross-over type of GEI

b) Some advanced lines under investigation have desirable yield and stability or adaptability in given environments

c) There have been positive genetic gains in soybean grain yield and related traits from soybean breeding between 1996 and 2007.

d) Some traits, having a secondary association with yield had directly or indirectly contributed to increased grain yield over years of breeding.

e) Close to ideal test locations i.e., most representative of other locations and most discriminative among genotypes exist for evaluation of the soybean lines used in the study.

1.7 Structure of the dissertation

The dissertation has the following structure.

Chapter One: Introduction

This chapter presents a brief background to the study undertaken, outlining the problem to be addressed by the study, the objectives to be met and the hypothesis behind each objective. Through this chapter, the gaps in research on the topic at hand are identified.

Chapter Two: Literature Review

This chapter reviews the origin and botany of soybean and defines key concepts pertinent to the study. Furthermore, a review has been conducted addressing production at a global, regional and national level (including constraints); genetic gains in soybean breeding; genotype x environment interaction in yield of soybean; the relationship between yield and its secondary component traits as well as different methods that are used to evaluate this relationship.

Chapter Three: Methodology.

This chapter outlines the different materials and methodologies to be employed to meet the set objectives in Chapter One as well as the methods used in the analysis of the field data.

Chapter Four: Results

Results of the field trials and their analysis are outlined in this chapter

Chapter Five: Discussion of results

A critical discussion and interpretation of the results obtained from the study has been conducted with reference to comparative studies.

Chapter Six: Conclusion and Recommendations

This chapter relates the findings of the study to the objectives set in chapter one as well as make some recommendations for future breeding programmes.

Chapter Seven: Recommendations

Based on the objectives and findings of the study, these recommendations are outlined for future soybean breeding programmes.

2 LITERATURE REVIEW

2.1 Introduction

This chapter reviews literature on topics that are relevant to the study to be undertaken. It establishes the botany, origin and development of soybean from its centre of origin to sub Saharan Africa. The objective of this chapter is to provide insight into the trait of yield and its improvement and how the environment in which a genotype is grown influences its expression. It also establishes methods that have been used in past studies to analyse genotype and environment interaction as a way of defining a breeding strategy to employ in a breeding programme. Environmental factors have a large influence on quantitative traits while their influence on qualitative traits is minimal. Inconsistency in genotype performance is exhibited as changes in the ranking order of genotypes across environments or as changes in mean performance of genotypes while maintaining rank order (Crossa *et al.*, 1995). These two expressions of GEI are termed, qualitative and quantitative GEI (Crossa *et al.*, 1995) respectively. Qualitative GEI is alternatively referred to as cross over GEI and is important in plant breeding because it reduces selection gains, thereby retarding genetic progress (Annicchiarico, 2002); several genotypes must be selected from different test locations, thus lengthening the selection process.

2.2 Botany of soybean

In order to devise an effective breeding strategy, it is prudent to understand the biology of the crop. Soybean is a self-pollinating crop, classified under the family Fabaceae, genus *Glycine,* subgenus *soja* and species *max.* Recent studies on the *Glycine* genus (Singh, 2010) suggest it to have evolved from a wild ancestor giving rise to two sub genera, *Glycine*, perennial in nature, and *soja,* an annual native to China, Japan, Korea, Taiwan and Russia (Singh, 2010). G. *soja*; 2n=4x=40, is made up of two annual species, the wild *G. soja* Sieb and Zucc and the cultivated *G.max* (L.), Merr; 2n=4x=40 (Hymowitz, 2008). The two members of subgenus *soja* are considered as one genome because they can be successfully crossed to produce viable hybrids, giving fertile F1 progeny. G. *soja* though possessing greater variability, has several undesirable growth characteristics (Hymowitz, 2004), hence the wide spread cultivation of G. *max*. However, some G. *soja* species are reported to provide useful sources of variation for trait improvement in cultivated soybean (Kanamaru *et al.*, 2006; Natarajan *et al.*, 2006; Krishnamurthy *et al.*, 2013).

2.3 Origin and development of soybean

The other important factor that strategists have to take into account is the origin of the crop. This has implications for breeding in that you need to understand the sources of diversity, natural pests and adaptation domains of the crop, among other factors. Scholars are of the consensus that China is the centre of diversity and origin of soybean cultivation. It is believed to have been first domesticated in northern china and spread to other parts of South East Asia (Singh, 2010) . It was later introduced to the United States where it was initially grown as a forage crop (Hymowitz, 2004). The development of lodging and shatter resistant varieties was responsible for changing soybean from a forage to an oilseed crop in that region (Bilyeu *et al.*, 2010).

In Africa, according to one account by Burkill (1966), soybean was introduced in the late 1800s. Its cultivation increased as demand for the crop grew in Europe. Turning to their African colonies as potential cultivators of the crop, the Europeans facilitated its spread across the continent (Shurtleff and Aoyagi, 2009). South Africa became the first sub-Saharan African country to implement trials at Cedara in Natal (Smit, 1987). This was the beginning of the soybean germplasm introduction programme which initiated soybean breeding in South Africa (Jarvie, 2008).

Outside South Africa, the most active soybean breeding programme in Southern Africa is in Zimbabwe which started in 1963. This was a direct result of collaboration with the South African Soybean Breeding Programme and introductions from the United States. These introductions resulted in the cultivar Hernon 147 (Gwata *et al.*, 2005). Soybean varieties bred in Zimbabwe have been introduced and adapted for commercial cultivation in the Zambian environment. The latest variety register (SCCI, 2014) shows that between 1973 and 2012, 36 soybean varieties were released in Zambia and only two of these; Lukanga and Mulungushi were developed by the government breeding programme. Of the 36 varieties, only three are still being cultivated today. Soybean variety development in Zambia, is dominated by the private sector, as indicated by the proportion of varieties released by seed companies (SCCI, 2014). Emphasis is placed on yield enhancement and rust disease resistance. Varieties with potential yields of up to 5t/ha have been reported and significant genetic gains have been attained (Shurtleff and Aoyagi, 2009). However, breeding gains for yield are predicted to decline so that productivity will depend on the ability of plant breeders to constantly adapt new varieties to changing environmental conditions. Therefore, evaluation of genotype and environment interactions, the primary purpose of this study, is a vital component of soybean breeding.

2.4 Importance of soybean

The introduction and wide spread of Soybean across the African continent, Sub-Saharan Africa in particular, is a testament to the usefulness and importance of the crop in on the continent. The grain is used in the extraction of edible oils while soybean cake a by-product of the oil extraction is utilised as a high-protein animal feed. However, many commercial varieties have shown higher protein value than oil in Nigeria (Giami, 2002), Brazil (Goldsmith, 2008), United States (Wilson, 2004) and even sub Saharan countries as South Africa, Zambia and Zimbabwe (Mushoriwa, 2013). Therefore, it is an important crop in the combating of protein deficiency diseases and malnutrition that plague the country of Zambia and many other developing countries (Keatinge *et al.*, 2011). As an animal feed, soy cake provides a relatively low cost, high quality protein feed source. Soybean meal is the preferred source for poultry feed and contains between 40% and 48% crude protein, after oil extraction (Ravindran and Blair, 1993).

Pimentel and Patzek (2005), compared ethanol production using various feedstock to biodiesel production using soybean. They concluded that biodiesel can be produced at lower cost using soybean. However, soybean is not yet utilised for production of biofuels in Sub-Saharan Africa. Soybean also improves soil fertility through nitrogen fixation. This is of tremendous benefit in African farming systems, where soils are highly deficient in nitrogen due to nutrient mining and inadequate nutrient replenishment (Laker, 2013). Soybean is therefore, a very important legume crop whose uses are wide and is an important dietary component for both human and livestock consumption.

2.5 Soybean production

In response to the growing demand for soybean products, production of the crop has shown tremendous expansion on a global scale. The largest producer of the crop is currently the United States of America (Ullah *et al.*, 2012), closely followed by Brazil, Argentina (Table 1), China and India (Wilcox, 2004).

Zambia's position as a soybean producing country in the region has allowed for the development of a breeding programme with full-fledged objectives to guide its breeding activities.

2.6 Soybean breeding objectives in Zambia

The soybean breeding programme in Zambia has a number of breeding objectives. These include breeding for increased resistance to diseases as rust and frog eye leaf spot. Frog eye leaf spot is a common disease of soybean caused by *Cercospora sojina* (Mian et al., 2008).

Table 1 Global soybean production of top 20 producers (000' metric tons).

Country	2010	2011	2012	2013
USA	90,605	84,192	82,055	89,483
Brazil	68,756	74,815	65,849	81,724
Argentina	52,675	8,889	40,100	49,306
China	15,083	4,485	12,800	11,951
India	12,736	2,214	14,666	11,948
Paraguay	7,460	8,310	4,345	9,086
Canada	4,345	4,246	5,086	5,198
Uruguay	2,000	1,830	3,000	3,200
Ukraine	1,680	2,264	2,410	2,774
Bolivia	1,693	1,861	2,061	2,347
Russian	1,222	1,756	1,806	1,636
South Africa	566	710	650	785
Indonesia	907	851	843	780
Italy	553	565	422	625
Nigeria	365	493	650	600
Serbia	541	441	281	385
Korea	350	350	350	340
Zambia	112	117	203	261
Total	**262,077**	**258,831**	**238,031**	**272,875**

Incidence of the disease is known to be prevalent in the wetter, humid regions of Zambia such as Mpongwe and Mkushi. Yield losses up to 60% have been reported (Tchagwa, *personal communication*). Therefore, the best method of control is to plant resistant cultivars. Like soybean rust, *Cercospora sojina* is a dynamic pathogen with extensive virulence or race diversity. So far 11 races have been identified making breeding for resistance complicated (Mian *et al.*, 2008). Another important objective is the improvement of quality traits such as oil and protein percentages. Above all these objectives is the development of high yielding and stable cultivars. However, yield is a qualitative trait for which direct selection is complicated by

environmental interaction. To combat this problem, indirect selection for yield as well as other traits may be necessary.

2.7 Path analysis

The indirect selection of a quantitative trait such as yield requires a complete determination of how a component trait will affect the success of the selection process. Path analysis generates path coefficients which partition the correlation coefficients into their direct and indirect influences on a dependent variable such as yield (Cramer and Wehner, 1998, 2000). Direct or indirect causal effect is implied.

Researchers have successfully used this method to predict how a component trait affects the success of the selection process in soybean breeding. Mushoriwa (2013), observed that the traits with the highest positive and significant direct effect on yield potential among 42 genotypes in Zambia and Zimbabwe, included number of nodes per plant (0.48); followed by plant height (0.27); and 100 seed weight (0.20). These findings suggest number of pods per plant as the most significant component for efficient selection for yield improvement. These observations are similar to those reported by Arshad *et al.* (2014) for positive direct effect of 100 seed weight (0.292) and those observed by Malik *et al.* (2007), for direct effect of plant height and number of pods per plant. In addition, Malik *et al.* (2007) observed maximum direct effect on yield for days to flowering completion (32.75) followed by days to pod initiation (19.46), he concluded that, days to flowering completion was the most important selection criteria. Machikowa and Laosuwan (2011), observed high direct and indirect contributions via pods per plant; pods per plant could be used directly or indirectly as the selection criterion for identification of high yielding genotypes in early maturing soybean.

Variations in path coefficients and overall findings can occur due to the genotype and environmental differences involved. In light of the foregoing, it is necessary to carry out path analysis, as opposed to correlation analysis alone, in order to determine the most effective traits to emphasise in the indirect selection for yield potential for future improvement with these genotypes, across the varied soybean growing environments in Zambia. Effective selection is crucial to raising genetic gains of a breeding programme. That is, Increase in performance that is achieved through artificial genetic improvement programmes after one generation or cycle of breeding.

2.8 Yield genetic gain studies in soybean

Genetic gains are an important part of a breeding programme. These gains are constantly evaluated after a period of time. Researchers have in the past, demonstrated a relative

increase in grain yields of Soybean over years of breeding (Egli, 2008a; Lange and Federizzi, 2009). This is attributed to increased interest in the activity of farming, advances in agronomic practices as well as new technologies. The increase in the number of seed companies' engaged in crop improvement through genetic advances also means that current soybean cultivars have greater yielding potential.

Crop improvement is achieved through a series of cycles during which desirable genes are exchanged by cross-pollination and fixed by selfing in a variety. In soybean this may be by single seed descent advancement to ensure that these genes are fixed. Their compounded effect over time is exhibited in the productivity of the soybean variety developed. Large positive increments improve the productivity of the crop allowing for market demand to be met, while small or negative changes impact negatively on productivity and consequently, on food security. Hence, evaluation of the breeding gains attained over time remains critical.

Breeding gains are always based on a given period and differences in breeding strategies used in cultivar development (Duvick, 2005; Egli, 2008a; Lange and Federizzi, 2009; Tefera *et al.*, 2010). In Brazil, Lange and Federizzi (2009) predicted the genetic gains of three maturity groups of soybean from four breeding programmes, over a 20 year period. Yield gains were found to be between 0.87% and 3.49% per annum. In a separate study, similar results were observed by Egli (2008a), when yields of soybean were evaluated in six American states from 1950 to 2005. It was found that soybean yields increased at a rate of 1.5% per year within the first 40 years and declined to 1.4% per year in subsequent years. The decline was attributed to intensification of selection using common elite parental lines, which narrowed the genetic base and caused a reduction in the genetic gains.

In Africa, Nigeria in particular, Ogoke *et al.* (2003) implemented a study over two cropping seasons under different fertilizer management systems. Four varieties were evaluated and the results showed that grain yield, for the new varieties was 58% higher than the old less improved varieties at different levels of phosphorus so that genetic gain was 0.6%. This may seem like a small increment but it is significant in a self-pollinating crop like Soybean. Increased performance may have been attributed to the better response of the new varieties to P fertilizer application. This study showed that yield gains can be influenced by changes in growing environments, as well as the contribution of plant breeding. The two categories interact as was the case in Ogoke *et al.* (2003) findings; improved varieties out-yield older ones when evaluated in the same environment (Tukamuhabwa *et al.*, 2012), therefore yield response is a function of both varietal improvement and the growing environment (GEI).

Although genetic gains have been reported in the reviewed literature, the global rate at 1.3%, is insufficient to meet the United Nations target of doubling crop yields by 2050 (Ray *et al.*,

2013), in order to meet the needs of a growing population. The world population is estimated to reach 9.7 billion by the year 2050. According to Tilman *et al.* (2011), crop demand is estimated to increase by 100% to 110% by the same year. Crop yield growth is, thus the most sustainable way to ensure future food security. As yield growth plateaus have not yet been reported in soybean breeding (Egli, 2008b), it can be concluded that there is still significant variation available to increase yield potential. However, increasing annual rates of yield improvement will require technological developments and innovations as suggested by Specht *et al.* (1999).

In order to adequately estimate genetic gain, a breeding programme must determine the genetic variation that can be passed on and expressed in an end product.

2.9 Heritability

Genetic variation of a trait is used to calculate heritability, a breeding tool to help implement effective selection strategies. Heritability is defined as the proportion of observed variation that can be genetically passed on to the next generation with each successive breeding cycle (Hallauer *et al.*, 2010). It is important as it allows the breeder to determine the best method of effectively transferring desirable genes and effecting genetic gains in a trait of interest.

Heritability values range from 0.0 (genes do not contribute at all to phenotypic individual differences) to 1.0 (genes account for all individual differences). A high heritability value would favour direct selection for a given trait and produce better response to selection. A low heritability on the other hand would require indirect selection for most effective response to occur.

Heritability is known to occur as broad sense and narrow sense heritability. Broad sense heritability is the ratio of genetic variance to phenotypic variance, expressing the extent to which individual phenotypes are determined by genotypes (Nyquist and Baker, 1991). Narrow sense heritability, on the other hand, is the ratio of additive variance to phenotypic variance, expressing the extent to which phenotypes are determined by the genes transmitted additively from the parents to offspring's (Fehr, 1991). In this study, the broad sense heritability was estimated using variance components. Previous studies have been done to estimate heritability for yield and other important traits in soybean. Karasu *et al.* (2009) as well as Malik *et al.* (2007), both observed low to moderate heritabilities for the traits of yield, pods per plant, branching and hundred seed weight. Mushoriwa (2013), who also investigated heritability later, observed moderate level of heritability particularly for yield and low heritability for protein and oil content.

2.10 Genotype and environment interaction in soybean

Another important factor in determining genetic gain is the environment in which the varieties are being evaluated. However, the environment in itself cannot be considered in simplicity because interactions occur between genotypes and the concerned environment for expression of a trait. Therefore, a clear understanding of this interaction is vital to the determination of selection strategies to employ in a breeding programme. It is widely accepted that multi-location data constitutes pattern, and noise with the noise being part non-structural and part due to genotype and environment interaction (G x E) (Crossa, 1990). Breeders, therefore, strive to increase the structural pattern which represents interpretable and predictable response of genotypes (Crossa, 1990) and reduce noise from GEI by breeding for stability. Annicchiarico (2002) cited consistency in performance as necessary between the components of the breeding target, components in this case referring to not only location but also practices, seasons and aspects that can be controlled or predicted. Based on results of previous studies (Cucolotto *et al.*, 2007; Tukamuhabwa *et al.*, 2012; Ngalamu *et al.*, 2013), the larger the GEI the lower the stability of the genotypes under evaluation.

2.11 Yield stability studies and variability in the medium altitude environment

The diverse soybean growing environment under investigation in this study warrants an extensive genotype x environment interaction analysis to determine the best strategy to aim for in the breeding programme (Crossa, 1990).

2.12 Adaptation strategies

There are two possible strategies that may be implemented from multi-location trial data (Crossa, 1990). These are specific and wide adaptation strategies (Annicchiarico, 2002). However, there is a gap in knowledge regarding the type of adaption that is important in soybean for the medium altitude environments in Zambia, Malawi and Zimbabwe. Nonetheless, these two concepts of adaptation are discussed in the next section of this chapter, while the type of adaption that condition performance of soybean in the three countries have been investigated with respect to 30 advanced soybean lines in the current study (See Chapters 3 and 4).

2.12.1 Specific adaptation strategies

There are genotypes that may be specifically adapted to certain environments. A genotype that is consistently well ranked across a limited number of locations is said to be specifically adapted to those locations (Fox *et al.*, 1997). This occurs due to significant qualitative GEI between genotypes and location (G x L). Seed companies and other large breeding programmes aiming at releasing cultivars in several countries or an entire region exploit this phenomenon. By exploiting the positive relationship between genotype and environment, the breeder can develop homogeneous zones of genetic response, lowering GEI and increasing genetic gains from selection in those locations (Annicchiarico, 2002; Annicchiarico *et al.*, 2005); as a result, yields are increased. In their evaluation of the 24 wheat cultivars in three seasons over 47 environments, showed that specific adaptation increased genetic yield gains by 2% to 7% above that of wide adaptation. In soybean breeding, Gurmu *et al.* (2009) reported significant GEI for protein and oil content showing specific adaptation to particular locations. The yield studies that have been conducted elsewhere (Karasu *et al.*, 2009; Tukamuhabwa *et al.*, 2012) have also shown preference for this type of strategy.

