Biometrical approaches for investigating genetic improvement in wheat breeding in South Africa

by

Mardé Booyse

A thesis
submitted in fulfilment of
the requirements of the degree

Philosophiae Doctor

in the Department of Plant Sciences (Plant Breeding)
Faculty of Natural and Agricultural Sciences
University of the Free State
Bloemfontein

January 2014

Promoter: Prof M.T. Labuschagne

Co-promoter: Prof K.W. Pakendorf

Table of Contents

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public. I further cede copyright of the thesis in favour of the University of the Free State.

Signed……..

Mardé Booyse

Acknowledgements

I wish to express my sincere appreciation to the following individuals for their various contributions to this study:

My promoters, Prof. Maryke Labuschagne and Prof. Klaus Pakendorf for their guidance, enthusiasm and motivation.

Dr Cobus le Roux, General Manager of the Agricultural Research Council (ARC) Grain Crops Division for the privilege to use the data of ARC Small Grain Institute (ARC-SGI) to conduct the study.

Mrs Hesta Hatting, Dr Rachel Oelofse, Mr Diederick Exley, Dr Eben von Well and Mr Petrus Delport of ARC-SGI for their assistance with the data.

My employer, ARC Biometry, and colleagues for the support.

The librarians, Mrs Juliette Killian of ARC-SGI and Mrs Annemari du Preez of the University of the Free State's library for their assistance.

Mrs Sadie Geldenhuys for all the administration and motivation.

The statisticians and biometricians, Prof. Dirk Van Schalkwyk, Mrs Marie Smith, Dr José Crossa, Prof. Brian Cullis, Dr Alberto de la Vega, Dr Kym Butler, Dr Roger Payne of Genstat and Mr Bongani Ndlovu of SAS Technical Support.

Mrs Irene Joubert from ARC Institute for Soil, Climate and Water (ARC-ISCW) for the Google maps.

Ms Anélia Marais for the language editing of the thesis. Mrs Selene Delport for her assistance with the formatting, layout and graphical presentation of the thesis.

Mrs Hannél Ham and Mrs Zelda Bijzet for their continuous assistance and patience with the formatting and layout of the thesis when I have disrupted the previous work.

My parents, sister, family and friends for their motivation and patience.

But most of all to my Father in Heaven.

Dedication

I dedicate this thesis to my parents and sister, without their continuous love, support, motivation and prayers I could not do without.

O LORD, You have searched me and known *me.*
[2] You know my sitting down and my rising up;
You understand my thought afar off.
[3] You comprehend my path and my lying down,
And are acquainted with all my ways.
[4] For *there is* not a word on my tongue,
But behold, O LORD, You know it altogether.
[5] You have hedged me behind and before,
And laid Your hand upon me.
[6] *Such* knowledge *is* too wonderful for me;
It is high, I cannot *attain* it.
Ps 139:1-5

List of abbreviations and symbols

a	Intercept of the linear regression
AMMI	Additive main effects and multiplicative interaction
ANOVA	Analysis of variance
ARC	Agricultural Research Council
ASV	AMMI Stability Value
b	Slope of the linear regression
BLUE	Best Linear Unbiased Estimator
BLUP	Best Linear Unbiased Predictor
CA	Cluster analysis
Check	Check mean
CV	Coefficient of Variation (%)
CV1	Canonical variate 1 (in discriminant analysis)
CV2	Canonical variate 2 (in discriminant analysis)
DA	Discriminant Analysis
E	Environment
G	Genotype
ΔG	Genetic Advance
GA	Genetic Advance
GCV	Genetic Coefficient of Variation (%)
GEI	Genotype-by-Environment Interaction
GGE	Genotype + Genotype-by-Environment Interaction
GLM	General Linear Model
h^2	Heritability
HLM	Hectolitre Mass
i	Standardised selection differential
IPCA	Interaction Principal Component Analysis
LDA	Linear Discriminant Analysis
Mc	The difference between the mean of the five best genotypes and the check mean
MT	The difference between the mean of the five best genotypes and the trial mean
MET	Multi-environment trial(s)
MIXED	Linear Procedure with both fixed and random effects
MS	Mean Squares
P/p	Probability=Significance level
PC	Principal Component
PCA	Principal Component Analysis
PCV	Phenotypic Coefficient of Variation (%)
PROC	Procedure
r	Pearson's product -moment correlation coefficient
R	Response to selection
R^2	Regression coefficient of determination
RCBD	Randomised Complete Block Design
REML	Restricted maximum likelihood estimation
S	Standard deviation of the sample
S^2	Variance of the sample
SD	Standard deviation of the sample
SE	Standard Error
SREG	Site Regression (model)
SS	Sum of Squares
TRET (Ratio)	The ratio of the mean of the five best genotypes and the trial mean
TM	Trial mean
σ	Standard deviation of the population
σ^2	Variance of the population
σ_p	Phenotypic standard deviation
μ	Mean of the population
μ_0	Mean of the new generation
μ_p	Mean of the population selected