2.12.2 Broad adaptation strategies

Another strategy that can be implemented is the broad adaption. In this case genotypes that are well ranked across all environments are said to possess wide adaptation (Annicchiarico, 2002). This occurrence is a result of significant but quantitative GEI across locations. Therefore, genotypes performing well in one environment will perform well in other environments as well (Falconer, 1960). This strategy is cheaper and easier to implement as less seed is required and testing can be done at fewer locations. With the target of breeding for wide adaptation Cucolotto *et al.* (2007) selected for genotypes of high and predictable yields with wide adaptability of 30 soybean cultivars from three different maturity groups over three seasons and 30 environments. Of the 30 genotypes, only four, CD 202 (early), M SOY 7202 and CD 206 (semi-early), and M SOY 7602 (medium) had wide adaptation and high yield. This shows how small the level of observable wide adaptation is. The heterogeneity of most growing environments favours a specific adaptation strategy over a wide one. Therefore, it is prudent to conduct multi-location and over season trials to determine the type of adaption for each advanced line in the breeding programmes.

2.13 Genotype x environment interaction - for economic traits (yield, oil, and protein)

Yield, oil content and protein content being quantitative traits are strongly influenced by the environments in which genotypes are evaluated. Significant genotype x environment interaction for yield has been shown in many studies around the world (Cucolotto *et al.*, 2007; Gurmu *et al.*, 2009; Karasu *et al.*, 2009; Tukamuhabwa *et al.*, 2012). In Brazil, Cucolotto *et al.* (2007), in their assessment of 30 soybean cultivars for adaptability and stability reported significant qualitative interaction over three seasons in sixteen locations. With only four cultivars displaying wide adaptation. These findings are similar to those done in Uganda (Tukamuhabwa *et al.*, 2012) across five locations. However, Tukamuhabwa *et al.* (2012), observed significant quantitative GEI and recommended that the genotypes be bred for general as opposed to specific adaption for grain yield.

Oil and protein content evaluations have shown significant interactions with environments in which selection occurs. According to Gurmu *et al.* (2009), there was strong interaction between the environment and three traits; oil content, protein content and grain yield. With genotypes selected for high oil content and protein content showing inconsistencies in ranking across environments. Gurmu *et al.* (2009), therefore recommended for specific locations. Considering how varied the African production environment is, Zambia in particular, it is important to determine the expression of these traits in the genotypes under investigation across locations.

2.14 Methods of evaluating GEI

There are many methods of evaluating GEI from multi-location data. These have been reported in previous studies (Gauch Jr, 1992; Fox *et al.*, 1997). However, in this study, the Lin and Binns (1988) measure of superiority stability statistic additive main effect and multiplicative interaction (AMMI) and GGE biplots were used. These methods including the traditional regression approach are reviewed in the current study.

2.14.1 Regression analysis

Despite its apparent limitations, many researchers have recently used the traditional linear regress approach for G x E study. The linear regression analysis, which was developed by Finlay and Wilkinson (1963), is sometimes used in the analysis of multi-location data for genotype adaptation to growing environments. Even though it was not used in this study, its importance in multi-location evaluation cannot be overlooked. This method involves the regression of a single genotype's yield on the mean yield of all genotypes involved in the study

for each test locations (Romagosa and Fox, 1993). It therefore uses environmental means alone as a measure of genotype adaptation response. The limitations are explained by the fact that genotypes are confounded in the environmental mean so that regression lines cannot be considered completely independent (Kearsey and Pooni, 1998). The test genotypes, therefore, must be replicated in the same order with all other genotypes, making evaluation over seasons difficult, this would limit further use of results generated in this study to one season. However during the past decade it has been used in soybean breeding (Cucolotto *et al.*, 2007; Gurmu *et al.*, 2009), but hardly in isolation with the other methods. The regression method is criticised for its poor precision over other methods and its inability to handle multiple trait analysis. Multiple trait analysis is a necessary part of GEI analysis (Kearsey and Pooni, 1998).

2.14.2 Additive main effect and multiplicative interaction (AMMI)

Given the limitations of the linear regression, in this current study, the Additive Main Effect and Multiplicative Interaction (AMMI) model was used to determine yield stability of the soybean lines and to estimate the pattern of interaction between test locations and genotypes through its ranking function. AMMI combines a univariate method for additive effects of genotypes and environments, with a multivariate method for the multiplicative effect which is also the interaction of the genotypes with the environment (Gauch Jr, 1988). This model has a three-fold function of addressing the following: firstly it looks at the which-won-where pattern of data; secondly the AMMI biplots allow visualization of the mean genotype performance; and finally test environment evaluation (Yan *et al.*, 2007). It is increasingly used in soybean breeding data analysis (Cucolotto *et al.*, 2007; Gurmu *et al.*, 2009; Tukamuhabwa *et al.*, 2012; Amira *et al.*, 2013).

2.14.3 GGE biplots

The other increasingly common approach for G x E data analysis has been the GGE-biplot. The GGE biplots methodology for graphical analysis of multiple environment yield data was developed by Yan *et al.* (2000). The abbreviation GGE refers to the genotype main effect (G) plus the genotype x environment interaction (GE). These are the two important sources of variation in genotype evaluation. A biplot is a plot that shows both the genotypes and environments under evaluation. It is constructed by plotting the first two principal components (PC1 and PC2). If the principal components of the biplot are significant, that is, explain much of the variation observed, a GGE biplot analysis (Yan and Tinker, 2006) can be used.

This method has been used successfully in soybean breeding to generate groups showing: “which-won-where” pattern (Tukamuhabwa *et al.*, 2012) trait comparisons and comparison of

genotypes on the basis of yield and stability (de Oliveira *et al.*, 2005) as well as the best test environment (Yan and Rajcan, 2002). In this study, it was necessary to identify the most ideal test environment (representative and able to discriminate performance of genotypes) for future evaluation or advancement of the genotypes as well as to determine the most stable of genotypes. Therefore, a methodology such as GGE was needed (see Chapter 3.9.1).

2.14.4 AMMI model vs GGE biplot

There was need to compare the G x E study tools in order to select the most appropriate for the current study. In this regard, the Additive Main Effect and Multiplicative Interaction (AMMI) biplot has been criticised for its lack of an inner product property which some researchers believe to be the most important property of a true biplot (Yan *et al.*, 2007). However, this argument is not entirely valid as the vectors of an AMMI biplot are a function of the vector length and angle between vectors, the equivalent of an inner product.

$$Cos(\theta) = \frac{Inner\ Product}{Length\ of\ vector} = \frac{X^T Y}{\|X\|\|Y\|}$$

The GGE biplot has been reported to explain more genotype and GE than the AMMI1 graph (Yan *et al.*, 2007). In the analysis of a rice dataset in Uganda, Samonte *et al.* (2005) employed both the GGE biplot and the AMMI1 graph which explained 77.3 and 64.6% of the total G+GE, respectively. This implied that the GGE-biplot was more effective. In contrast, a comparison of GGE and AMMI genotype discriminating powers by Amira *et al.* (2013) showed that both methods gave strongly reliable results. However, Ye *et al.* (2001) observed better representation from GGE biplot analysis arguing that test-environment evaluation has not been thoroughly researched in AMMI analysis. Yan *et al.* (2007), later reported that the mean performance and stability view of the GGE biplot was superior to the AMMI biplot as it explained more genotype and genotype x environment effects giving a more accurate presentation of the data. Based on past research, it is clear that GGE and AMMI biplot are invaluable tools for GEI analysis and their use seems to be a matter of preference. For this current study, both methods were used in order to get the most out of the data generated.

2.14.5 Superiority measure of cultivar performance

The fourth method that was considered and used in the current study was the cultivar superiority index. This is a measure of yield stability across locations which was coined the superiority measure as proposed by Lin and Binns (1988). This method presents an easier way to identify specific adapters in given locations. It is based on setting a maximum response, M_j, being set across locations or seasons and comparing this to the mean, X_i, of a genotype across all locations. A superior genotype will have a smaller superiority index compared to

less superior genotypes. In sugarcane multi-location trials, de Oliveira *et al.* (2005) were able to select the same clones using the superiority measure as with other measures of stability. In soybean breeding, this method was successfully employed by Jarvie and Shanahan (2009) to determine superior and adapted genotypes under rust disease stress. The three methods were therefore integrated to investigate G x E of advanced soybean lines in the current study.

2.15 Summary

The literature reviewed showed that production of soybean on the African continent, particularly Sub-Saharan Africa is still very low relative to global production, accounting for less than 1% of all production. This is attributed to the low yields realised by growers of the crop who are facing a number of constraints among which is the poor availability of well adapted genotypes.

Genetic gain and its importance in plant breeding and genotype improvement were also discussed. It was clear from the literature that genetic gains have been achieved in past studies; however, the annual gains are not high enough to meet the needs of the growing population as predicted for 2050. Therefore, breeders must continue striving to obtain higher genetic gains in their breeding programmes.

Protein and oil content are the two most important seed composition traits in soybean grain. Tremendous progress has been made in improving the world's soybean genotypes for these traits. Literature reviewed showed that oil content is easier to improve due to its positive relationship with yield. Protein content on the other hand has an inverse relationship with yield so that increasing its levels leads to a reduction in yield. Both of these traits are quantitatively inherited and highly affected by the environment in which they are selected.

Soybean rust disease was also discussed in the literature. The multiple virulence gene action of the causal organism has so far prevented the development of genotypes with complete resistance to the disease. Cases of success that soon broke down were reviewed. This occurrence has perpetuated the continued breeding for complete resistance. In the meantime, tolerance is the first line of defence for the disease as a means of reducing complete losses to soybean producers.

Literature reviewed also showed that yield is quantitatively inherited with low heritability, which makes selection for high yielding genotypes complex. Many studies were reviewed which indirectly selected for other agronomic traits that unlike yield were highly heritable and positively correlated to yield. These were shown to be easier to select for. The PathSAS macros of the SAS computer software were shown to successfully determine traits directly or indirectly contributing to increased yield.

Another important phenomenon discussed in the literature reviewed was genotype x environment interaction. Yield is highly affected by GEI. Therefore, multi-location trials should be used to extensively ascertain the nature of this relationship. The literature reviewed showed that a significant GEI may require the implementation of a specific adaptation strategy. However, another possibility is a wide adaptation strategy. Breeders must therefore know what strategy best suits the genotypes being evaluated.

Literature available did not show up to date data on G X E studies done in Zambia, Malawi and Zimbabwe, however, it did show that yields and production are still lower than other major soybean producing countries in the world. It is not clear what kind of genetic gains have been realised for the Zambian, Malawi and Zimbabwe medium altitude breeding programmes. However, it can be speculated from the low yield that they are likely to be low. This therefore justifies dedicating resources to research study that is reported in the next chapters.

3 METHODOLOGY

3.1 Introduction

This chapter outlines the different materials and methodologies that were employed to meet the set objectives in Chapter One as well as the methods used in the analysis of the field data.

3.2 Germplasm

Thirty advanced soybean germplasm lines derived from the crossing of parental lines in the soybean-breeding programme between 1996 and 2007 were used in the study. For the convenience of the study, these lines were coded G1-G30 and evaluated in the 2015/2016 cropping season. Among them, seven commercial cultivars, were included in the trials as checks, they were coded G3, G7, G10, G11 G15, G18 and G29 for ethical reasons. The commercial checks were selected for their high yielding potential and known representation of different maturities and superior agronomic traits.

3.3 Test Locations

The yield trials were conducted in three countries, Zambia, Malawi and Zimbabwe, across 16 locations (Table 2). These represented agro-ecological regions of soybean production in the three countries.

a) Zambian growing environment

In Zambia, the growing environment is divided into three agro ecological regions I, II and III. These divisions are based on the amount of rainfall received annually and the average temperatures expected in a growing season. Region I, also known as the low rainfall region receives the lowest amount of rainfall with a mean of <800 mm per annum. Rainfall is erratic allowing for long dry spells over a short cropping season of 80-120 days. Region II; the medium rainfall region of Zambia is characterised by mean annual rainfall of 880 to 1000 mm, allowing for a longer cropping season of 100 – 140 days. It covers much of central Zambia, with the most fertile soils and accommodates most of the country's commercial farms. Distribution of rainfall is not as erratic as in Region I, but dry spells are common and affect crop yields. Average mean daily temperatures range from 23- 26°C in the hottest month October to 16-20°C in the coldest months of June and July. Region III is the high rainfall region of Zambia, it covers the northern region of the country and is characterised by annual rainfall of >1000 mm. It also has the longest rain fed growing period of 120-200 days, characterized by extreme acidity and Aluminium toxicity due to the excessive rainfall. The Zambian sites used in this

study are spread out across region II and III. Mpongwe and Somahwe are high potential sites in region III while Lusaka west and Kabwe are in region II though Kabwe is closer to the belt of region III and may on occasion receive higher annual rainfall than Lusaka. Mkushi on the other hand is partly within regions II and III. Therefore, the degree of environmental variation is expected to be high from one site to the next.

b) Zimbabwe growing environment

In Zimbabwe, the country is divided into five main agro-ecological regions according to differences in rainfall. The sheer number of regions highlights the expected amount of variation for the different locations at which multi-location trials can be set up, if they are to be representative of all growing environments. Annual rainfall is highest in region I, >1000 mm, and covers approximately 2% of the land area. Natural region II, receives between 750 mm and 1000 mm of rainfall and is be most suited to intensive farming based on crop production. Natural region III is a semi-intensive farming region with moderate rainfall of between 650 mm and 800 mm. However, severe mid-season dry spells are not uncommon in this region. Natural region IV is a semi-extensive farming region characterized by seasonal droughts and severe dry spells during the rainy season. Rainfall received is the lowest here, between 450 mm and 650 mm. Crop production is therefore risky. Natural region V is an extensive farming region covering only 27% of Zimbabwe. Rainfall in this region is too low and erratic for the reliable production of even drought resistant grain crops. Extensive cattle or game ranching is the only sound farming system for this region. Nine of the 16 locations used in this study fall across at least three of these agro ecologies.

c) Malawi growing environment

Like Zimbabwe, Malawi is also divided into five agro-ecologies. These include the highlands, escarpments, plateaus, the lakeshore and upper Shire region and finally the Lower Shire valley. The climate in Malawi changes from semi-arid in the Lower Shire Valley, semi-arid to sub-humid on the plateau and sub-humid in the highlands. Most of the country receives rainfall of between 763 mm -1,143 mm per annum.

The highlands consist of isolated mountains between 1,320-3,000 masl. The escarpments are associated with major fault lines along the edge of the Rift Valley, they are also found around the highland plateau and mountains. Three quarters of Malawi consists of plateau at elevations of 750-1300 masl. The topography is flat to rolling, with scattered rock. The soil is deep and well drained on higher parts, with poorly drained sand and clay in the hollows.

The Lakeshore and Upper Shire Valley is flat to gently undulating, with deep soils in the hollows. Soils are similar to those along the lakeshore. The Lower Shire Valley extends from

Kapachira falls to Nsanje at the bottom of the country at less than 180 masl. Makoka research station, a location used in this study experiences annual average minimum and maximum temperature of 15.6^{0}C and 25^{0}C, respectively. The station receives an annual average rainfall of 1,044 mm most of which falls in five months from November to March. Bvumbwe on the other hand falls within the highlands agro ecology. It is characterised by annual rainfall of about 1,219 mm most of which falls mainly in the months of December to April. Frequent mist and drizzles and occasional frosts during the months of May to August.

Table 2 Description of trial locations used in yield evaluation during the 2015/16 season

Country	Trial Site	Code	Date planted	Latitude	Longitude	Altitude
Zimbabwe	ART Farm	E1	18-11-15	17°43'00'S	31°0500'E	1527
Zimbabwe	A.U	E2	03-12-15	19°10′00″S	32°25′00″E	976
Zimbabwe	Banket	E3	29-12-15	17°22′59″S	30°23′59″E	1277
Zimbabwe	Bindura	E4	06-01-16	17°18′06″S	31°19′50′E	1118
Malawi	Bvumbwe	E5	15-12-15	15°55'00'S	35°04'00'E	1228
Zambia	Kabwe	E6	28-12-15	14°19′59″S	28°25′00″E	1174
Zimbabwe	Kadoma Research Center	E7	12-12-15	18°19′59″S	29°54′55″E	1176
Zambia	Lusaka West	E8	24-11-15	15°67′00''S	28°33'00'E	1300
Malawi	Makoka	E9	21-12-15	14°41′59″S	35°36′00′E	837
Zimbabwe	Mazowe	E10	03-12-15	17°10′00″S	31°00′00″E	1249
Zambia	Mkushi	E11	07-12-15	13°01′00″S	28°46′00'E	1276
Zambia	Mpongwe	E12	05-12-15	13°30′32″S	28°09′18'E	1195
Zimbabwe	Panmure	E13	08-01-16	17°16′37″S	31°36′31′E	925
Zimbabwe	RARS	E14	09-12-15	17°40'00'S	31°14'00'E	1341
Zambia	Somahwe	E15	02-12-15	13°32′00'S	28°09′00'E	1182
Zimbabwe	Stapleford	E16	09-12-15	17°49′39″S	31°03′12″E	1494

RARS- Rattray Arnold Research Station; ART- Agricultural research trust; AU- Africa University

3.4 Experimental design

The experiments were laid out in a 6 X 5 α-lattice design replicated three times at each location. Each entry served as a treatment. Six row plots, 5 m long with an inter-row spacing of 0.45 m (gross plot area- 13.5 m^2) and intra-row spacing of 0.06 m were used, providing a plant population of 370, 000 plants/ha at seeding rate of 80 kg/ha.

3.5 Management

Recommended cultural practices for soybean production were followed at all the trial locations. Land was prepared to a fine tilth for increased seed soil contact. The seed of each entry was

inoculated with *Rhizobium japonicum (Kirchner)* strain at a rate of 300 g/ha before planting by hand. This was done to facilitate biological atmospheric nitrogen fixation into the soil as a source of nitrogen for seed development. The soil at all locations was further furnished with a basal fertilizer, Soya blend (5%N; 12%P_2O_5; 24%K_2O; S5%; Zn0.4%; B0.10%), applied at a rate of 300 kg/ha.