List of figures

List of tables

Chapter 1
GENERAL INTRODUCTION

Breeding plays a major role in obtaining adapted and superior genotypes for production. The use of science for selecting and producing genotypes bearing desirable characteristics, based on knowledge about the heredity of such characteristics, is called genetic improvement. Genetic improvement in grain yield is a primary objective of all wheat breeding programmes. Periodic evaluation of breeding progress allows for quantification of the magnitude and rate of genetic change that has been accomplished by breeders over a given period. Yield components, as well as agronomic and morphological traits, should be periodically analysed to evaluate breeding progress and to determine which traits confer the greatest contribution to yield.

Wheat is one of the most important grain crops of South Africa (SA). Compared with the other agricultural crops, wheat is surpassed by only maize in terms of the growing area and its production share. In the 2012/13 season, wheat contributed approximately 11% to the gross value of field crops. The average annual gross value of wheat for the past five years up to 2011/12 is four times less than maize (Department of Agriculture, 2012).

Wheat is planted mainly between mid-April and mid-June in the winter rainfall area, and between mid-May and the end of July in the summer rainfall area. The crop is harvested from November to January. Most of the wheat produced in SA is bread wheat, with small quantities of durum wheat being produced in certain areas. Wheat is generally classed as “hard” or “soft”. Hard wheat tends to have higher protein content than softer wheat and is used mainly for bread. Soft wheat, on the other hand, is more suitable for confectionery (Department of Agriculture, 2012).

The estimated area planted is 511200 hectare, which is 15.5% less than the 604700 hectare of the previous season. The actual production for 2012 was 1.8 million ton. The national bread consumption is estimated at 3.2 million ton (2.8 billion loaves) per annum or approximately 62 loaves per person per annum in 2012 (Department of Agriculture, 2012). This leaves South Africa with a shortfall of 1.3 million ton that has to be imported on a yearly

basis. This number is expected to grow in future as the population and the buying power increases (Pakendorf, 2013).

Consistent wheat production is necessary for food security and is therefore of extremely high agricultural and economic significance. Future production increases depend on the ability to improve, or at least maintain, the rate of increase to feed the population of 52 million (growing with approximately 1 million per year) in SA.

Over the past 30 years there has been an increase in efficiency, productivity and quality due to dedicated scientific inputs from various research disciplines including plant breeding, crop physiology, agronomy and statistics. This manifested in a significant decline in the total area planted, from 1.63 million ha in 1980 to a mere 511 000 ha in 2012.

Wheat yield showed a notable increase over this period, from a mere 1 ton per hectare in 1980 to approximately 3.45 ton per hectare in 2012 (Department of Agriculture, 2012). Although the total area planted decreased, the yield increased, probably due to better planning and planting methods. For example, Pakendorf (1977) documented a yield increase of 2.8% per year over 24 years (1945-1971) in the Western Cape breeding trials; Van Lill and Purchase (1995) reported a yield improvement of 1.35% per year from 1930 to 1994; and Van Niekerk (2001) recorded an incremental yield increase of 1.31% per year from 1979-2001 without the loss of quality or quantity of protein. In her study about the physiological change in wheat, Barnard (2012) reported a 0.1 ton per hectare per year global increase from 1980 to 2011. Although this is a favourable picture, producers are still in need of improved wheat cultivars for the changing climate and demand.

In order to determine the efficiency of breeding programmes, it is important to evaluate the strategy applied and to utilise the available resources better. Thus, before new cultivars are released, researchers need to know whether the newly developed lines are actually genetically more advanced than the existing commercial cultivars and to determine their superiority. Estimates of the progress achieved by breeding programmes are essential tools in quantifying genetic progress made, and thus ascertain how efficient the inputs have been. Various methods exist to quantify this progress, such as comparing historic cultivars with those recently released in specially designed yield trials, and then estimating the progress made. The problem in this case, however, revolves around the availability of seed of historic

cultivars, disease susceptibility of these cultivars or that their straw lengths do not comply with modern fertiliser recommendations. Another method would involve comparing historic data from previous trials, and assessing the progress made over certain periods by employing conventional statistical means. In that case, however, problems revolve around the continuity of cultivars that have been repeated over a period of time and that can act as a basis for comparison (Cargnin *et al*., 2008).

It is therefore necessary to search for alternative methods to monitor genetic progress. An alternative method would make use of the information as it becomes available throughout multi-environment trials, i.e. data from mostly non-recurrent genotypes and localities in the experimental years.

The aim of this study was to provide novel and conventional biometrical or statistical information on wheat yield improvement in the past, and to explore what it may mean for yield improvement in the future. Various biometrical techniques were used to determine the trend in grain yield and two quality traits [hectolitre mass (HLM) and protein content] from 1995-2010 for the three production environments (dryland Western Cape, dryland Free State and the irrigation regions) of the elite and cultivar trials of the Agricultural Research Council: Small Grain Institute (ARC-SGI).

The objectives of the study were to:

i. evaluate the wheat yield improvements achieved over the last 16 years (1995-2010) through the breeding programmes by various biometrical techniques;
ii. determine the trends of yearly yield by regression methods and other biometrical techniques;
iii. demonstrate the direction of yield progress during the last 16 years by different biometrical/statistical techniques;
iv. compare the AMMI and the GGE analyses in assessing Genotype-by-Environment interaction for yield and the two quality traits;
v. study the relationship between wheat grain yield and quality traits by different statistical techniques.

References

Barnard A (2012) Physiological changes in the wheat crop (Part 2). South African Grain 14:32-35

Cargnin A, De Souza MA, Fronza V, Fogaca CM (2008) Genetic and environmental contributions to increased wheat yield in Minas Gerais. Crop Breeding and Applied Biotechnology 8:39-46

Department of Agriculture (2012) Trends in the agriculture sector in 2012. In Agricultural statistics. Pretoria pp 1-111

Pakendorf KW (1977) A study on the efficiency of current methods of breeding and testing for wheat improvement in the Western Cape Province. Dissertation, Stellenbosch University

Pakendorf KW (2013) Withering wheat production in South Africa? South African Grain 20:50-51

Van Lill D, Purchase JL (1995) Directions in breeding for winter wheat yield in South Africa from 1930 to 1990. Euphytica 82:79-85

Van Niekerk HA (2001) The South African wheat pool. In Bonjean A, Angus W (eds) The world wheat book: A history of wheat breeding. Lavoisier, Paris, pp 923-936

Chapter 2
LITERATURE REVIEW

2.1 Introduction

Wheat is by far the biggest winter cereal crop planted in South Africa (SA). Other winter crops are barley and canola. Summer field crops (e.g. maize) are better suited for the SA climatic conditions. The three main wheat production regions are the:

i. Western Cape region (winter rainfall) where spring wheat is planted;
ii. Free State region (summer rainfall) where winter and intermediate wheat are cultivated;
iii. Northern region (mainly irrigation) where spring wheat is grown.

2.2 Concepts

Selection is a process whereby only a portion of the population is chosen due to their superior appearance of a trait. The difference between the population mean (μ) and the mean of the portion selected (μ_p), is called the *selection differential*. The change between the parent mean (μ_p), and the offspring mean (μ_0), is called the *response to selection*. The appearance, the *phenotype,* results from the interaction of the individual's genetic make-up (*genotype*) with its environment. In performance testing, the organism eligible for selection is measured phenotypically for a particular trait. Because organisms have different parents, variation exists between genotypes. Production environments also vary. Therefore, variation in phenotypes for measured traits will be found.

To improve selection accuracy, environmental variation needs to be minimised so that differences between individuals are genetic in nature to the greatest extent possible. The proportion of the phenotypic variation due to genetic variation is called *heritability.* Response to selection (R) is a function of three entities, namely:

i. heritability (h^2),
ii. the standardised selection differential (i), and
iii. phenotypic standard deviation of the parental population σ_p.

The formula is as follows: $R = \mu_0 - \mu_p = ih^2\sigma_p = \Delta G$. This is called Genetic Advance (ΔG). **Figure 2.1** shows it more explicitly.