Planting was done on the dates that are indicated in Table 2. Seed was sown in furrows at a depth of 20 mm and then covered with soil. The experiments were implemented under rain-fed conditions at all sites. Supplementary irrigation was applied where rainfall did not occur immediately after planting, in order to bring the soils to field capacity. Integrated pest management, which involved the use of pre- and post-emergence herbicides as well as pesticides at the rates that are indicated in Table 3, was employed at all test locations. No fungicides or additional chemicals were used. This was to allow the genotypes expression in each environment to be fully observed without external manipulation. The net plots were hand harvested to minimize mechanical shattering

Table 3 Chemicals used and their rates according to country

Country	Pre-emergence herbicide	Rate	Post- emergence herbicide	Rate	Pesticides	Rate
Zambia	Dual Magnum Zephyr	1l/ha 0.5l/ha	Fusillade Chloromuron	2l/ha 17.5g/ha	Cyperforce	0.75l/ha
Zimbabwe	Metolachlor	1.5l/ha	Basagran	3l/ha	Thionex	35-60g/15l

3.6 Data Collection

Over the course of the season, data were collected from the middle four rows which made up the net plot area of 7.92 m^2 ([0.45 m * 4rows] *4.4 m); each row was adjusted by 30 cm on either end to minimise border effects. Data was collected using standard procedures that are used at Seed Co. Two data sets were collected for performing different analyses

3.5.1 Yield stability data:

Data for the following yield traits (Table 4) were collected from all plants within each net plot with the aid of field recorder electronic devices powered with an android operating system

Table 4 Description of yield stability traits collected from all trial locations

Trait	Description of trait	Method of collection
Days to flowering	Days until 50% of the plants had flowered.	Counting the days from planting until 50% of the plants in a plot had an open flower
Days to maturity	Days until 95% of the pods had dried.	Counting the number of days from planting to when 95% of the pods had dried.
Days to shattering	Days from physiological maturity to first pod shattering	Counting the number of days from when 95% of the pods had dried to shattering of the first pod
Lodged percentage	Percentage of lodged plants at maturity	Visual estimate of plants leaning more than 45° to the soil surface in each plot
Seed appearance Sores	Visual quality of seed in terms of colour and shape.	Scale of 1 to 5 where 1 was very good quality and 5 was very poor quality with much discoloration, mold and cracking
Percentage purple stain	Number of purple stained seed in a sample of 100 seed.	Counting a sample of 100 random seeds from the net plot and recording the purple stained seed as a percentage of the seed sample.
Plot yield (PYLD)	weight of seeds per plot in grams	Weighing all seed harvested from a net plot
100 seed weight	Weight of 100 dry seed	randomly weighing 100 seeds in grams on a scale at maturity at standardized to 11% moisture content.
Pod height	Average height of 5 plants from the ground surface to lowest pod	Measuring stick (cm)
Plant height	Average height of 5 plants from the ground surface to the top leaf	Measuring stick (cm)
Red leaf blotch score	Appearance of red blotches on leaves	Scored at the R6 stage. Scale: 0-6 where 0 = resistant or absent of symptoms and 6= very susceptible

Trait	Description of trait	Method of collection
Rust disease scores	Appearance of soybean rust symptoms on leaves	Scored at R6 stage using a dual digit scoring system of 1-3 adapted from Shanmugasundaram (1977)
Bacterial blight score	Appearance of blight spots on leaves	1-5 scale where 1 is resistant and 5 very susceptible.
Protein content	Percentage protein content of seed	measured by Near Infrared Reflectance Spectroscopy method with an Inframatic 9500 machine
Oil content	Percentage oil content of seed	Near Infrared Reflectance Spectroscopy method with an Inframatic 9500 machine
Moisture content	Percentage moisture in grain	Inframatic 9500 machine
Plot yield	Weight of grain in the net plot	Weighed in grams and adjusted as in Equation 4.

Plot yield (PYLD) adjusted to Kg ha^{-1} at 11% moisture (SYLD) using the following formulae;

$$SYLD = \frac{PYLD\ (g/net\ plot) \times 10 \times (100 - \%MC)}{100} \times 1.1 \qquad \text{Equation 1}$$

Where:

%MC = Grain moisture in percentage

PYLD= Plot yield

3.5.2 Correlation and path analysis data

Data for traits in Table 5 were collected from 25 randomly sampled plants per net plot and the mean recorded for each entry following the protocols used at Seed Co. Due to logistical reasons, the samples were taken from two sites only; Rattray Arnolds research station (RARS) and Stapleford (SRC),

Table 5 Description of correlation and path ananylis data collected from two trials

Trait	Description of trait	Method of collection
Pods per plant	Number of pods per plant	Counted and recorded as average of 25 plants/ net plot at maturity
seeds per plant	Number of seeds per plant	Counted and recorded as average of 25 plants/ net plot at maturity
Seeds per pod	Number of seeds in a pod	Quotient of number of seeds per plant and number of pods per plant
branches per plant	Number of branches per plant determined through a visual count at maturity	Counted and recorded as average of 25 plants/ net plot at maturity
nodes per plant	Number of nodes per plant will be recorded after a visual count at reproductive stage	Counted and recorded as average of 25 plants/ net plot at maturity
Weight of seed per plant	Weight of seed per plant in grams calculated as an average of five plants	Counted and recorded as average of 25 plants/ net plot at maturity
Bio-Mass	Above ground dry weight per plant	Average weight of 5 plants within a net plot (grams)
Harvest index	Ratio of weight of seed /plant and above ground dry weight of plant	Determined by dividing weight of seed per plot by above ground dry weight of plant

Harvest index was calculated for each entry as:

$$Harvest\ Index\ (HI) = \frac{weight\ of\ seed\ per\ plant}{Above\ ground\ dry\ weight\ of\ plant}$$ Equation 2

This reading was necessary to ascertain the reproductive efficiency of each genotype.

3.7 Data analysis

3.7.1 Analysis of Variance

A combined analysis of variance (ANOVA) was carried out for grain yield and related traits (Table 6) as a mixed effects model with genotypes as fixed effects and locations as random effects using the breeding management system software (BMS, 2014) as follows:

$$Y_{ijkr} = \mu + g_i + e_j + r(e)_j + \beta(er)_{ik} + (ge)_{ij} + \varepsilon_{ijkr} \quad \text{Equation 3}$$

Where:

Y_{ijkr} = Mean yield of *i*th genotype in *j*th environment in *k*th block within the rth replication; μ= the overall mean; r (e)$_i$ =the effect of the rth replication within jth environment; $\beta(re)_j$= the effect of the *k*th block in the rth replication and *i*th environment; g_i= the main effect of the ith genotype; e_j= the main effect of the *j*-th environment; $(ge)_{ij}$ = the interaction of the *i*th genotype with the *j*th environment; ε_{ijkr}= the random experimental error term associated with the mean of the *i*th genotype in the *j*th environment in the *k*th block within rth replication. The skeleton of ANOVA is shown in Table 6.

Table 6 Skeletal analysis of variance components for multilocation data

Source	**df**	**MS**	**EMS**
Environment	e-1	Menv	
Replication (Environment)	e(r-1)	$M_{r/env}$	
Block (Replication*Environment)	er(b-1)	$M_{b/r/env}$	
Genotype	(g-1)	M_g	$\delta^2e + r\,\delta^2ge + er\,\delta^2g$
Genotype X Environment	(g-1)(e-1)	M_{gei}	$\delta^2e + r\delta^2ge$
Error	e(g-1)(r-1)	M_e	δ^2e

The significance of genotypes, locations and genotype X location/ environment was determined at the 0.05 and 0.01 probability levels using appropriate F-values (Fisher, 1925).

Mean separation was conducted using least significant difference (LSD) test at 0.05 probability level (Steel and Torrie, 1980).

3.7.2 Estimation of genetic parameters

The expected mean squares under the assumption of mixed effects model were computed from linear combinations of the mean squares. Phenotypic and genotypic variances were computed (Wricke and Weber, 1986) as follows;

$$\text{Genotypic variance: } \delta^2 g = \frac{Mg - Me}{r}$$

$$\text{Phenotypic variance: } \delta^2 p = \delta^2 e + \delta^2 g$$

Where;

Mg and Me are the mean sum of squares for the genotypes and error mean square from the analysis of variance.

3.7.3 Genotypic and Phenotypic coefficient of variation

The genotypic (GCV) and phenotypic (PCV) coefficient of variation were calculated for all quantitative traits, as a relative indicator of trait variability and which effect (Genetic or environmental) had a greater impact on expression of the trait, according to Singh and Chaudhary (2010), using the equations:

$$GCV\ (\%) = \frac{\sqrt{\delta^2 g}}{\bar{X}} * 100 \qquad \text{Equation 4}$$

$$PCV\ (\%) = \frac{\sqrt{\delta^2 p}}{\bar{X}} * 100 \qquad \text{Equation 5}$$

Where;

$\delta^2 g$ = genotypic variance, $\delta^2 p$ =phenotypic variance and $\bar{X}$= Grand mean.

3.7.4 Heritability

Broad sense heritability based on fixed genotypes across random locations was estimated as a percentage using variance components of ANOVA (Hallauer and Miranda, 1988):

$$H = \frac{\delta^2 g}{\delta^2 g + \frac{\delta^2 ge}{e} + \frac{\delta^2 e}{re}} * 100$$ Equation 6

Where; δ^2g = total genotypic variance, e = environment, r = replications, δ^2ge = genotype X location variance. The heritability was estimated on a mean entry basis.

3.8 Breeding Gain

Twenty-seven of the thirty advanced soybean germplasm lines coded G1-G27 were evaluated for genetic gains achieved between 1996, when the oldest of the 27 entries was constituted and 2007, the year when the latest entry was constituted. The five commercial cultivars, coded G24- G27 were used as benchmarks for calculation of genetic gain in these trials.

The realized genetic gains were determined by the following formula as used by Souza *et al.* (2009) :

$$GA(\%) = 100 * \frac{(Mean\ of\ best\ genotypes - Mean\ of\ commercial\ checks)}{Grand\ mean}$$ Equation 7

Where; GA (%) = Genetic advance or percentage genetic gain.

i. *Realized gain (RG2): genetic gains relative to mean of best commercial check*

$$RG2 = \left(\frac{\mathrm{MS} - \mathrm{MBC}}{\mathrm{MBC}}\right) * 100$$ Equation 8

ii. *Realized gains (RG3): genetic gains relative to mean of commercial checks*

$$RG3 = \left(\frac{\mathrm{MS} - \mathrm{MC}}{\mathrm{MC}}\right) * 100$$ Equation 9

3.9 Genotype X environment interaction and yield stability analysis

3.9.1 GGE biplot analysis

A GGE biplot analysis (Yan and Tinker, 2006) was performed on yield data using Breeding view, a component of the breeding management system (BMS, 2014) and GenStat statistical software . Multi-location data for the 30 genotypes was analysed for stability and yield across the four locations (GEI) using the GGE biplot model (Yan *et al.*, 2001; Yan and Rajcan, 2002) in Equation :

$$Y_{ijk} - \mu - \beta j = \sum_{l=1}^{k} \lambda l\, \xi il\, \eta jl + \varepsilon ij \qquad \text{Equation 10}$$

In this model:

Y_{ijk}; the mean yield response, of the i^{th} genotype in the j^{th} environment and k^{th} block, μ; the grand mean of the responses, βj; the environment effect, λl ξil ηjl are collectively called the principal component (PC), λ_l is the singular value of the i^{th} PC, ξil is the PC score, for genotype i and ηjl is the PC score for environment j, $\boldsymbol{\varepsilon}_{ij}$ is the residual associated with genotype i in the environment j. The bi-plots were generated by employing singular value decomposition (SVD) on multi-location trial data using a site regression model 2 ($SREG_2$) for yield and stability (de Oliveira *et al.*, 2005).

3.9.2 Additive main effects and multiplicative interaction model (AMMI)

Grain yield was analysed using the AMMI model that combines into a single model analysis of variance (ANOVA) for genotype and environment main effects with principal component

$$yij = \mu + gi + eij + \sum_{k=1}^{t} \lambda k\ \alpha ik\gamma jk + \varepsilon ij \qquad \text{Equation 11}$$

analysis (PCA) for the GEI. The AMMI model implemented is shown below (Crossa, 1990):

Where, yij =is the mean grain yield (ton ha−1) of the i'th genotype in the jth environment. μ is the overall mean, gi and eij are the main effects of the genotype and environment respectively, t is the number of PCA axes considered, λ k is the singular value of kth PCA axis, αik and jk are scores for the ith genotype and jth environment on the kth PCA axis, and εij is the residual term which includes experimental error.

3.9.3 Superiority measure (Cultivar superiority index)

The superiority measure (Pi) proposed by Lin and Binns (1988) was calculated in BMS on yield data using the formula:

$$P_i = \sum_{j=1}^{n} \frac{\left(X_{ij} - M_j\right)^2}{2n}$$ Equation 12

Where:

Xij = the ith genotype yield in the jth season, n= the number of locations, and Mj= mean yield of all check genotypes. Using this equation, the most consistently superior genotype was selected on the basis of having the lowest Pi value. That is, the smallest difference from the mean of checks Mj.

3.10 Correlation and path coefficient analysis for grain yield

Simple correlations and path coefficients was determined in combination, following the procedures of Singh and Chaudhary (2010). These were computed using the PathSAS (Cramer and Wehner, 2000) macros in SAS software, version 9.3 (SAS Institute, 2012), to show the direct and indirect effects of each secondary trait on grain yield, oil and protein content. PathSAS performed a correlation analysis to establish the degree of linear relationship between the independent variates. The independent variates were then regressed on yield to obtain direct effects in the form of path coefficients. Path coefficients when multiplied by the simple correlations determined the indirect effects of secondary traits on grain yield.

Conclusion

The trait of grain yield was evaluated across 16 sites, while all other agronomic data was collected separately for sites in Zimbabwe and separately for Zambia and Malawi. This was necessitated by the use of local checks, which were different for the three countries but similar in Zambia and Malawi. The results of yield evaluation are presented in the results section (Chapter 4).

4 RESULTS

4.1 Introduction

This chapter outlines the results of the study undertaken. It presents all the finding before and after analysis, in relation to each objective stated earlier in chapter 1.

4.2 Analysis of Variance (ANOVA)

There were significant (Fpr<.001; Fpr<.05) differences for traits of grain yield, plant height and pod height across sites in Zambia and Malawi in Table 7. The coefficients of variation were all below fourteen, an indication of the reliability of the data used in analysis of the trials. Two of the three traits measured across seven sites revealed presence of significant Genotype X Environment interaction. The mean square values for the environments also showed significant differences in all three traits.

Table 7 Mean squares for yield and agronomic traits across seven locations in Zambia and Malawi

Source of Variation	d.f.	Grain Yield	Plant Height	Pod Height
Env	6	42126997.70***	4416*	1276.66***
Env.rep	14	607669.30**	3096*	31.82
Env.rep.blk	105	408175.60**	1437.00	19.59
Genotype	29	1616053.80***	5903***	104.90
Genotype.Env	174	564290.60***	1189.00	12.69***
Residual	301	257432.00	1590.00	29.95
Mean		3464.19	73.23	14.29
% CV		14.65	7	12
SE		504.38	3.494	1.215
LSD (5%)		543.20	10.11	3.514

*** p<0.001; **p<0.05; *p<0.01

Analysis of variance for the nine sites in Zimbabwe (Table 8) exhibited highly significant genotypic and environment (Fpr<.001) differences for grain yield, across all sites. The genotype x environment differences were also seen to be significant for all traits measured. Coefficients of variation were relatively low for quantitative traits as yield, pod height, plant height and seed mass as well as qualitative traits such as crude oil and protein content.

Table 8 Mean squares for yield and agronomic traits across nine locations in Zimbabwe

Source of Variation	d.f.	Grain Yield	Crude Protein%	Crude Oil%	Pod Height	Plant Height	Seed Appearance	Seed Mass
Env	8	28464062***	663.03***	292.42***	734.44***	28400.78***	6.65***	1155.45***
Env.rep	18	392430***	4.45***	0.89***	13.49*	290.25***	0.84***	5.96
Env.rep.blk	135	207319***	2.97***	1.27***	14.83***	346.30***	0.48***	17.61***
Genotype	29	734602***	34.90***	19.99***	83.58***	4510.52***	1.40***	244***
Genotype.Env	232	217218***	3.16***	0.73***	14.27***	153.47***	0.60***	8.46***
Residual	387	121072	1.40	0.34	7.90	59.93	0.25	4.33
Mean		2859	45.86	22.62	16.33	92.84	2.59	22.64
%CV		12.17	2	2	14	8	19	8
SE		347.95	0.36	0.18	0.77	2.59	0.16	0.30
LSD (5%)		259	1.00	0.49	2.14	7.21	0.45	1.68

*** p<0.001; **p<0.05; *p<0.01

Analysis of variance for common traits across all sites in Zambia, Malawi and Zimbabwe revealed significant genotype and environment differences as seen in Table 9 below. Genotype X Environment interaction was also seen to be significant for grain yield and pod height. The combined coefficients of variation were high for all traits.

Table 9 ANOVA table for 28 genotypes across 16 locations in Zambia, Malawi and Zimbabwe

Source of Variation	d.f.	Seed Yield	Plant Height	Pod Height
Env	15	41415917***	24532.3	891.92**
Env.rep	32	506992	1521.7	20.41
Env.rep.blk	240	289790	834.9	15.4
Genotype	27	1580429***	8606.6**	156.23
Genotype.Env	405	387399**	703.3	13.85*
Residual	708	453978	894.1	19.4
Mean		3056	86	16
% CV		22.29	35.08	28.28
SE		673.8	29.9	4.405
LSD (5%)		270.4	10.11	3.514

*** p<0.001; **p<0.05; *p<0.01

4.3 **Mean Performance of soybean lines**

The mean performance of the 30 genotypes for which agronomic data was collected at 9 sites in Zimbabwe is shown in Table 10. Pod height ranged from a high of 19.1 cm to a low of 12.14 cm with G25 showing the highest pod height while G18 had the lowest. For plant height, the results show that G27 was the tallest standing at 119.6 cm followed closely by G5. G29 was the shortest at 68.83 cm. The 100 seed mass (SDMA) ranged from 29.06 to 17.08 g. Seed quality properties (oil and protein), were also evaluated and shown in Table 10. The highest oil content was observed in G18 with 24.99% while the lowest was in G16 with 20.7%. Crude protein content ranged from 48.17% to 43.81%. The grain yield ranged from 3197 kg ha^{-1} to 2441 kg ha^{-1}. The top yielding genotype in Zimbabwe was an experimental line G14 followed by a standard G15 and another experimental line G2.

Results from Zambia and Malawi sites were displayed in Table 11. Only three traits, plant height, pod height and yield were measured at all sites. Pod height ranged from a high of 18.51 (G27) to a low of 11.86 cm (G15). The genotypes exhibited a shorter stature at these sites on average when compared to the sites in Zimbabwe. This was shown by a lower plant height range of 95.14 cm (G5) to 52.44 cm (G22). Grain yield ranged from 3530 kg ha^{-1} to 2396 kg ha^{-1}. The top yielding genotype was a standard G10 while the lowest was an experimental line G16.

Table 10 Mean performance of genotypes for grain yield, quality and agronomic traits at nine locations in Zimbabwe

Entry #	Genotype	Pod Height (cm)	Plant Height (cm)	Seed Appearance	Grain Yield (Kg/ha)	Seed Mass (grams)	Crude Protein (%)	Crude Oil (%)
14	G14	17.94	101.6	2.492	3197	25.36	46.08	23.28
15	G15	13.37	82.41	2.259	3175	29.06	43.81	24.63
2	G2	16.39	78.37	2.14	3127	22.37	47.71	22.23
11	G11	17.6	94.97	2.073	3036	28.37	44.83	23.64
22	G22	13.2	69.51	2.998	3034	17.21	46.28	21.98
9	G9	16.29	96.75	2.749	3032	25.34	45.23	23.02
18	G18	12.14	80.96	2.522	2990	26.09	44.51	24.99
29	G29	13.44	68.83	2.673	2968	18.59	44.87	23.71
3	G3	18.18	81.32	2.14	2961	25.91	44.55	23.25
10	G10	18.75	117	2.736	2959	20.69	44.39	21.84
7	G7	16.23	94.76	3.017	2945	23.89	44.48	22.38
26	G26	16.12	96.95	2.708	2944	24.27	44.02	23.82
19	G19	16.12	87.96	2.236	2928	27.75	45.91	22.65
4	G4	17.67	82.37	2.713	2876	22.55	46.77	22.5
24	G24	17.16	113.3	2.682	2868	19.84	45.87	21.82
27	G27	17.97	119.6	2.952	2854	24.03	45.17	22.59
23	G23	18.07	98.63	2.701	2845	18.5	46.13	22.7
1	G1	16.68	89.1	2.372	2841	26.57	46.16	22.51
17	G17	13	70.36	2.524	2825	21.04	48.12	21.9
21	G21	18	94.22	2.454	2824	20.48	47.42	21.48
12	G12	15.2	93.62	2.813	2812	24.61	44.32	22.93
30	G30	18.67	104.4	2.634	2804	17.08	45.29	22.69
20	G20	13.57	75.68	2.82	2747	24.36	47.03	22.72

Entry #	Genotype	Pod Height (cm)	Plant Height (cm)	Seed Appearance	Grain Yield (Kg/ha)	Seed Mass (grams)	Crude Protein (%)	Crude Oil (%)
6	G6	15.25	94.46	2.644	2736	21.12	45.36	22.95
25	G25	19.1	104.2	2.39	2726	19.55	46.75	22.6
8	G8	18.58	108.1	2.737	2621	20.63	47.29	21.61
13	G13	16.9	88.21	2.654	2617	19.42	45.68	22.33
28	G28	15.33	81.95	2.651	2571	19.93	48.17	21.26
16	G16	16.18	96.59	2.586	2457	20.79	47.08	20.7
5	G5	16.9	119	2.67	2441	23.85	46.46	21.83
MEANS		16.33	92.84	2.591	2859	22.64	45.86	22.62
LSD (5%)		2.141	7.208	0.4481	259.6	1.681	1.006	0.4945
C.V.		14	8	19	10	8	2	2
Signific		***	***	***	***	***	***	***

*** p<0.001; **p<0.05; *p<0.01

Mean performances of 28 genotypes across the combination of sites used in the trial are presented in Table 12. Results arranged in descending order of yield performance show that genotype 2 with mean yield at 3,524 Kg/ha is the highest yielding across the 16 sites followed by Genotypes 15 and 10. It is relatively short in stature at 69.64 cm and pod height of 14.92. The lowest yielding across all sites was shown to be genotype 16. The genotype height ranged from 69.64 cm to 111.70 cm.