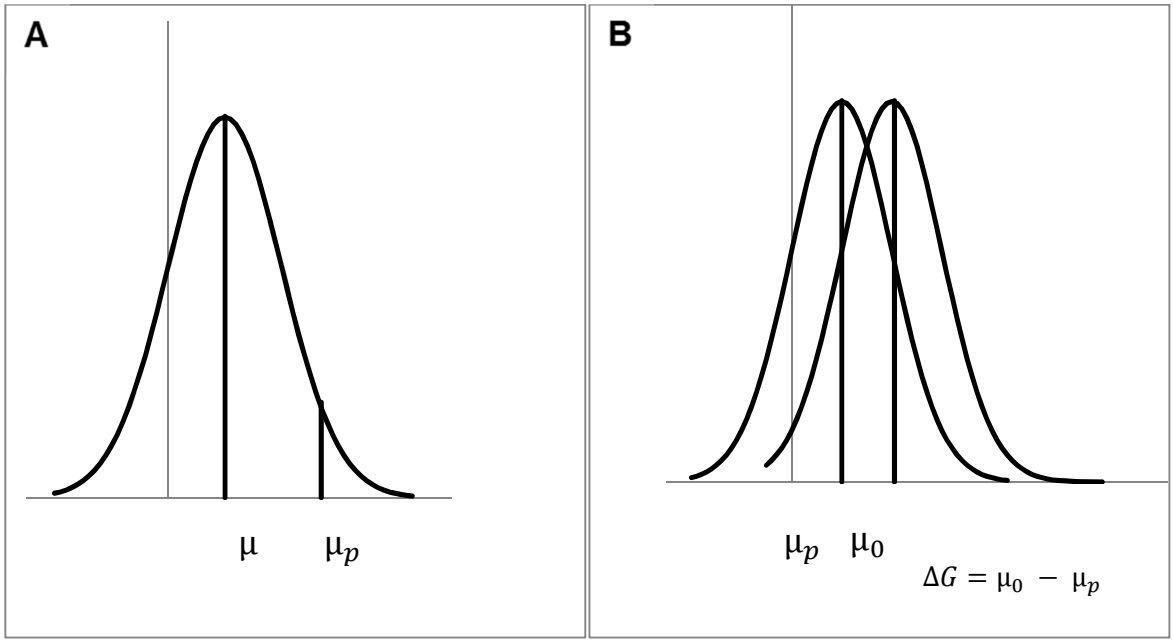


Figure 2.1 Genetic advance from selection indicates the progress made from one generation to another. (A) μ = population mean and μ_p = mean of proportion selected. (B) μ_p = mean of proportion selected and μ_0 = new generation (modified from Figure 8.4 in Acquaah, 2007).

The genetic variance, genetic gain and heritability estimates are of great importance in plant breeding programmes. Plant breeders estimate genetic variances in their populations so that they can predict the response to selection and determine the best selection and breeding procedure for the populations. The magnitude of heritable variability and more particularly its genetic components is clearly the most important aspects of the genetic constitution of the breeding material, which has a close bearing on its response to selection (Falconer and MacKay, 1996). Breeding programmes depend on the knowledge of the inheritance of key traits. These are controlled by genetic and environmental factors that influence their expression.

The progress of a breeding programme is conditioned by the degree and the nature of the genotypic and non-genotypic variation in various characters. Most of the economic characters, e.g. yield, hectolitre mass (HLM) and protein content, are complex in inheritance

and are greatly influenced by various environmental conditions. The study of heritability and genetic advance is very useful in order to estimate the scope for improvement by selection. Heritability levels show the reliability with which the genotype will be recognised by its phenotype expression (Bilgin *et al.*, 2011).

To plan an efficient breeding programme, it is necessary to statistically analyse data from the breeding programme. Analysis of variability among the traits and the association of a particular character to other traits contributing to the yield of a crop would be of great importance in planning a successful breeding programme.

The success of a breeding programme in a certain period can be assessed by the genetic advance observed. Besides quantifying the progress obtained in a certain period, the genetic advance analysis also enables aggregation of other information, such as comparison of the advance obtained with the use of different breeding strategies or in different environments. This kind of information contributes to the understanding of past events. This allows elaboration of new strategies, adoption of corrective methods and more efficient resource allocation. The result is an increase in the breeding programmes' efficacy (Lange and Federizzi, 2009).