Table 11 Mean performance of genotypes for grain yield and agronomic traits at 7 locations in Zambia and Malawi

Entry #	Genotype	Pod Height (cm)	Plant Height (cm)	Grain yield (Kg/ha)
10	G10	18.07	91.86	3530
2	G2	12.31	59.22	3376
27	G27	18.51	91.44	3328
15	G15	10.86	70.18	3296
11	G11	13.17	72.75	3162
8	G8	17.25	89.56	3150
23	G23	14.52	71.57	3072
26	G26	12.26	79.79	3056
9	G9	13.39	75.74	3038
21	G21	17.69	71.03	3002
24	G24	13.1	84.43	2993
19	G19	14.09	66.39	2982
25	G25	17.91	78.74	2979
5	G5	15.85	95.14	2964
6	G6	11.75	62.86	2940
1	G1	15.84	76.72	2896
14	G14	15.44	85.22	2882
30	G30	16.09	86.9	2880
20	G20	11.86	58.18	2878
7	G7	14.47	71.96	2862
3	G3	18.2	86.94	2800
22	G22	12.2	52.44	2796
29	G29	11.55	73.48	2785
28	G28	12.39	65.29	2764
12	G12	12.57	71.97	2762
18	G18	13.4	58.33	2756
4	G4	14.34	64.26	2692
13	G13	12.85	65.58	2688
17	G17	11.84	52.96	2678
16	G16	15.09	66.12	2396
MEANS		14.29	73.23	2946
5% LSD		3.514	10.11	543.2
C.V.		12	7	9
SIGN		***[1]	***	NS

*** p<0.001; **p<0.05; *p<0.01; NS= Non-significant

Table 12 Mean performance of 28 genotypes across sixteen locations

Genotypes	Pod height (cm)	Plant height (cm)	Grain yield (Kg/ha)
2	14.92	69.64	3523.87
15	12.65	77.82	3490.32
10	18.94	108.49	3487.46
11	16.22	85.03	3348.17
9	15.74	89.27	3321.64
14	17.28	96.74	3304.64
27	18.47	109.35	3298.21
26	14.88	90.04	3226.16
21	18.35	111.70	3196.31
24	16.24	102.06	3164.55
23	16.46	84.63	3129.28
19	15.86	78.74	3125.91
1	16.85	84.25	3123.96
7	15.72	86.30	3116.17
25	18.79	92.42	3094.50
6	14.11	79.77	3084.18
18	13.45	72.34	3082.95
22	13.03	61.11	3081.22
8	18.51	100.43	3073.12
4	16.47	73.72	3043.73
20	13.65	67.83	3036.48
30	17.94	97.74	3024.20
17	12.89	61.78	2936.53
12	14.74	83.71	2923.91
5	17.25	110.81	2912.32
13	15.72	76.79	2873.15
28	14.57	74.76	2869.79
16	16.17	82.83	2608.41
Mean	16.02	86.00	3056.00

Summary statistics taken at each test location (Table 13) show that all the coefficients of variation were below 20%. Heritability values were moderate to high at all sites, with the exception of E4, E5, E9, E11 and E13. The data or results collected can therefore be accepted as replicable and useful.

Table 13 Summary statistics for grain yield at each test location

Environment	Mean	Min	Max	Range	Median	LSD	CV	Heritability
E1	3244.333	2193	4397	2204	3219.5	569.5897	10.7343	0.826634
E2	3003.988	2141	4279	2138	3007	693.4346	14.01203	0.584555
E3	1720.905	1185	2454	1269	1681.5	404.2692	14.19664	0.630192
E4	2596.738	1459	3895	2436	2592	685.6593	16.28941	0.376048
E5	3563.726	1400	5104	3704	3539	929.4819	16.09021	0.32151
E6	2813.929	1100	4750	3650	2775	841.4179	18.44692	0.841903
E7	2814.833	1723	3880	2157	2761	696.3915	15.21515	0.461532
E8	4081.429	3010	5560	2550	4050	548.4097	8.289305	0.860231
E9	3563.726	1400	5104	3704	3539	929.4819	16.09021	0.32151
E10	3242.393	2172	4115	1943	3265.5	527.3493	10.03363	0.543237
E11	4582.381	3210	5900	2690	4615	952.3732	12.82157	0.248465
E12	2674.286	1600	4920	3320	2640	840.401	19.3867	0.660793
E13	2907.762	2087	3891	1804	2836.5	623.6375	12.87137	0.39331
E14	3676.262	2728	4769	2041	3704	607.3517	10.1543	0.574566
E15	3052.738	1670	5090	3420	2945	841.593	16.5362	0.650855
E16	2456.071	1804	3287	1483	2418	448.3186	11.26084	0.572429

4.3.1 Estimation of genetic parameters from ANOVA

Estimates of genotypic variance, phenotypic variance, genotypic coefficient of variation (GVC) and Phenotypic Coefficient of Variation (PCV) as well as heritability are presented in Table 14. Heritability values for pod height and plant height were high (H^2>0.5) according to the classification by Robinson *et al.* (1949). However, Grain yield was only moderate (H^2=0.5). PCV values were found to be higher than GVC values for all traits.

Table 14 Estimates of genotypic parameters from ANOVA across 16 locations

Parameter	Mean	δ2g	δ2p	H2	GVC (%)	PVC (%)
Seed Yield	3056.00***	375483.67	829461.67	0.45	20.05	29.8
Plant Height	86.00***	2570.83	3464.93	0.74	58.96	68.45
Pod Height	16.00***	45.61	65.01	0.70	42.21	50.39
Crude Protein%*	45.86***	11.17	12.57	0.89	7.29	7.73
Crude Oil%*	22.62***	6.55	6.89	0.95	11.31	11.60
Seed Appearance*	2.59***	0.38	0.63	0.61	23.90	30.73
Seed Mass*	22.64***	79.89	84.22	0.95	39.48	40.54

δ^2g: Genotypic variance, δ^2p: Phenotypic variance, GCV: genotypic coefficient of variation, PCV: phenotypic coefficient of variation, *** p<0.001. *measured only at nine sites in Zimbabwe

4.4 Genetic gain

The mean yields of all the commercial varieties used in the trials across sixteen sites were presented in Table 15. The results showed that the commercial line G15 was the genotype to be exceeded in the breeding programme on the basis of yield alone. Subsequently, the mean yields of the five best performing soybean experimental lines across sixteen sites in the trials are shown in Table 16 below. G2 had the highest yield performance among experimental lines.

Table 15 Means for commercial checks used in genetic gain analysis across all locations

Genotype	Purple stain%	Pod height (cm)	Plant height(cm)	Grain yield (Kg/ha)
G15	4.97	12.65	77.82	3490.32
G10	1.04	18.94	108.49	3487.46
G11	0.36	16.22	85.03	3348.17
G7	10.15	15.72	86.30	3116.17
Mean	4.13	15.88	89.41	3360.53

Using the means of commercial lines (McLean and Byth, 1980) and selected lines (MSL) and means of the selected population (MP), genetic gain values generated were presented in Table 17. The genetic gain value comparing selected lines to commercial lines, GG1, for the trait of yield was negative. GG2, Genetic gains relative to the selected population mean was however positive. With the best commercial check being G15 as presented in Table 15, GG3; Genetic gains relative to best check was also found to be negative across all 16 sites.

Table 16 Means for best experimental lines

Genotypes	Purple stain%	Pod height (cm)	Plant height(cm)	Grain yield (Kg/ha)
G2	0.34	14.92	69.64	3523.87
G14	3.21	17.28	96.74	3304.64
G9	5.48	15.74	89.27	3321.64
G27	1.55	18.47	109.35	3298.21
G26	12.73	14.88	90.04	3226.16
Mean	4.66	16.26	91.01	3334.90

Table 17 Realised genetic gains

Trait	MCL	MSL	MP	GG1 %	GG2 %	GG3 %
Grain yield (Kg/ha)	3360.53	3334.90	3125.04	-0.76	6.72	-4.45
Purple stain %	4.13	4.66	2.85	12.83	63.42	-6.24
Pod height (cm)	15.88	16.26	15.92	15.60	2.11	28.54
Plant height (cm)	89.41	91.01	86.08	1.79	5.73	16.95

Realized gains GG1%: genetic gains relative to mean of commercial lines. GG2%: genetic gains relative to mean of Population. GG3%: genetic gains relative to mean of best commercial check.

Graphical presentation of yield regression over time is presented in Figure 2. Unfortunately, the R^2 value was too low to be adequately conclusive. The first cross constitution (1996) represented in this study as well as the last (2007) were both shown in the chart. It showed that since 1996, when the best performing genotype, now a commercial line, G15 was constituted, yields have fallen. However, a spike in the graph, slightly above G15 was seen in one genotype (G2) in 2007. From 1996 to 2007, soybean yields dropped at a linear rate of 14.1 Kg ha^{-1}. However, the data is not significant due to the small coefficient of determination (R^2).

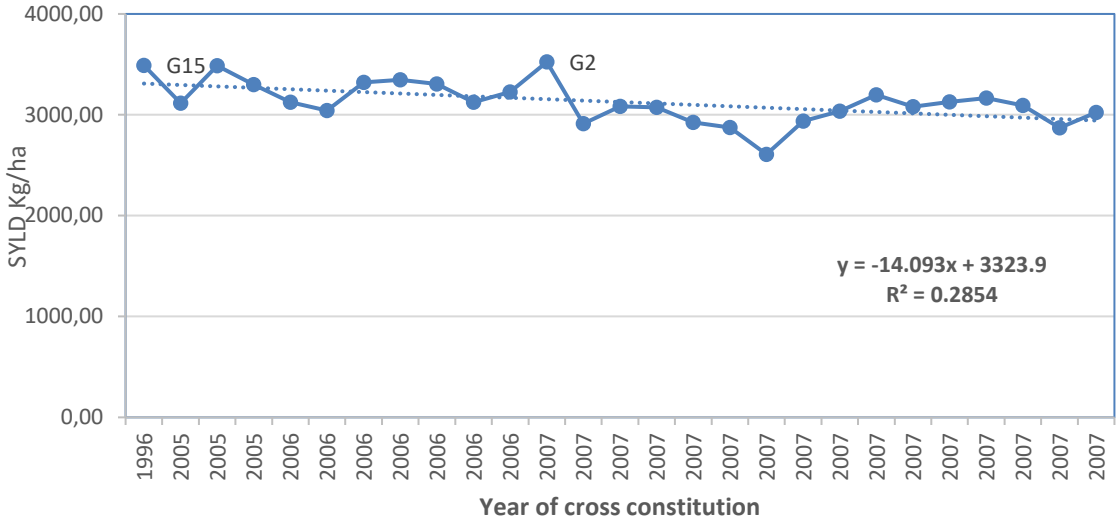

Figure 2 Regression of soybean yields against year of constitution across 16 locations

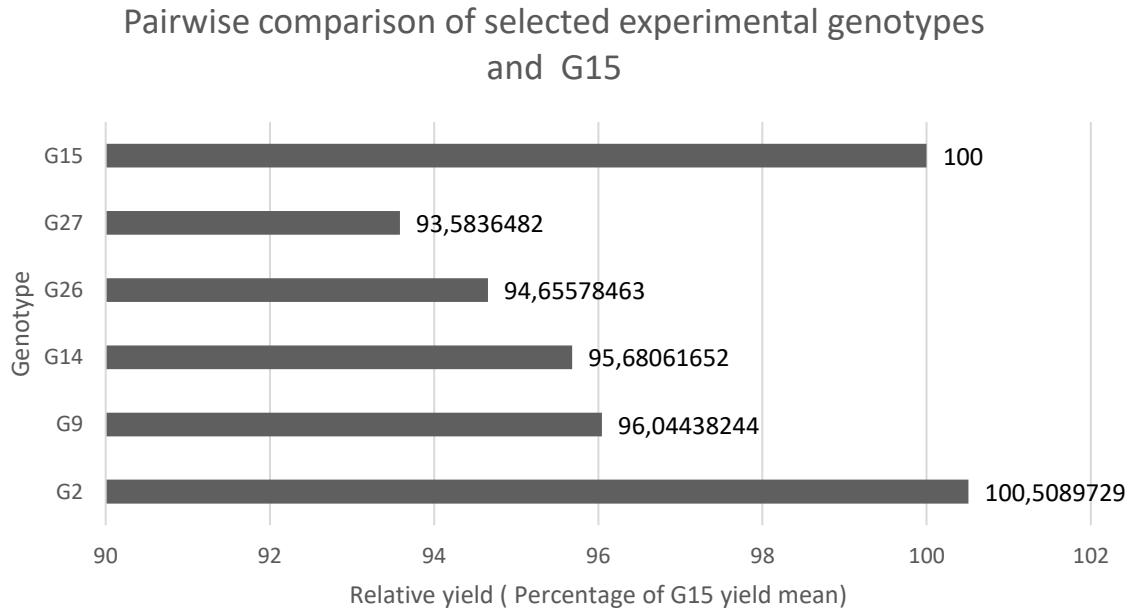

Figure 3 Pairwise comparison of five selected genotypes and the best yielding commercial check, G15 across all the 16 locations evaluated.

Results of pairwise analysis over all study locations between the benchmark commercial line, G15 and the five best performing experimental lines in Figure 3, revealed that, overall, only G2 was able to yield higher than G15 across all 16 sites. Giving a positive relative yield advantage of 1% over the benchmark (G15) at 100% relative yield.

The benchmark variety G15 compared to the best yielding genotype G2 across 9 environments in Zimbabwe is presented in Table 18. The results showed that G15 was outperformed at 5 sites. The yield advantage exhibited in Table 18 was not apparent in E1 to E3 and E8.

Table 18 Pairwise analysis of standard G15 and experimental line G2 across nine locations in Zimbabwe

	Environments								
Genotype	**E1**	**E2**	**E3**	**E4**	**E5**	**E6**	**E7**	**E8**	**E9**
G2	3760	2744	3457	2732	3186	3364	2220	3025	3682
G15	3988	3959	3483	2678	2820	3014	1861	3426	3344
Differential (Kg/ha)	-228	-1214	-26	54	365	350	359	-401	338
% Yield advantage	-6%	-31%	-1%	2%	13%	12%	19%	-12%	10%

This could be an indication that the sites in question were in a different mega environment where G2 is not adapted but the check G15 is. However, positive yield advantage of 2-19% was clearly observed in E4, to E7 and E9, showing specific adaptation of G2 to these enviroments. The switch in superiority between the two genotypes is further evidence of existing GXE on yield performance.

Further analysis displayed in Table 19 showed that the breeding programme had an annual genetic yield gain of only 2.80 kg ha^{-1} $year^{-1}$ translating into 0.08% (Almost negligible) increase in yield, over a breeding period of 12 years as shown in Table 19 below.

Table 19 Annual genetic gains of best experimental line (G2) compared to the best commercial variety over a period of 12 years

Trait	A Mean values for G15	B Mean for G2	C Differential (B-A)	Annual Genetic gain (Realised/Actual) C/12 years	Annual genetic Gain (%) [(C/A) * 100]/12years
CROIL (%)	24.63	22.23	2.4	0.2% $year^{-1}$	0.81
CRPRO (%)	43.81	47.71	(3.9)	(0.325%) $year^{-1}$	(0.74)
YIELD (kg ha^{-1})	3490.32	3523.87	33.55	2.80 $year^{-1}$	0.08

G15: Benchmark variety, G2: New experimental line, 12 years: Breeding period under evaluation

The other commercially important genotype (G3) in Zambia and Malawi was also depicted against the best five lines in Figure 4 below. All the selected lines showed higher yield performance. Genotypes G2 and G27 were the highest yielding across the 7 sites. Significant yield advantage was registered above the best check (G3). This advantage ranged from 1% to 21% across the seven sites.

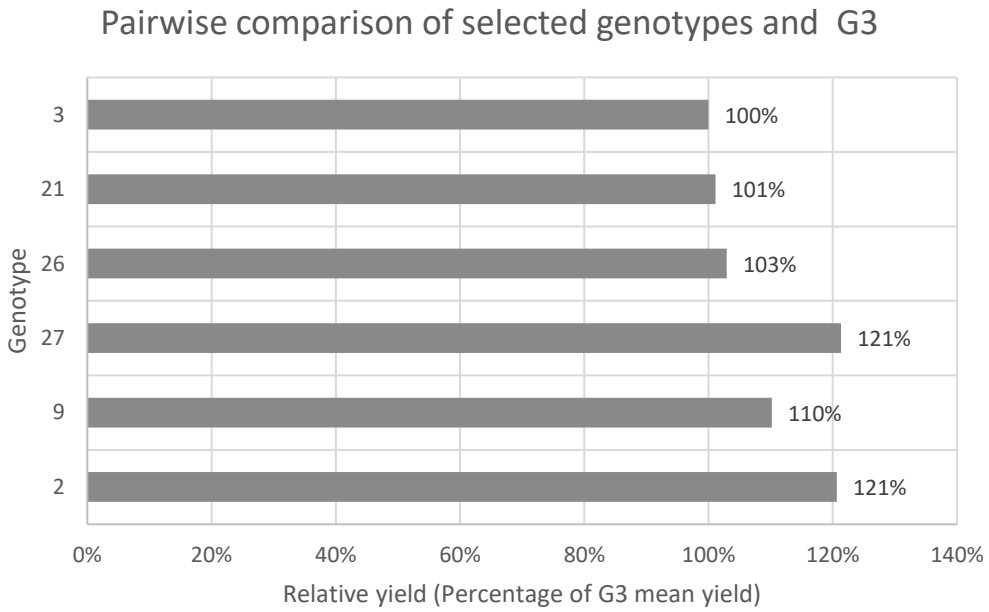

Figure 4 Pairwise comparison of best five selected genotypes against the commercial line G3 at 7 sites in Zambia and Malawi

A closer look at genotype performance in each of the seven sites in

Table 20, showed that the yield advantage was even higher between G2 and G3 in 6 of the 7 sites. This seems to indicate better adaptation of the experimental genotype G2 to all but one of the sites used in the trial. The switch in superiority was therefore, lower than that in the Zimbabwe sites leading to the possible conclusion that two mega environments may exist in the two countries (Zambia and Malawi and therefore, lower GXE interference.

Table 20 Pairwise analysis of standard G3 and the experimental line G2 across 7 locations in Zambia and Malawi

	Environments						
Genotype	**E10**	**E11**	**E12**	**E13**	**E14**	**E15**	**E16**
G2	5196.67	3276.087	2810.85	4693.33	3706.22	4199	4199
G3	3813.33	4108.73	2446.3	3863.33	2821.58	3070.67	3266
Differential (Kg/ha)	1383.34	-832.643	364.55	830	884.64	1128.33	933
% Yield advantage	36%	-20%	15%	21%	31%	37%	29%

4.5 Genotype and genotype X environment interaction

4.5.1 Additive main effects and multiplicative interactions

Table 21 AMMI-9 ANOVA for yield of the 28 soybean lines analysed across sixteen sites

Source	**df**	**SS**	**MS**	**% Interaction SS Explained**	**F probability**
Treatments	**447**	**806200732**	**1803581**		**0.00000**
Genotypes	27	54532013	2019704		0.00000
Environments	15	581035857	38735724		0.00000
Block	32	15949205	498413		0.00000
Interactions	**405**	**170632862**	**421316**	**100**	**0.00000**
IPCA	41	47784512	1165476	28.00	0.00000
IPCA	39	33066562	847861	19.38	0.00000
IPCA	37	19334777	522562	11.33	0.00000
IPCA	35	17397613	497075	10.20	0.00000
IPCA	33	11437102	346579	6.70	0.00196
IPCA	31	10971574	353922	6.43	0.00184
IPCA	29	7892515	272156	4.63	0.04867
IPCA	27	6293983	233110	3.69	0.16122
IPCA	25	5151132	206045	3.02	0.30646
Residuals	**108**	**11303091**	**104658**		**0.99982**
Error	**864**	**158317665**	**183238**		*
Total	**1343**	**980467601**	**730058**		*

The results of the AMMI analysis are shown in Table 21. According to Table 21, treatments accounted for 82.2% of the total grain yield sums of squares using approximately 33.3% of the total degrees of freedom. The genotypes alone, accounted for captured 6.7% of the treatment sums of squares while the environments explained up to 72% of the treatment sums of squares. The interactions explained 17.4% of the total sums of squares and 21% of the treatments sums of squares. This proves that the environments accounted for more variation followed by GXE interactions and finally, genotypes captured the least variation.

The AMMI analysis of variance showed significant effects of the genotypes, environments and the G x E Interaction. IPCA 1 showed significance (P≤0.001), however it accounted for only 28% of the interaction sum of squares. On addition of IPCA 2, the two IPCAs explained only 47.38% of the interaction sum of squares. IPCAs 3, 4, 5 and 6 were significant, when added to the model, IPCAs 1 to 6 accounted for 82% of the G x E interaction sum of squares. Therefore, AMMI-6 was used to describe the G x E interaction.

Table 22 AMMI ranking of first four best performing genotypes in each environment across the three countries in the subtropical medium altitude environments

Number	Environment	Mean	Score	1	2	3	4
3	E11	4582	0.25	G23	G9	G3	G8
15	E8	4081	-0.79	G2	G8	G5	G13
6	E14	3676	-11.15	G13	G14	G8	G21
12	E5	3564	6.56	G9	G2	G26	G19
16	E9	3564	6.56	G9	G2	G26	G19
1	E1	3244	-14.55	G13	G8	G14	G17
2	E10	3242	-6.92	G10	G26	G18	G16
7	E15	3053	27.08	G9	G2	G14	G7
9	E2	3004	-11	G2	G21	G8	G25
5	E13	2908	-12.59	G10	G16	G2	G13
14	E7	2815	-8.52	G13	G14	G22	G23
13	E6	2814	47.91	G8	G25	G10	G4
4	E12	2674	-6.61	G10	G1	G26	G14
11	E4	2597	-4.33	G2	G9	G6	G22
8	E16	2456	-4.11	G26	G14	G6	G2
10	E3	1721	-7.78	G2	G10	G20	G21

AMMI rankings of the best performing genotypes across sites are presented in Table 22. Sites E11, E8 and E14 were ranked as high yielding environments and E3, E16 and E4 as low yielding environments. When performance was ranked in the sites, Genotype G2 was found to be in the top four ranks at nine sites. At all except one of the nine sites (E16-low yielding site) G2 out-performed G15. On the other hand, G15 out-ranked G2 in low yielding environments, save for E14. G10 was also seen to rank well (first or second) in five sites. Overall, the inconsistencies in ranking of superior performance of genotypes among the 16 sites showed that GXE was at play in the outcomes observed.

4.5.2 GGE biplot analysis

Results of GGE biplot analysis showed that, compared to 46.45% of interaction variation explained by the AMMI-9 model, GGE-2 analysis model explained 49.46% of the sum of squares with two principal components, PC1= 31.67% and PC2= 17.79% of the GGE sum of squares.

Based on the polygon view (Figure 5), the test sites fell into five sectors and at least three mega environments. All the sites in the same sector, share the same winning genotypes that are also the vertex genotypes. The most significant mega-environment contained all but four test sites and was spread across all five sectors. The best performing genotype at sites E5, E6 and E15 was G10.

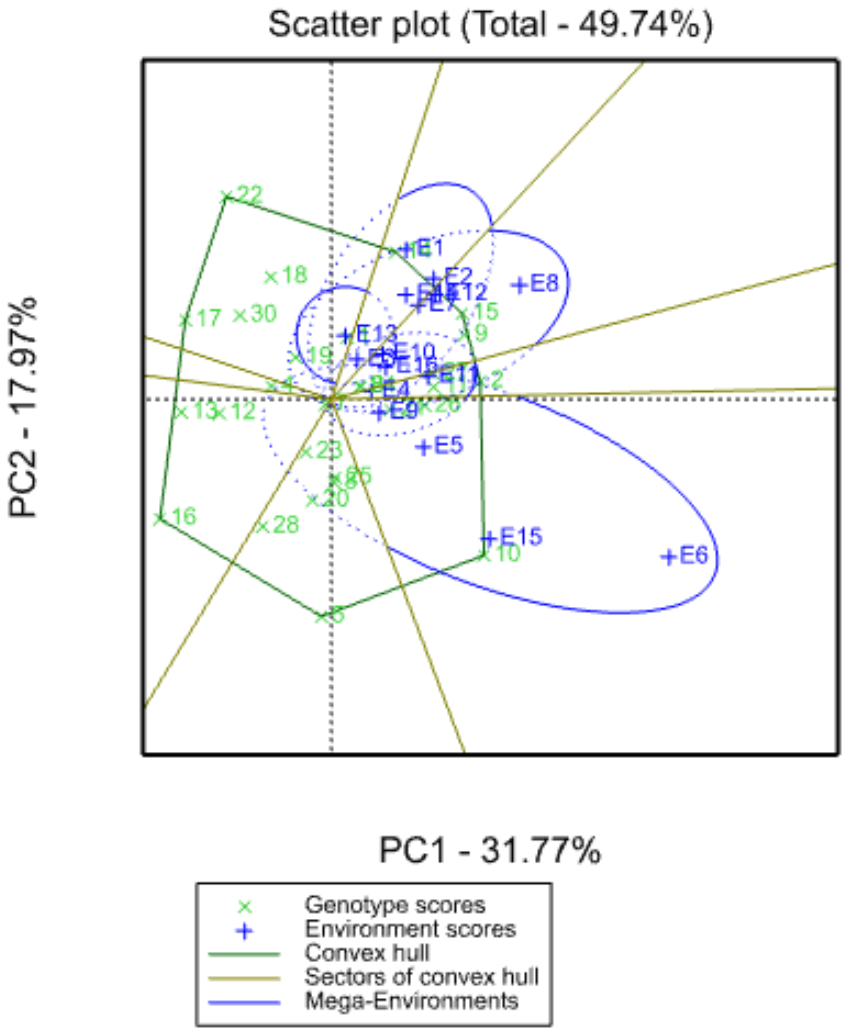

Figure 5 Which won where GGE Biplot of PC1 scores against PC2 scores for 30 genotypes

The best genotype at sites E4 and E11 was G2. Note that Genotype 2 was also the best performer for E9, E10 and E16 because markers of these sites were on G2's side of the perpendicular to the line that connects G 2's marker and that of G15. At sites E1, the best performer was G11. In the sector housing the largest cluster of test locations (E7, 12 and 2), which was also an intersection of the other two mega-environments the commercial check, G15 was the best performing genotype located at its vertex. A smaller mega environment containing genotypes G2, G9, G27 and G11, was situated between the two relevant mega-environments and exhibited general adaptation to most test locations.

The changes in superiority between sectors as well as the large number of sectors indicates a high degree of cross over genotype environment interaction.

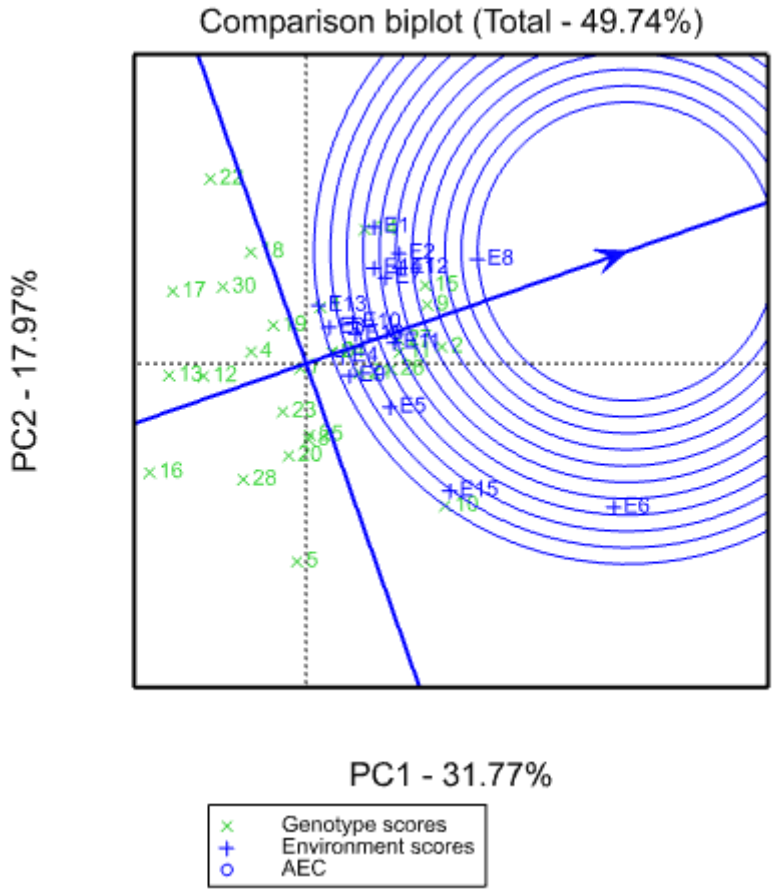

Figure 6 Ideal environment GGE biplot biplot of PC1 scores against PC2 scores for 16 environments

Most of the test locations were clustered together as seen in Figure 6. According to Yan *et al.* (2000), an ideal testing environment should have a large PC1 score (More discriminating of genotypes) and small PC2 score (More representative of test locations). Figure 6 shows that environment E4 had the smallest PC2 score, this made it the most representative location but small PC1 score as well, therefore, less discriminative. Most of the test locations were separated from E4 by an acute angle (<90^{o}) and therefore, very close to it. Location E6 was connected to E4 by an obtuse angle (≥ 90^{0}) and was identified as the most discriminating of test locations by virtue of having a high PC1 score. The ideal test location was identified as E11 because it had a good balance of discrimination and representativeness.

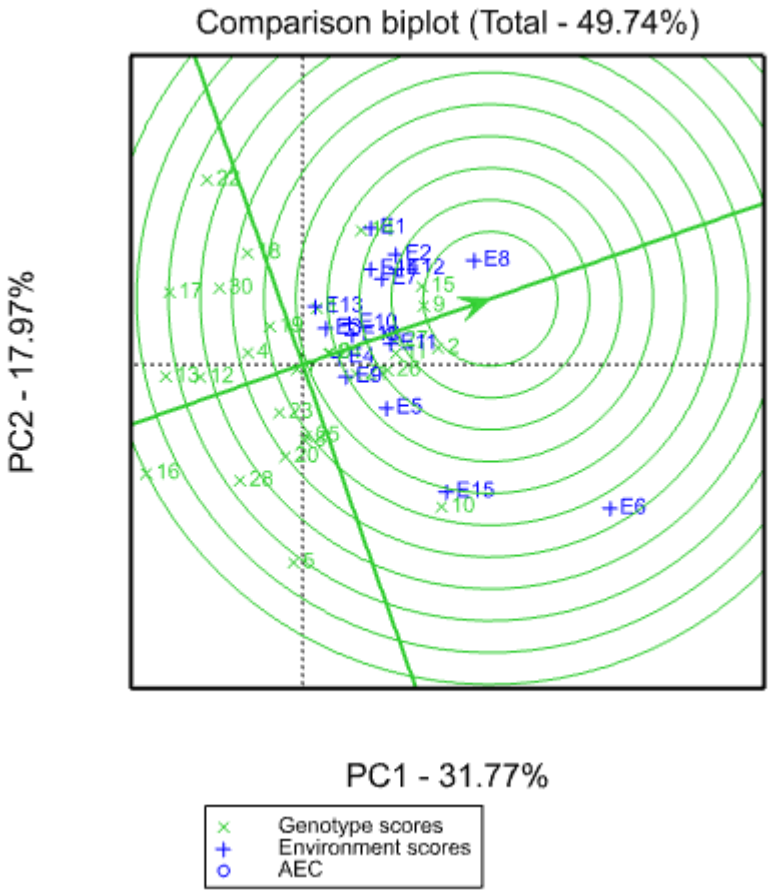

Figure 7 Ideal genotype GGE biplot of PC1 scores against PC2 scores for 16 environments

The ideal genotype is defined by Yan and Rajcan (2002) as having the highest mean performance and absolute stability. This ideal genotype is represented by the AEC that is marked by a blue circle with an arrow pointing from it In Figure 7. Such a genotype has a small PC2 score (stable) and a high PC1 score (high yielding). Concentric circles were drawn to help visualise the distance between each genotype and the ideal genotype; a genotype is more desirable if it is located closer to the ideal genotype,

According to Figure 7, G2, G10, G15 and G9 all had high PC1 score, therefore high yielding. However, G2 out-yielded them all. In terms of stability, G2 had the lowest PC2 score and was therefore, the most stable. G2 was identified by the biplot as the ideal genotype, followed by G9 and the commercial check G15.

4.5.3 Cultivar Superiority Measure

The cultivar superiority index ranged from 134,512 to 999,964 as shown in Table 23. The lower value representing the best combination of stability and productivity in terms of yield was

associated with the commercial check, entry G15, while the largest index value was associated with entry G16.

Table 23 Cultivar superiority indices and associated mean yields of all 30 entries across 16 locations

Ranking	Genotypes	Cultivar superiority index	Means (Kg/ha)
1	15	134512	3490
2	2	173083	3524
3	10	214317	3487
4	11	297492	3348
5	27	314541	3298
6	9	348268	3322
7	14	353788	3305
8	26	357835	3226
9	21	389836	3196
10	8	400017	3073
11	25	410692	3094
12	24	414798	3165
13	23	431112	3129
14	1	435772	3124
15	19	440870	3126
16	7	459852	3116
17	20	490601	3036
18	6	496481	3084
19	18	503705	3083
20	4	549894	3044
21	30	556508	3024
22	5	633217	2912
23	28	635202	2870
24	22	650340	3081
25	12	676821	2924
26	17	706943	2937
27	13	745578	2873
28	16	999964	2608

Associated mean yields across the 16 environments showed a similar pattern; the lower superiority index was associated with the largest mean yield and increasing stability across sites while higher indices were associated with lower mean yields and decreasing stability.

4.6 Frequency distribution of secondary traits

The frequency distribution of secondary traits in the test environments is shown in Figure 8 and Figure 9. When plotted, the data showed normal distribution and positive skewness for number of nodes per plant, pods per plant, and seeds per plant. Approximately 25% of the genotypes produced 25 nodes per plant and 50 pods per plant. However, number of seeds per pod had discontinuous distribution with 80% of the genotypes having two seed per pod. Yield was shown to have outliers with 70% of the genotypes producing 20 g of seed per plant. Harvest index, had outliers with 60% having a score of 0.6. For the number of branches per plant, the distribution showed an average of three branches per plant with a positive skewness. Biomass also showed a positive skewness with a number of outliers.

Table 24 shows the descriptive statistics, further emphasizes the positive skewness of all the secondary traits analyzed in relation to yield. Heritability values ranged from 0.0 to 0.82 (0% to 82%). According to Robinson *et al.* (1949), all the traits displayed high heritability (>50%) with the exception of Yield, biomass and harvest index. These three traits all showed low heritability (<50%). None of the analyzed traits showed moderate heritability (=50%).

Table 24 Discriptive statistics for yield and secondary traits for 30 genotypes at one site.

Trait	Mean	Min	Max	%cv	Skewness	Heritability
Number of Nodes	23.99	14.00	41.00	22.61	0.52	0.62
Number of Pods	52.47	30.00	77.00	19.53	0.18	0.64
Seeds per plant	110.56	74.00	172.00	20.99	0.59	0.69
Seeds per pod	2.07	1.00	3.00	15.89	1.29	0.62
Yield per plant (grams)	22.91	16.42	78.50	29.97	6.00	0.00
Biomass	48.34	29.45	110.95	23.51	2.140	0.33
Number of Branches	3.91	1.00	7.00	34.84	0.48	0.82
Harvest Index	0.49	0.30	1.60	27.59	6.41	0.12

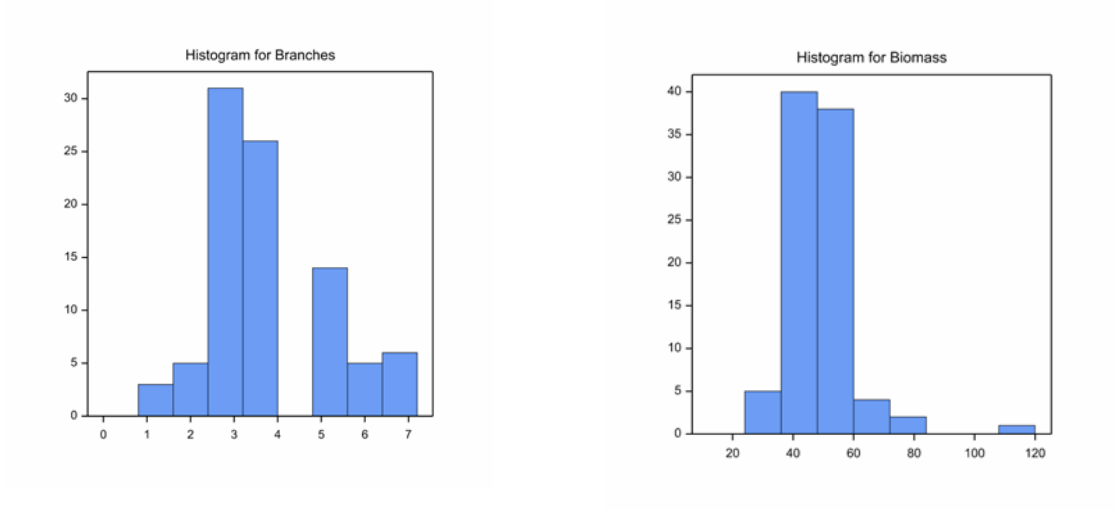

Figure 8 Histogram for number of branches per plant and biomass per plant for 30 genotypes planted at one location in Harare, Zimbabwe.

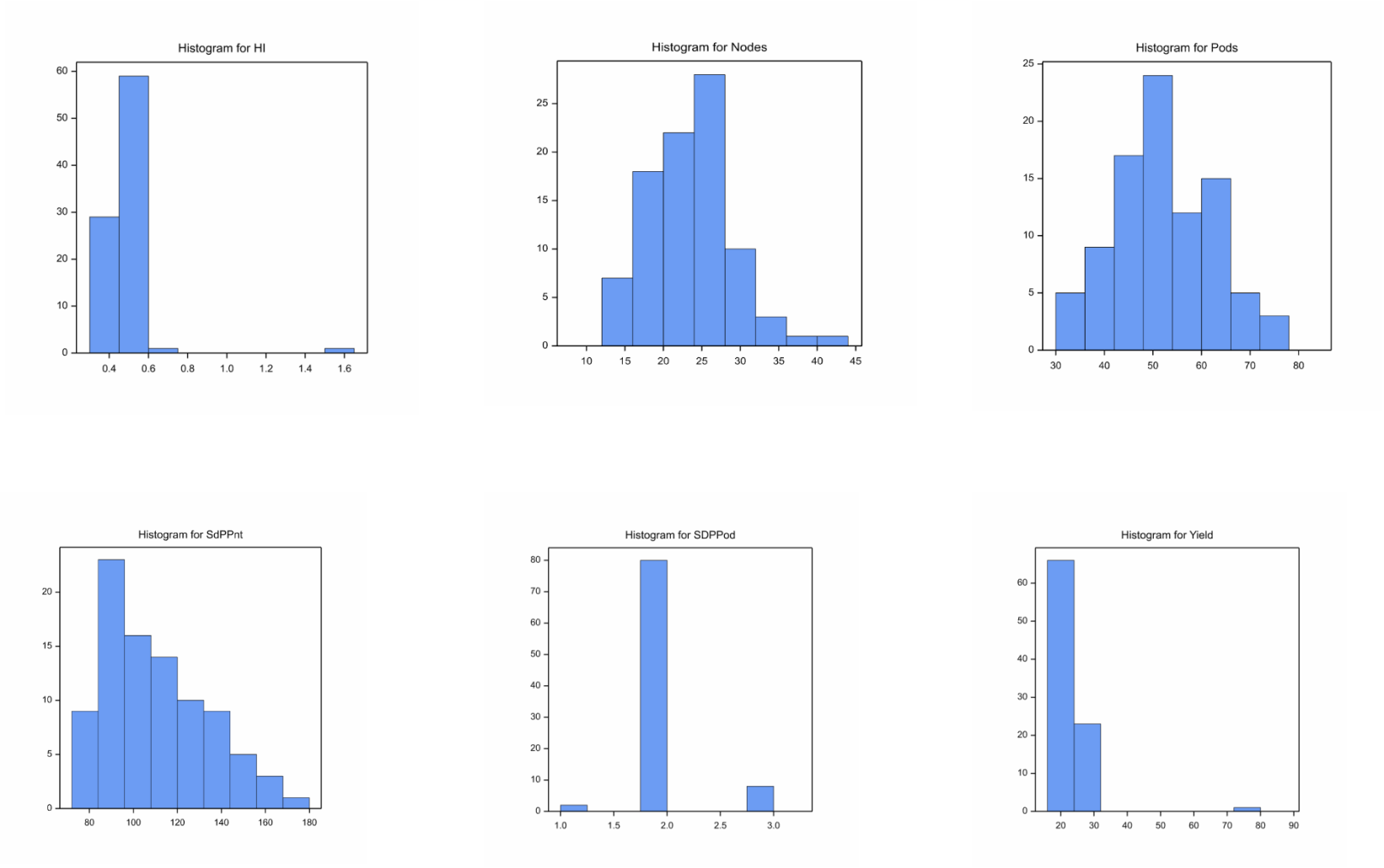

Figure 9 From top left to right: Histograms for harvest index (HI), nodes per plant and pods per plant. From bottom left to right: Seeds per plant, seeds per pod and yield per plant for 30 genotypes planted at one location in Harare, Zimbabwe.

4.7 Phenotypic correlation of grain yield and its secondary traits

Table 25 displays the mean performances of each genotype for the secondary traits analysed. The coefficient of determination (R^2) derived was highly significant at 0.93. Highly significant variations ($p<0.001$) were shown among the tested genotypes for traits of number of nodes per plant, number of branches per plant, number of pods per plant, number of seeds per plant and seeds per pod. Non-significant differences were registered for grain yield, biomass and harvest index.

Results of correlation analysis presented in Table 26 showed positive correlation between number of nodes per plant and all the secondary traits except harvest index and number of seeds per pod. Biomass and number of branches per plant showed positive correlation to all traits except harvest index, biomass and number of branches per plant. Another notably high correlation was seen between number of nodes per plant and number of pods per plant. Number of seed per plant was strongly and positively correlated to number of nodes per plant. Grain yield displayed positive correlation to all traits analysed except number of seed per pod. However, the correlation strength was weak between yield and all traits except harvest index.

Table 25 Mean performance of genotypes in respect of secondary traits analysed for correlation and path analysis

Genotypes	Nodes	Pods	Seed/ Plant	Seed/ Pod	Yield	Biomass	Branches	Harvest Index
G1	15.19	31.61	83.33	2.82	19.67	40.19	1.33	0.50
G2	15.73	46.05	95.67	2.05	19.55	42.40	4.33	0.47
G3	19.22	36.29	84.00	2.31	19.82	40.99	3.00	0.50
G4	19.94	45.99	117.67	2.62	23.48	44.84	3.33	0.50
G5	23.56	52.29	96.33	1.98	22.03	52.55	5.33	0.47
G6	24.30	51.97	110.33	2.03	19.43	41.57	3.00	0.47
G7	24.60	49.98	109.00	2.03	22.55	46.28	4.00	0.50
G8	32.80	63.70	116.00	2.04	26.75	62.41	5.33	0.43
G9	20.71	49.69	102.67	2.02	23.43	45.51	2.67	0.50
G10	29.07	61.69	147.00	1.97	25.90	53.31	3.00	0.50
G11	22.92	56.28	100.00	1.97	25.59	54.27	3.33	0.50
G12	29.03	63.68	119.33	2.04	27.44	62.37	5.33	0.43
G13	26.35	63.67	115.33	2.01	20.53	46.82	5.67	0.40
G14	22.76	49.39	97.00	2.01	21.94	45.70	3.33	0.50
G15	25.60	50.32	86.67	1.62	23.37	46.93	4.00	0.50
G16	26.02	58.66	121.00	2.05	22.89	55.43	4.33	0.40
G17	26.57	47.29	93.00	1.96	18.74	33.56	5.67	0.57
G18	28.99	49.34	99.00	2.05	21.66	67.01	4.67	0.43
G19	19.30	47.01	86.00	2.03	23.28	48.07	2.00	0.50
G20	18.58	45.60	98.33	2.05	19.55	40.36	2.67	0.50
G21	27.13	64.70	148.00	2.04	24.71	54.79	3.67	0.47
G22	24.64	52.98	105.67	1.62	38.71	42.18	5.67	0.83
G23	23.55	53.94	132.67	2.05	21.18	50.24	3.33	0.40
G24	24.81	48.39	111.00	2.37	20.32	44.83	4.00	0.47
G25	22.38	53.02	121.67	2.02	21.77	51.78	3.33	0.43
G26	20.64	53.32	105.00	1.96	22.16	44.58	3.33	0.53
G27	27.90	47.96	107.67	1.96	24.39	46.91	4.00	0.53
G28	26.40	59.69	128.67	1.97	22.33	52.61	6.33	0.40
G29	22.48	58.07	135.33	2.30	22.74	40.41	4.00	0.57
G30	28.51	61.43	143.33	2.04	21.31	51.30	3.33	0.40
Mean	23.99	52.47	110.56	2.07	22.91	48.34	3.91	0.49
LSD (5%	7.18	13.33	28.92	0.41	11.48	17.24	1.40	0.21
Heritability	0.62	0.64	0.69	0.62	0.00	0.33	0.82	0.12
p-value	***	***	***	***	NS	NS	***	NS

*** p<0.001, significant at 1% probability. NS= Non significant

Table 26 Correlation coefficients between grain yield and secondary yield components analysed

	Nodes per plant	Number of pods per plant	Number of seeds per plant	Number of seeds per pod	Yield per plant	Biomass	Branches	Harvest Index
Nodes per plant	-							
Number of pods per plant	0.7519***	-						
Number of seeds per plant	0.60575***	0.79826***	-					
Number of seeds per pod	-0.2077	-0.31645	0.0688	-				
Yield (grams)	0.29154**	0.39252**	0.16912	-0.29615**	-			
Biomass	0.61432***	0.56312***	0.41284***	-0.08027	0.25344*	-		
Number of Branches	0.6139***	0.54361***	0.30927**	0.26275*	0.21562*	0.3105**	-	
Harvest Index	-0.17455	-0.10812	-0.24862	-0.23445	0.76317***	-0.30662	-0.04954	-

*p≤0.05, ** p≤0.01, ***p≤0.001

4.8 Regression Analysis

Results of regression analysis are shown in Table 27 below. Results showed that biomass and harvest index had a highly significant (P<0.001) effect on yield. However, the coefficients of determination values were very low (<20%) for all but harvest index. This indicates that trends observed are weak.

Table 27 Regression of secondary traits on yield at Ratray Arnold Research and Stapleford Research stations

Trait	F pr	R^2 (%)	Regression coefficient	SE	Pr(t)
Nodes per plant	0.6042	8.50	0.04	0.079	0.6042
Number of pods per plant	0.005**	15.41	0.18	0.063	0.0050**
Number of seed per plant	0.8539	2.86	0.00	0.022	0.8539
Number of seeds per pod	0.4236	8.77	0.77	0.958	0.4236
Biomass	<.0001***	6.42	0.22	0.030	<.0001***
Branches	0.778	4.65	-0.07	0.239	0.7780
Harvest Index	<.0001***	58.24	47.13	1.997	<.0001***

*p≤0.05, ** p≤0.01, ***p≤0.001

4.9 Path coefficient analysis

Results of the path analysis are displayed in Table 28. Among all yield components measured, harvest index per plant had the only strong direct and positive effect on plant yield; correlations between other yield components and yield per plant were weak. The direct effects of the number of nodes per branch, nodes per plant, seeds per pod and biomass per plant on the grain yield were weak. With respect to grain yield per plant, the indirect effects of yield components on each other were weak for all yield components. The observed weak indirect correlations were a result of weak correlations among yield components. Therefore, selection for an increased harvest index per plant could be instrumental in improving plant yield, whereas selection for the other yield component traits may not have a positive effect on fruit yield.

Table 28 Direct, indirect and total correlation or contribution of secondary traits on yield

Trait	Nodes	Seed/ plant	Seed/pod	Biomass	Branches	Harvest index	Total to Yield	Pods
Nodes	0.06	0.13	0.02	0.25	-0.03	-0.14	0.3	0.12
Seed per plant	0.04	0.22	-0.04	0.17	-0.01	-0.2	0.17	0.07
Seed per pod	-0.01	0.08	-0.11	-0.08	0.02	-0.2	-0.3	-0.12
Biomass	0.04	0.09	0.02	0.41	-0.02	-0.29	0.25	0.1
Branches	0.04	0.07	0.04	0.14	-0.04	-0.03	0.2	0.08
Harvest index	-0.01	-0.05	0.02	-0.12	0	0.95	0.79	0.31

4.10 Conclusion

The results of the study undertaken, showed approximately 25% of the genotypes produced 25 nodes per plant, 50 pods per plant, 80% of the genotypes having 2 seed per pod and yielding 20g of seed per plant. All the traits analyzed displayed high heritability (>50%) with the exception of Yield, biomass and harvest index which instead showed low heritability (<50%). Correlation analysis showed a weak association between yield and all traits except harvest index. Path coefficient analysis showed that indirect selection for increased harvest index would have a positive effect on grain yield.

Genetic gain results showed that the commercial line G15 was the genotype to be exceeded in the breeding programme based on yield alone. Since the constitution of G15 in 1996, yields have fallen at a linear rate of 14.1 Kg ha^{-1}. A genetic gain of 10 -21% was observed over the best check in Malawi and Zambia. However, no significant gain was observed in Zimbabwe, overall. Looking at specific locations, four showed the best experimental G2 as being below the best check, G15. In the other five locations in the same country, the G2 performed 2-19% better than the best check.

The switch in superiority between G2 and G15 from one location to another is characteristic of crossover interaction and requires the implementation of a specific adaptation strategy. According to GGE analysis, location E11 was identified as the ideal test environment while genotype G2 was the ideal genotype, yielding highest and being most stable. However, Cultivar superiority index identified G15 as the most productive and stable.

An overall analysis across the 16 sites resulted in the disappearance of all observed genetic gains. The complexity of G x E was validated by GGE biplot analysis, which separated the test sites into five separate sectors and AMMI showed up to seven significant IPCAs. The observed trends are discussed at length in the next chapter.

5 GENERAL DISCUSSION

5.1 Introduction

This chapter critically discusses and interprets the results and trends observed from the study with reference to comparative studies and existing literature. It is from this chapter that certain outcomes will be explained before general conclusions and recommendations can be made.

5.2 Mean performance

Significant genotype mean squares ($P \leq 0.001$) were observed for all the evaluated traits indicating the presence of adequate variability among the studied genotypes. From the 30 genotypes evaluated in this study, seven top yielding genotypes were identified. These included two commercial lines, G15 and G10 as well as five experimental lines, G2, G9, G27, G26 and G14, which are at the advanced stage of breeding. The LSD of 259kg at 5% did not show significant differences among these genotypes for grain yield. G15 was found to be the top yielder across all 16 sites, despite having been bred much earlier, 1996. Based on its superior performance, this genotype can be recommended for use as a parent in future crosses to allow transfer and preservation of favourable genes. However, this is a genotype that is not widely grown in the locations or countries used in the study. G15 has wide adaptation in Zimbabwe as exhibited by extremely high yields at the Zimbabwe sites, which is why it is currently only released in this country. The experimental line, G2, was ranked highest for grain yield at 12 of the 16 sites, coming second only to G15. It is therefore a good candidate for genetic advancement and performed well at nearly all sites. In addition to being high yielding, G2 also exhibited one of the lowest plant heights, standing at an average of 70.23 cm, while all shorter genotypes as G17, G20 and G29 were much lower yielding. With this height, yields can further be increased by altering the plant density to increase plant population.

Important seed quality attributes, as crude oil and crude protein percentage when analysed were high in the experimental line G2 (22.23% and 47.71%, respectively). However, the standard genotype G15 did have higher oil content but lower protein value (24.63% and 43.1%, respectively). The fact that the breeding programme has developed such a line is an indication of genetic improvement.

5.3 Analysis of Variance

Significant differences were detected amongst genotypes for grain yield and secondary traits, across sites, an indication that the environment had an influence on the expression of each trait so that individual genotypes did not necessarily perform the same at each site. Therefore, it was prudent to go ahead and analyse the interaction of the genotypes and their respective environments particularly for the traits of yield and seed content (oil and protein percentage). The differences in yields obtained where not very wide as the yield ranged between 3526 kg/ ha and 2642 kg/ ha. This could be attributed to a low diversity for the trait of yield emanating from having a narrow genetic base from which crosses are developed. A low diversity leaves very narrow opportunity for genetic improvement of the trait in question. The narrow range was also observed for crude oil and crude protein percentage measured at nine sites in Zimbabwe were differences as low as 0.5% were observed. On a positive note, the significant genotype mean squares seen at all sites showed that the test locations were able to adequately discriminate between genotypes.

5.4 Heritability, Genetic and Phenotypic coefficient of variation

The overall heritability estimate for the trait of yield in this study was found to be moderate at 50%. This classification is according to that of Robinson *et al.* (1949). Where high heritability is >50%, moderate= 50% and low heritability is <50%. This finding was in agreement with an earlier study by Mushoriwa (2013) on some of the key test locations, the study found grain yield heritability to be moderate (49.88%), leading up to 50% as in this study. The estimates at individual sites ranged from high to low, with the highest being at E1 and E8. Both of these sites are research stations where as expected, management of trials and data collection was close to optimal. The differences in heritability values can therefore be explained by the differences in test locations and the efficiency of trial management.

Some sites were severely affected by the poor rainfall pattern of the 2015/2016 season that led to moisture stress and in some cases, poor plant stands so that the full genetic potential of the material could not be expressed. This was true for nearly all the sites used in Malawi that exhibited low heritability values. According to Sleper and Poehlman (2006), a higher heritability increases the effectiveness of selection as it signifies a lower environmental variation. Heritability estimates for the secondary traits however were extremely high. Crude oil and seed mass for example were as high as 95%; this suggests that high genetic variation more than environmental variation was responsible for the outcome.

Some studies have shown similar findings where high GXE interaction lead to lower heritability values. Karasu *et al.* (2009), in their study of the heritability of soybean yield reported low to moderate heritability. They concluded that seed yield being a quantitative trait with a complex character controlled by many genes, had a larger environmental influence.

According to the classification system of Shivasubramanian and Menon (1973), GCV and PCV values are classified as low (0 to 10%), moderate (10 to 20%) and high (>20%). High PCV and GVC observed for traits of grain yield, plant height, pod height, seed appearance and seed mass indicates adequate variability to be exploited in selection. Crude oil had moderate values while crude protein was low. Higher PVC than GVC values for all traits under analysis indicates the importance of environmental influence on trait expression as indicated in the ANOVA findings of significant GXE Interaction. However, the narrow differences between the two parameters show low environmental variation. In combination with heritability values, yield was clearly the most influenced by the environment with the lowest heritability value and larger difference between its PCV and GCV values.

5.5 Genetic gains

Though grain yield exhibited moderate heritability, realized genetic yield gains of the five selected genotypes were found to be positive across all sites over the population mean. This shows that selection was generally successful ln improving grain yield. However, the genetic gains over commercial varieties and over the best commercial variety (G15) were negative assuming that selection did not significantly improve grain yield above what was already obtaining on the market. Further breeding is therefore required to bring these negative findings into positives.

Results from Zimbabwe test locations showed G2 to have a yield advantage of up to 19% over G15 at five of the eight locations. In Zambia and Malawi, gains ranging from 1- 21% were registered at six of the total seven locations evaluated. G2 therefore displayed specific adaptation to these six sites that included both high and low potential sites for the trait of yield. Overall, the study revealed that over a period of 12 years of breeding and selection, the breeding programme showed an annual genetic yield gain of only 2.8 kg $ha^{-1}year^{-1}$ translating into 0.08% annual rate of increase in yield. The results also showed a yield advantage over the earliest line (G15) of 1%.

Using the same method of analysis, Lange and Federizzi (2009), registered positive soybean yield genetic gains ranging between 1.01 and 1.27%. However, they also registered no genetic gains for early maturity groups. It was thought that the lack of genetic progress in yield was

due to a shift in breeding objectives over the period concerned (Lange and Federizzi, 2009). The breeding programme under evaluation in the recent past, prioritized breeding for Soybean rust resistance (Levy, 2005), at the expense of increased yield. This may have led to a fall in yield potential since the constitution of G15 in 1996, while having developed rust tolerant lines. Breeding efforts to combine high grain yield with resistance to soybean rust resulted in a linkage drag, which led to lower yield potential. Genotypes constituted post 1996 against a rust tolerance background showed lower yields relative to the non-rust tolerant ones.

Another explanation for a low or lack of genetic progress offered by some scholars is that the selection environment does not necessarily equally favor discrimination between genotypes of high and low yield potential. The poor rainfall pattern experienced at some test sites in the 2015- 2016 season would have increased the GEI. The crossing strategies used to create variability in the lines may have also contributed to such an outcome by not being sufficient for the objective of increasing yield, as they have not changed with the change in objectives from breeding for rust resistance back to breeding for yield against a rust tolerant background.

As expected, Crude protein being negatively correlated to grain yield showed negative genetic gains; therefore, it was reducing as grain yield was increasing. This has been the case in many other studies in which grain yield and seed composition traits were evaluated (Gurmu *et al.*, 2009; Karasu *et al.*, 2009; Mushoriwa, 2013; Ngalamu *et al.*, 2013)

The 1% yield advantage registered by G2 over thee best commercial line, G15, in this study shows that it is still possible to develop competitive lines for the desired market.

5.6 Variation of genotypes for grain yield and secondary traits

A separate analysis of yield and its secondary component traits revealed highly significant differences among the evaluated genotypes for mean values of secondary traits analysed except grain yield, harvest index and biomass. This turnout suggests a low level of genetic diversity among the genotypes for these three traits. The test genotypes showed superior mean values for all other yield components, indicating that the traits were improving progressively over time. This meant that the modern genotypes have better performance in these traits compared to the earlier genotypes, which were constituted before 2007. The superior traits may have been used for selection but did not significantly increase the yield potential in the experimental lines.

5.7 Correlation analysis

According to correlation analysis, number of branches per plant, biomass, number of seeds per plant, harvest index, number of nodes per plant and number of pods per plant were positively correlated, thus exhibiting a degree of association among them. In particular, number of nodes per plant and number of pods per plant showed positive, strong and significant correlation to all traits but number of seed per pod and harvest index for which they displayed negative correlation. Associations between characters are important because they determine which traits can be improved simultaneously. In this study, the traits under evaluation with the exception of harvest index and seeds per pod may have been selected for and improved together in the breeding programme.

All the traits analysed were positively correlated to grain yield except number of seed per pod. This would signify a proportional increase in yield for each increase of the secondary traits analysed. However, the fact that their correlation was weak for all except harvest index meant that, higher mean values for these traits might not necessarily increase the grain yield. In addition, their negative correlation to harvest index meant that their improvement might produce a decrease in the harvest index. Since harvest index showed a strong positive and significant correlation to yield, increasing harvest index is expected to cause a proportional increase in mean grain yield as well.

5.8 Regression of traits on grain yield

Regression analysis determined that correlation between yield and harvest index was the most significant. The high coefficients of determination and highest level of significance from harvest index exhibits its importance in improving yield. This finding is in agreement with the strong and positive association established from correlation analysis. Secondary traits with low coefficients of determination (<20%), had negligible direct contribution to grain yield, even though significant associations with yield were detected in correlation analysis. This means that those traits had less direct influence on yield; however, they cannot be ignored, because their cumulative contribution to yield could have incremental effects. As a result, number of pods per plant and biomass, having significant regression on yield should be included in the selection index for grain yield.

5.9 Path Coefficient Analysis

The direct effects of number of pods per plant, number of branches per plant, biomass, and number of seeds per plant, nodes per plant and harvest index were positive while the

remaining characters exhibited negative direct effects. The highest direct effect was exhibited by harvest index and it was followed by biomass; hence, this trait may be given more emphasis in indirect selection for high yielding soybean lines.

Harvest index being a ratio of weight of seed per plant and above ground dry weight is a measure of a plant's efficiency at converting photosynthates into yield. Therefore, direct selection for harvest index might result in increased yield gains than direct selection for grain yield itself, particularly since harvest index exhibited higher heritability than grain yield.

The trait of pod number per plant is recognized as an important factor affecting yield based on the results of this study. It would therefore, be expected that the number of nodes which hold the pod have the same effect on yield (Egli, 2013). However, the number of pods are affected by management practices during the reproductive period (Board and Tan, 1995), such an occurrence may explain the low indirect contribution of number of pods via nodes per plant. In the same vain, the dry spell experienced during the grain filling stage may have affected the number of seeds in each pod translating into an un-proportional number of seed per plant. The results show that effective selection for superior genotypes is possible if harvest index and number of pods per plant are considered.

5.10 Adaptive Main Effects and Multiplication Interaction (AMMI)

Additive AMMI analysis of variance showed evidence of significant interaction between genotypes and the environments in which they were grown. AMMI ranking further classified the interaction as cross over or qualitative interaction. This was evident from the differential ranking of genotypes among locations for the trait of yield (Appendix 2). The complex interactions where further explained by the large number of significant IPCA scores. Crossover interaction is known to have negative implications for breeding progress. This is because, such an interaction requires selection of many genotypes from different test locations, thus lengthening the selection process and reducing genetic gains, thereby retarding genetic progress (Annicchiarico, 2002). These findings are similar to those found by Cucolotto *et al.* (2007). Much earlier, Mushoriwa (2013) evaluated the breeding programme which is the subject of this study and reported cross over interaction for 42 genotypes across 13 test locations. This form of interaction implies that the genotypes generally have specific adaptation to certain locations and cannot be recommended as the best across all sites. The genotypes may be highly responsive to changes in the environment in which they are grown.

AMMI identified genotype G2 as the most widely adapted genotype. G2 was among the top performers in 9 of the total 16 test locations spread out across the three countries represented

in the study. These locations represented stress prone environments (hot regions on Zimbabwe; E4, E5 and E3) and cooler regions (E2, E3, E8, E9, E15 and E16). The second most adapted genotype was G15 being among top four performers in E1, E7, E12, E14, E15 and E16. These are locations where G2 was either a better performer or completely outside the top genotypes. In addition, G15 performed best in high management areas. Another genotype to note was G10 a commercial genotype, which performed well in five locations (E4, E5, E9, E11 and E15). These three genotypes make good candidates for further testing and possibly release in the three countries due to their wide adaptation.

5.11 GGE biplot analysis

5.11.1 Genotype adaptation

The polygon view of a GGE biplot is used to identify genotypes that are best adapted to test locations. It attempts to group test locations into mega environments together within boundaries known as sectors. The biplot analysis divided the test locations among five sectors and at least three mega environments.

In this study, the biplot singled out genotype G10 as the most responsive in the sector holding locations E5, E15 and E16. It showed specific adaptation to location E15. Genotype G15 was the most widely adapted as it was responsive to the sector that contained the largest cluster of test locations. However, it was most adapted to locations E12 and E8. Genotype G2 was most adapted to the highest yielding environment E11. No locations fell in the sectors housing genotypes G6, G12, G13, G16, G17, G18, G22, G23 and G28. Therefore, the genotypes did not show adaptation to any test location.

The GGE biplot analysis also identified G2 as the ideal genotype followed by G9 and the commercial check G15. An “ideal” genotype is defined as one that is high yielding and stable across test locations. In reality, such a genotype may not exist, but serves as a point of reference for genotype evaluation.

5.11.2 Discrimination and representativeness of test environments

An ideal environment GGE biplot graphically depicts the discriminating ability and representativeness of locations used in genotype evaluation. An ideal location is one that is most discriminating of genotypes and is representative of test locations. In this study, E4 was identified as the most representative location but less discriminative. Most of the test locations where separated from E4 by an acute angle (<90^{o}) and therefore, very close to it. Location E6 was identified as the most discriminating of test locations by virtue of having a high PC1 score.

This meant that E6 allowed genotypes to fully express themselves so that material of higher yield potential could be distinguished from those of lower yield potential. It was the best for genetic differentiation better than any other location. However, E6 was not representative of other locations as it was further from the AEC. E11 was the most representative of locations even though it showed lower discrimination among genotypes, it had a good balance of discrimination and representativeness.

The angle between vectors connecting two sites can be used to determine the existence of relatedness between sites (Yan and Tinker, 2006). Relatedness allows the breeder to use their discretion and drop some sites in order to reduce costs of testing without compromising the value of the data realised. Many of the locations used in this study showed correlation or relatedness by having small acute angles between them, with the exception of E5, E6 and E15. A strong relationship was also seen between E5, E6 and E15. This resulted in two sets of unique locations. The lack of correlation between the two sets meant that they did not differentiate the genotypes equally; therefore, completely dropping one set of sites would result in loss of data so that meaningful conclusions cannot be made from it. Therefore, in future trials the two sets of locations should be represented, as high variability exists among them.

6 GENERAL CONCLUSION

6.1 Introduction

This chapter relates the findings of the study to the objectives set in chapter one. These were:

- To assess the nature of genotype x environmental interactions of soybean grain yield,
- To identify consistently well ranked advanced soybean lines in medium altitude and subtropical environments of Zambia, Zimbabwe and Malawi.
- To identify the most ideal test environment for the genotypes under evaluation
- To determine the genetic gains achieved in breeding for high yield and stability of 30 advanced soybean lines in the Zambian breeding programme between 1996 and 2007.
- To determine the secondary traits that made direct and indirect contributions to increase in yield potential realised from the soybean breeding programme in Zambia between 1996 and 2007.

It therefore provides a summary of essential components of the study for which all detailed processes and findings have been presented in each chapter preceding the general conclusion.

6.2 Summary of findings from the study

The following conclusions were made from the study undertaken:

- Genotype environment interactions were evaluated and found to significantly affect the performance of genotypes at each site. AMMI showed that the interaction was of the cross over type as reflected by the difference in genotype ranking from one test location to another. AMMI also had 7 significant IPCA scores, these findings together with a large number of sectors (>5) and mega environments from the GGE biplot further emphasised the complexity of GXE interaction that was at play in the expression of grain yield among the genotypes.
- Genotype stability was evaluated and according to GGE analysis, the most stable genotypes were G2, G9 and G15. This was in agreement with the AMMI finding of G2, G10 and G15 being the most stable. GGE biplot analysis also revealed that G2 was

the ideal genotype due to its closeness high yielding ability and increased stability. The cultivar superiority index (CSI) identified G15 as the most stable genotype.

- Specific adaptation was observed by G2 to the ideal test location E11. The genotype showed specific adaptation to E15 while G11 was specifically adapted to E1.
- Test locations showed good discrimination among genotypes. However, the most discriminative location was E6, while E4 was the most representative of test locations. Ideal location for testing was E11 and it was also the highest yielding location to which G2 was most adapted.
- The lines G2 and G15 showed wide adaptation in all the countries represented in this study.
- Two sets of related test locations were identified from the biplot pattern so that each set will need to be represented in future testing. This will be necessary to preserve the integrity of the data to be collected. Repetition over seasons may also be necessary to determine which locations can then be overlooked in further testing in order to reduce testing costs to the breeder.
- Grain yield was moderately heritable (0.2-0.86) across sites. This indicated that environmental variation was higher than genetic variation in this study.
- The genetic gain study showed a gain of 10-21% of the genotype G2 over the best check G3 in Zambia and Malawi. However, in Zimbabwe, no significant genetic gain was registered. G2 showed a 2-19% advantage over G15 at five sites. But this yield advantage was not apparent in the remaining 4 sites.
- Overall, all observed genetic gain in the one season dissapeared when the 16 sites in the three countries where analysed together. Indicating that the GXE observed may have lowered the heritability of yield and negatively affected genetic gain. The breeding programme had an annual genetic yield gain of only 2.80 kg ha^{-1} $year^{-1}$ translating into 0.08% increase in yield, over a period of 12 years.
- A 6.5% gain over the population mean was observed, showing that selection was successful in increasing yield. However, there was no significant gain that was observed relative to the current commercial cultivars, indicating limited breeding progress.
- The study also revealed high genetic variability of traits among genotypes, which can be exploited to obtain further breeding gains.
- Analysis of secondary yield traits revealed that heritability of yield was 0 while that of harvest index was higher at 0.12.

- Harvest index was the most important trait for indirect selection of grain yield. It showed strong and significant correlation to grain yield (0.8), and a positive and high direct (0.95) and indirect (0.79) path to effecting higher mean yield. Traits of number of pods per plant and biomass may also be useful in developing a yield selection index due to their significant coefficient of determination on yield.

6.3 Summing up

The study was successful in addressing the objectives set out in chapter one.

- G x E interaction was highly significant for the trait of grain yield and was of the crossover type.
- The lines G2, G10 and G15 were consistently well ranked in medium altitude and subtropical environments of Zambia, Zimbabwe and Malawi.
- The ideal test environment for the genotypes under evaluation was identified as E11 for its representativeness and ability to discriminate among genotypes.
- No significant genetic gains were achieved in breeding for high yield and stability of 30 advanced soybean lines in the soybean breeding programme between 1996 and 2007.
- Harvest index had positive and strong direct and indirect contributions to increase in yield potential.

Following the success of the study, recommendations that can be made from the findings are outlined in the next chapter.

7 RECOMMENDATIONS

7.1 Introduction

This chapter draws on the conclusions and findings of the study outlined in previous chapters to outline what measures can be taken to address the issues raised and their implications for the breeding programme

- The lines G2, G10 and G15, which had good general adaptation, should be considered for release in the three countries, Zambia, Malawi and Zimbabwe. Genotype 2 could also be specifically released in E11.This will help to reduce complex GXE effects on grain yield
- Considering E11 was the ideal testing location, further tests involving the genotypes evaluated should prioritize this location as a way of saving on resources and obtaining the most representative data. The locations around E11 should be considered as one mega environment with similar patterns of GXE.
- In order to increase genetic yield gains above those of commercial varieties, a revision of the current crossing strategy should be considered. The new strategies should match the current breeding objective of the breeding programme, which is the improvement of grain yield potential. One possibility could be the crossing of elite lines to other elite lines with known and verified high yield performance.
- Increase selection intensity in order to skew the breeders' equation in the positive direction.

$$\Delta G = h^2 \times (Selection\ differential)$$

$Selection\ differential$ = Mean of selected- Mean of population

- Management of trials should continuously be improved in order to raise heritability of yield as this will reduce on environmental interactions and positively influence yield gains.
- This study should be repeated over a number of seasons in order to observe the effects of time on GXE interaction.
- Harvest index that translates to increased weight of seed per plant (Hundred seed weight) should be considered for indirect selection for higher grain yield.

REFERENCES

Alves, B. J., R. M. Boddey, and S. Urquiaga. 2003. The success of BNF in soybean in Brazil. Plant and Soil 252:1-9.

Amira, J., D. Ojo, O. Ariyo, O. Oduwaye, and M. Ayo-Vaughan. 2013. Relative discriminating powers of GGE and AMMI models in the selection of tropical soybean genotypes. African Crop Science Journal 21:67-73.

Annicchiarico, P. 2002. Defining Adaptation Strategies and Yield-stability Targets in Breeding Programmes. Quantitative genetics, genomics, and plant breeding:p. 365-383.

Annicchiarico, P., F. Bellah, and T. Chiari. 2005. Defining Subregions and Estimating Benefits for a Specific-Adaptation Strategy by Breeding Programs. Crop Science 45:1741-1749.

Arshad, W., M. Zeeshan, M. I. Khan, S. Ali, M. Hussain, and S. Rahman. 2014. Character association and causal effect analysis for yield and yield components among early maturing genotypes of soybean (*Glycine max* L. Merrill). Journal of Renewable Agriculture 1:4-11.

Bilyeu, K., M. Ratnaparkhe, and C. Kole. 2010. Genetics, genomics, and breeding of soybeans. Genetics, genomics, and breeding of soybeans.

BMS. 2014. Breeding Management System workbench. Version 3.0.8.

Board, J. E., and Q. Tan. 1995. Assimilatory Capacity Effects on Soybean Yield Components and Pod Number. Crop Science 35:846-851.

Burkill, I. H. 1966. A dictionary of the economic products of the Malay Peninsula. A Dictionary of the Economic Products of the Malay Peninsula. 2.

Cramer, C. S., and T. C. Wehner. 1998. Fruit yield and yield component means and correlations of four slicing cucumber populations improved through six to ten cycles of recurrent selection. Journal of the American Society for Horticultural Science 123:388-395.

Cramer, C. S., and T. C. Wehner. 2000. Path analysis of the correlation between fruit number and plant traits of cucumber populations. HortScience 35:708-711.

Crossa, J. 1990. Statistical analyses of multilocation trials. Advances in Agronomy 44:55-85.

Crossa, J., P. L. Cornelius, K. Sayre, R. Ortiz-Monasterio, and J. Iván. 1995. A shifted multiplicative model fusion method for grouping environments without cultivar rank change. Crop Science 35:54-62.

Cucolotto, M., D. Garbuglio, V. Pipolo, N. da Silva Fonseca Junior, and M. Kamikoga. 2007. Genotype x environment interaction in soybean: evaluation through three methodologies. Crop Breeding and Applied Technology 7:270-277.

de Oliveira, R. A., M. D. V. de Resende, E. Daros, J. C. B. Filho, J. Zambon, O. Ido*, et al.* 2005. Genotypic evaluation and selection of sugarcane clones in three environments in the state of Paraná. Crop Breeding and Applied Technology 5:426-434.

Duvick, D. 2005. Genetic progress in yield of United States maize (*Zea mays* L.). Maydica 50:193.

Egli, D. 2008a. Comparison of Corn and Soybean Yields in the United States: Historical Trends and Future Prospects.

Egli, D. 2008b. Soybean yield trends from 1972 to 2003 in mid-western USA. Field Crops Research 106:53-59.

Egli, D. B. 2013. The Relationship between the Number of Nodes and Pods in Soybean Communities. Crop Science 53:1668-1676.

Falconer, D. S. 1960. Introduction to quantitative genetics.

FAOSTAT. 2015. Area of Soybean planted area vs yields harvested in tonnes per hectare Food and Agriculture Organisation http://faostat3.fao.org/download/Q/QC/E (Accessed 30 May 2015).

Finlay, K., and G. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. Crop and Pasture Science 14:742-754.

Fisher, R. A. 1925. Theory of statistical estimation. Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press.

Fox, P., J. Crossa, and I. Romagosa. 1997. Multi-environment testing and genotype× environment interaction. Statistical methods for plant variety evaluation. Springer. p. 117-138.

Gauch Jr, H. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs Elsevier Science Publishers.

Gauch Jr, H. G. 1988. Model selection and validation for yield trials with interaction. Biometrics:705-715.

Giami, S. 2002. Chemical composition and nutritional attributes of selected newly developed lines of soybean (*Glycine max* (L) Merr). Journal of the Science of Food and Agriculture 82:1735-1739.

Goldsmith, P. D. 2008. Economics of soybean production, marketing, and utilization. Soybeans: chemistry production processing and utilization. AOCS Press, Urbana:117-150.

Gurmu, F., H. Mohammed, and G. Alemaw. 2009. Genotype x Environment interactions and stability of soybean for grain yield and nutrition quality. African Crop Science Journal 17:87-99.

Gwata, E., D. Wofford, K. Boote, A. Blount, and P. Pfahler. 2005. Inheritance of promiscuous nodulation in soybean. Crop Science 45:635-638.

Hallauer, A. R., J. Miranda Filho, and M. J. Carena. 2010. Hereditary variance: mating designs. Quantitative genetics in maize breeding. Springer. p. 81-167.

Hallauer, A. R., and J. Miranda. 1988. Quantitative genetics in maize breeding. Iowa State Univ Press, Ames, IA. Jenkins, JN, WL Parrott, JC McCarty Jr & RL Shepherd, 1988a. Registration of three noncommercial germplasm lines of upland cotton tolerant to tobacco budworm. Crop Science 28:90-91.

Hymowitz, T. 2008. The history of the soybean. Soybeans: chemistry, production, processing, and utilization:p.1-32.

Hymowitz, T. 2004. Speciation and cytogenetics. In: H. R. Boerma and J. E. Specht (eds.)

Soybeans: improvement, production and uses. American Society of Agronomy, Crop

Science Society of America and Soil Science Society of America:p. 97-136.

Iqbal, A., I. A. Khalil, N. Ateeq, and M. Sayyar Khan. 2006. Nutritional quality of important food legumes. Food Chemistry 97:331-335.

Jarvie, J., and P. Shanahan. 2009. Assessing tolerance to soybean rust in selected genotypes. Field Crops Research 114:419-425.

Jarvie, J. A. 2008. Optimising aspects of a soybean breeding programme.

Kanamaru, K., S. Wang, J. Abe, T. Yamada, and K. Kitamura. 2006. Identification and characterization of wild soybean (*Glycine soja* Sieb. et Zecc.) strains with high lutein content. Breeding Science 56:231-234.

Karasu, A., M. Oz, A. Göksoy, and Z. Turan. 2009. Genotype by environment interactions, stability, and heritability of seed yield and certain agronomical traits in soybean [*Glycine max* (L.) Merr.]. African Journal of Biotechnology 8:580-590.

Kearsey, M. J., and H. S. Pooni. 1998. The genetical analysis of quantitative traitsStanley Thornes (Publishers) Ltd.

Keatinge, J., W. Easdown, R. Yang, M. Chadha, and S. Shanmugasundaram. 2011. Overcoming chronic malnutrition in a future warming world: the key importance of mungbean and vegetable soybean. Euphytica 180:129-141.

Kokubun, M. 2011. Physiological mechanisms regulating flower abortion in soybean. Soybean-Biochemistry, Chemistry and Physiology. InTech.

Krishnamurthy, P., R. J. Singh, C. Tsukamoto, J. H. Park, J. D. Lee, and G. Chung. 2013. Kunitz trypsin inhibitor polymorphism in the Korean wild soybean (*Glycine soja* Sieb. & Zucc.). Plant Breeding 132:311-316.

Laker, M. C. 2013. Soil fertility in sub-saharan africa and the effect thereof on human nutrition.

Lange, C. E., and L. C. Federizzi. 2009. Estimation of soybean genetic progress in the South of Brazil using multi-environmental yield trials. Scientia Agricola 66:309-316.

Laurance, W. F., J. Sayer, and K. G. Cassman. 2014. Agricultural expansion and its impacts on tropical nature. Trends in Ecology & Evolution 29:107-116.

Levy, C. 2005. Epidemiology and chemical control of soybean rust in southern Africa. Plant Disease 89:669-674.

Lin, C.-S., and M. R. Binns. 1988. A superiority measure of cultivar performance for cultivar× location data. Canadian Journal of Plant Science 68:193-198.

Machikowa, T., and P. Laosuwan. 2011. Path coefficient analysis for yield of early maturing soybean. Sonklanakarin Journal of Science and Technology 33:365.

Malik, M. F. A., M. Ashraf, A. S. Qureshi, and A. Ghafoor. 2007. Assessment of genetic variability, correlation and path analyses for yield and its components in soybean. Pakistan Journal of Botany 39:405-413.

McLean, R., and D. Byth. 1980. Inheritance of resistance to rust (Phakopsora pachyrhizi) in soybeans. Crop and Pasture Science 31:951-956.

Mian, M., A. Missaoui, D. Walker, D. Phillips, and H. Boerma. 2008. Frogeye Leaf Spot of Soybean: A review and proposed race designations for isolates of Hara. Crop Science 48:14-24.

Miles, M., W. Morel, J. Ray, J. Smith, R. Frederick, and G. Hartman. 2008. Adult plant evaluation of soybean accessions for resistance to Phakopsora pachyrhizi in the field and greenhouse in Paraguay. Plant Disease 92:96-105.

Munawar, M., G. Hammad, and M. Shahbaz. 2013. Evaluation of Maize (*Zea mays* L.) Hybrids under Different Environments by GGE Biplot Analysis. Agriculture and Environmental Science journal 13:1252-1257.

Mushoriwa, H. 2013. Breeding Gains, Diversity Analysis and Inheritance Studies on Soybean [*Glycine max* (L.) Merrill] Germplasm in Zimbabwe. MSc thesis, University of KwaZulu-Natal, Pietermaritzburg.

Muthoni, J., H. Shimelis, and R. Melis. 2015. Genotype x Environment Interaction and Stability of Potato Tuber Yield and Bacterial Wilt Resistance in Kenya. Am. J. Potato Res. American Journal of Potato Research : The Official Journal of the Potato Association of America 92:367-378.

Natarajan, S. S., C. Xu, H. Bae, T. J. Caperna, and W. M. Garrett. 2006. Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. Journal of agricultural and food chemistry 54:3114-3120.

Ngalamu, T., M. Ashraf, and M. Silvestro. 2013. Soybean (*Glycine max* L) Genotype and Environment Interaction Effect on Yield and Other Related Traits. American Journal of Experimental Agriculture 3:977-987.

Nyquist, W. E., and R. Baker. 1991. Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Sciences 10:235-322.

Ogoke, I., R. Carsky, A. Togun, and K. Dashiell. 2003. Effect of P fertilizer application on N balance of soybean crop in the guinea savanna of Nigeria. Agriculture, Ecosystems and Environment 2:153-159.

Opperman, C., and S. Varia. 2011. Technical Report: Soybean Value Chain. AECOM International Development. (Dans le texte).

Pimentel, D., and T. W. Patzek. 2005. Ethanol production using corn, switchgrass, and wood; biodiesel production using soybean and sunflower. Natural Resources Research 14:65-76.

Ravindran, V., and R. Blair. 1993. Feed resources for poultry production in Asia and the Pacific. III. Animal protein sources. World's Poultry Science Journal 49:219-235.

Ray, D. K., N. D. Mueller, P. C. West, and J. A. Foley. 2013. Yield trends are insufficient to double global crop production by 2050.

Robinson, H., R. E. Comstock, and P. Harvey. 1949. Estimates of heritability and the degree of dominance in corn. Agronomy Journal.

Romagosa, I., and P. Fox. 1993. Genotype× environment interaction and adaptation. Plant Breeding. Springer. p. 373-390.

Samonte, S. O. P., L. T. Wilson, A. M. McClung, and J. C. Medley. 2005. Targeting cultivars onto rice growing environments using AMMI and SREG GGE biplot analyses. Crop Science 45:2414-2424.

SAS Institute. 2012. The SAS system for Windows. Release 9.3. SAS Institute Inc, Cary, NC, USA.

SCCI, S. C. a. C. I. 2014. Official variety register.

Schroeder, C., T. K. O. Onyango, R. B. Nar, N. Jick, H. Parzies, and D. Gemenet. 2013. Potentials of hybrid maize varieties for small-holder farmers in Kenya: a review based on Swot analysis. African Journal of Food, Agriculture, Nutrition and Development 13.

Shanmugasundaram, S. 1977. The international working group on soybean rust and its proposed soybean rust rating system. Workshop on Rust of Soybean-the Problem and Research Needs, Manila (Philippines), 28 Feb-4 Mar 1977.

Shurtleff, W., and A. Aoyagi. 2009. History of Soybeans and Soyfoods in Africa (1857-2009): Extensively Annotated Bibliography and Source book Soy info Center.

Singh, G. 2010. The soybean: botany, production and uses CABI.

Singh, R. K., and B. D. Chaudhary. 2010. Biometrical Methods in Quantitative Genetic AnalysisKalyani Publishers, New Delhi.

Sleper, D. A., and J. M. Poehlman. 2006. Breeding field crops. Fifth ed. Blackwell publishing, Iowa USA.

Smit, M. A. 1987. Soyabean production in South Africa,. In L.L Lötter (ed.).

Research Highlights, 1987.

Souza, A. R. R., G. V. Miranda, M. G. Pereira, and L. V. d. Souza. 2009. Predicting the genetic gain in the Brazilian white maize landrace. Ciência Rural 39:19-24.

Specht, J., D. Hume, and S. Kumudini. 1999. Soybean yield potential—a genetic and physiological perspective. Crop Science 39:1560-1570.

Steel, R., and J. H. Torrie. 1980. Principles and Procedures of Statistics: A Biometrical Approach. McGraw-Hill, New York, NY.

TechnoServe. 2011. Southern Africa Soy Roadmap-Zambia Value Chain Analysis. Technoserve.http://www.technoserve.org/files/downloads/technoserve-bmgf-zambia.pdf (Accessed 25 April. 2015).

Tefera, H., B. Asafo-Adjei, and K. E. Dashiell. 2010. Breeding progress for grain yield and associated traits in medium and late maturing promiscuous soybeans in Nigeria. Euphytica 175:251-260.

Tilman, D., C. Balzer, J. Hill, and B. L. Befort. 2011. Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences 108:20260-20264.

Tukamuhabwa, P., M. Asiimwe, M. Nabasirye, P. Kabayi, and M. Maphosa. 2012. Genotype by environment interaction of advanced generation soybean lines for grain yield in Uganda. African Crop Science Journal 20:107-116.

Ullah, H., I. H. Khalil, I. A. Khalil, M. Fayaz, J. Yan, and F. Ali. 2012. Selecting high yielding and stable mungbean [*Vigna radiata* (L.) Wilczek] genotypes using GGE biplot techniques. Canadian Journal of Plant Science 92:951-960.

Wang, T. L., C. Domoney, C. L. Hedley, R. Casey, and M. A. Grusak. 2003. Can we improve the nutritional quality of legume seeds? Plant Physiology 131:886-891.

Wilcox, J. R. 2004. 1 World Distribution and Trade of Soybean.

Wilson, R. F. 2004. Seed composition. Soybeans: improvement, production, and uses: p.621-677.

Wricke, G., and E. Weber. 1986. Quantitative genetics and selection in plant breedingWalter de Gruyter.

Yan, W., P. L. Cornelius, J. Crossa, and L. A. Hunt. 2001. Two types of GGE biplots for analyzing multi-environment trial data. Crop Science 41:656-663.

Yan, W., L. Hunt, Q. Sheng, and Z. Szlavnics. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Science 40:597-605.

Yan, W., M. S. Kang, B. Ma, S. Woods, and P. L. Cornelius. 2007. GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Science 47:643-653.

Yan, W., and I. Rajcan. 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. Crop Science 42:11-20.

Yan, W., and N. A. Tinker. 2006. Biplot analysis of multi-environment trial data: Principles and applications. Canadian Journal of Plant Science 86:623-645.

Ye, G., D. McNeil, and G. Hill. 2001. Methods for analysing multi-site plant variety trials. II. Selection for yield and stability. Agronomy New Zealand 31:25-33.

APPENDICES

Apendix 1 AMMI ranking of 28 genotypes in 16 test locations

Genotype	E1	Rank	E2	Rank	E3	Rank	E4	Rank	E5	Rank	E6	Rank	E7	Rank	E8	Rank	E9	Rank
1	3821.82	5	3182.10	11	1532.92	23	2197.90	27	3204.72	25	2763.29	18	2711.40	17	4151.58	11	3204.72	25
2	2665.82	24	3634.42	1	2207.46	1	3285.93	1	4118.33	2	3322.63	8	2934.44	12	5187.20	1	4118.33	2
4	3599.62	8	2916.57	16	1725.57	16	2552.68	15	3545.83	12	2522.64	19	2936.39	11	3469.83	25	3545.83	12
5	2319.04	28	2247.92	28	1671.11	17	2220.58	26	3510.49	14	3714.37	4	2344.18	27	3630.74	22	3510.49	14
6	2655.34	26	3046.74	15	1730.47	15	2465.33	20	3115.07	27	3031.91	12	3014.33	7	5012.13	3	3115.07	27
7	3278.38	12	3165.97	12	1811.85	9	3045.35	2	3506.92	15	2958.60	14	2833.04	16	4039.07	15	3506.92	15
8	3103.31	19	2691.97	23	1480.93	26	2291.77	25	3747.81	7	2900.49	16	2646.16	19	3945.73	18	3747.81	7
9	3988.07	2	3502.16	3	1837.43	8	2496.00	18	3233.03	24	3799.17	2	3004.11	9	5161.94	2	3233.03	24
10	3094.38	20	3283.99	10	1743.49	14	3036.11	3	4827.89	1	3678.77	5	2552.95	22	4109.76	12	4827.89	1
11	3353.53	10	2843.81	18	2054.70	2	2444.73	21	3656.84	11	3798.17	3	2908.71	13	3938.16	19	3656.84	11
12	3441.84	9	2856.22	17	1869.42	7	2532.69	16	3489.05	16	2234.02	23	2324.86	28	3369.98	26	3489.05	16
13	2472.33	27	2486.71	26	1628.26	19	2721.02	9	3518.42	13	1390.01	27	2898.90	14	4043.93	14	3518.42	13
14	4207.52	1	3290.92	8	1902.25	5	2734.50	8	3120.79	26	2999.25	13	3337.85	1	4713.14	4	3120.79	26
15	3914.39	3	3363.78	4	1769.91	12	2807.16	5	3873.70	5	3167.91	9	3312.05	2	4506.16	5	3873.70	5
16	2662.03	25	2402.04	27	1481.16	25	2557.26	14	2781.66	28	2395.11	21	2447.57	24	3367.05	27	2781.66	28
17	3224.85	14	2528.51	25	1777.38	11	2304.79	23	3484.37	17	1523.78	26	2600.21	20	3581.04	23	3484.37	17
18	3833.20	4	3357.45	6	1574.04	21	2755.77	6	3325.65	22	1909.58	24	3010.42	8	4022.47	17	3325.65	22
19	3147.95	16	3088.34	13	1882.06	6	2750.26	7	3738.97	8	2428.30	20	2394.55	26	4033.04	16	3738.97	8
20	3179.42	15	2707.71	21	1230.55	28	2693.07	10	3905.90	4	2870.13	17	2558.26	21	3844.17	20	3905.90	4
21	2834.80	22	3285.26	9	2042.86	3	2503.49	17	3302.59	23	3426.97	7	2979.18	10	4192.67	9	3302.59	23
22	3130.47	18	3555.08	2	1988.99	4	2677.58	12	3468.33	18	1267.93	28	2861.56	15	4384.50	8	3468.33	18
23	2824.77	23	2808.67	20	1792.52	10	3007.03	4	3726.88	9	2358.94	22	3191.13	3	3692.33	21	3726.88	9

24	3649.47	7	2530.40	24	1746.24	13	2369.77	22	3383.62	20	3074.73	11	3147.36	4	4398.61	7	3383.62	20
25	3135.39	17	2707.16	22	1546.63	22	2592.72	13	3802.57	6	3133.03	10	2551.16	23	4060.12	13	3802.57	6
26	3682.59	6	3357.75	5	1667.08	18	2683.64	11	3339.09	21	3809.37	1	3119.59	5	4188.25	10	3339.09	21
27	3262.97	13	3072.77	14	1350.29	27	2193.64	28	3992.61	3	3549.92	6	3084.08	6	4426.22	6	3992.61	3
28	3006.62	21	2833.07	19	1513.83	24	2495.16	19	3389.92	19	2939.43	15	2436.59	25	3316.79	28	3389.92	19
30	3332.12	11	3327.54	7	1606.71	20	2292.73	24	3673.26	10	1821.56	25	2668.76	18	3493.40	24	3673.26	10
Mean	3243.644		3002.68		1720.219		2596.738		3563.726		2813.929		2814.635		4081.429			

Apendix 2 AMMI ranking of 28 genotypes continued

Genotype	E10	Rank	E11	Rank	E12	Rank	E13	Rank	E14	Rank	E15	Rank	E16	Rank	Mean	Rank
1	3143.19	19	3972.54	25	3694.11	2	3019.49	13	3781.42	14	3394.80	6	2207.30	23	3123.96	13
2	3524.22	6	4800.32	8	2930.12	8	3317.78	1	3850.94	9	3744.73	2	2739.18	4	3523.87	1
4	3170.33	17	5033.71	3	2056.15	27	2792.85	20	3737.15	16	2599.65	25	2494.84	14	3043.73	20
5	3059.05	22	4567.44	18	2561.92	17	2825.74	18	3127.09	27	3249.78	8	2037.17	28	2912.32	25
6	3030.73	24	4695.82	14	2969.54	7	2931.09	15	3366.17	24	2910.90	16	2256.15	21	3084.18	16
7	3205.63	14	4295.13	22	2070.62	26	2731.24	22	3813.31	11	2805.93	21	2790.82	3	3116.17	14
8	3171.05	16	4739.55	11	2720.38	13	2866.14	17	3413.34	22	3569.50	5	2133.94	27	3073.12	19
9	3202.25	15	5021.10	4	2282.10	21	2983.95	14	3989.01	3	3068.12	13	2344.80	17	3321.64	5
10	3469.86	7	5075.57	2	2272.89	22	2894.92	16	3795.03	13	4551.02	1	2584.86	10	3487.46	3
11	3674.84	1	4583.33	17	3725.83	1	3223.66	3	3939.85	7	3051.00	14	2716.75	5	3348.17	4
12	3209.22	13	4498.77	19	1834.72	28	3088.56	9	3758.72	15	2627.00	24	2158.50	25	2923.91	24
13	3140.90	20	4717.22	12	2176.52	24	3033.28	11	3131.88	26	2866.81	18	2225.78	22	2873.15	26
14	3361.96	10	4680.34	15	2681.71	14	3158.81	5	4021.87	1	2883.14	17	2659.46	8	3304.64	6
15	3528.51	5	4784.08	9	3347.77	4	3063.15	10	4011.01	2	3686.31	3	2835.47	2	3490.32	2
16	2650.88	28	3644.81	28	2082.06	25	2610.73	24	3074.58	28	2660.90	23	2134.99	26	2608.41	28
17	3565.91	4	4354.15	20	2919.62	9	3261.06	2	3708.06	17	2371.63	27	2294.68	20	2936.53	23
18	3127.06	21	4183.54	24	2741.92	12	2821.19	19	3867.31	8	2862.69	19	2609.21	9	3082.95	17
19	3574.18	3	3950.41	27	2817.38	10	3104.31	7	3971.59	4	2814.94	20	2579.30	11	3125.91	12
20	3313.03	12	3971.45	26	2340.86	18	2387.40	27	3624.15	19	3371.47	7	2680.21	7	3036.48	21
21	2960.47	26	4948.22	6	3337.66	5	3100.20	8	3518.67	21	3108.29	11	2297.10	19	3196.31	9
22	3350.86	11	4759.08	10	2749.07	11	3177.92	4	3955.38	5	2040.42	28	2463.95	15	3081.22	18
23	3036.61	23	4864.72	7	2618.58	15	3032.51	12	3252.21	25	3619.77	4	2515.00	13	3129.28	11
24	3416.55	8	5203.94	1	2285.38	20	3131.44	6	3596.32	20	2995.03	15	2320.34	18	3164.55	10
25	3407.93	9	4350.53	21	2320.39	19	2726.96	23	3671.51	18	3187.66	9	2515.58	12	3094.50	15
26	2987.93	25	4699.79	13	2582.66	16	2552.57	26	3809.19	12	3096.13	12	2703.84	6	3226.16	8
27	3583.85	2	5001.43	5	3381.13	3	2308.75	28	3950.25	6	2696.99	22	2923.84	1	3298.21	7

28	2771.76	27	4248.11	23	2258.01	23	2580.93	25	3369.01	23	3175.22	10	2192.23	24	2869.79	27
30	3148.24	18	4661.57	16	3120.89	6	2792.70	21	3824.87	10	2594.93	26	2354.69	16	3024.20	22
Mean	3242.393		4582.381		2674.286		2911.406		3676.067		3057.314		2456.071			