Genetic progress can be estimated from the multi-environmental trials (MET) data. These trials are analysed by analysis of variance (ANOVA) that combines data of years, localities and genotypes. The combined ANOVA provides a system of assessment for the performance of genotypes over a range of environments (years, localities or both).

An important effect of genotype-by-environment interaction (GEI) is the reduction of the correlation between phenotype and genotype. Comstock and Moll (1963) have statistically shown the effect of large GEIs in reducing progress from selection. Genotype selection is more effective if there is consistency in yield of the best selection over a wide range of environments. However, it is hardly the case in numerous METs. GEI becomes the rule rather than the exception.

Two major statistical methods have been proposed to resolve the problems associated with GEIs. The most widely used technique in reducing GEIs is the regression method proposed by Yates and Cochran (1938) and amplified by Finlay and Wilkinson (1963). The second

method is the optimisation of blocks, localities, years and genotypes in determining variance components (Punto and Lantican, 1983). These two methods can also be used to calculate genetic gain. Both methods are subjected to certain assumptions.

These assumptions are:

i. error terms are randomly, independently and normally distributed;
ii. the variances of different samples are homogeneous.

2.3 Normality of residuals, deviations from and solutions

True normality is exceedingly rare in field trials. There are multiple options for dealing with non-normal data. Replacing outliers by the predicted values of the model fitted and transformation of the data are the most commonly used techniques. Three of the most common data transformations utilised for improving normality are square root, logarithmic and inverse transformations.

2.4 Homogeneity of variances versus heterogeneity of variances

Statistical heterogeneity manifests itself in the observed main or interaction effects being more different from each other than one would expect due to random error (chance) alone. Comstock and Moll (1963) stated that they "know from experience that the plot error variance is variable from one experiment to another … and there is nothing that compels the variances of the GEI effects to be homogeneous".

2.4.1 Implications of heterogeneity of variances in MET

According to Edwards (2007) the implications are that discrepancies in the heritability estimates (more accurately repeatibility) of observed data may occur. Another concern, however, is that genotype responses may be distorted by the influence of environments with less precise trials (Crossa, 1990). Furthermore, the estimation of *MSE* values may be inaccurate, suggesting more complex approaches based on their prediction as a function of external variables (Frensham *et al.*, 1998).

2.4.2 Solutions to heterogeneity

The primary method used to analyse MET is based on the ANOVA, which is a fixed effects model and requires homogenous variance-covariance of data.

Cochran (1937) proposed a weighted analysis of variance. Annicchiarico (2002) and Morris *et al*. (2004) found this type of analysis effective to contest heterogeneity of environments (years, localities or both). Evidence in the literature was found where logarithm transformation was used to remedy heterogeneity of variances (Sener *et al*., 2009; De Vita *et al*., 2010).

Hu and Spilke (2011) discussed several linear mixed models with different variance-covariance structures. They concluded that the problem of how the models should be assessed and which model is more suitable for a given trial's data has not yet been solved.

2.5 Unbalanced data

METs generally have highly unbalanced data structures in which a specific genotype is only observed in a subset of all environments for which data are available. Many statistical methods for analyses of MET data have been proposed that do not depend on balanced data. Various methods rely on various assumptions and variance-covariance structures in the data. Hence, the choice of the best model depends on the particular structure and statistical properties of the data and a statistical approach to select the best model (Smith *et al.,* 2005). Although a great deal has been written about analysis of METs from a theoretical perspective, very little has been done to compare broad classes of models empirically.

The choice of appropriate models depends on understanding the complexities in the data rather than the unbalanced nature of the data (So, 2009). For example, if data within regions has little heterogeneity of variances or covariances, methods for subdivided target regions could be applied without modeling heterogeneity of covariance structure within individual regions or localities (Piepho and Mohring, 2007).

2.6 Best linear unbiased predictors (BLUP)

The normal formula for genotype means is:

$$\bar{y}i = 1/n \sum_{j=i}^{N} yij$$

Genotype effect, either fixed or random, is more descriptive than a genotype mean. Best Linear Unbiased Estimator (**BLUE**) is a formula that corresponds to the 'fixed effect' analysis that is calculated in the same way as the genotype effect and is represented by the following formula:

$$ti = 1/n \sum_{j=i}^{N} (y_{ij} - \bar{y}_i)$$

This formula accurately represents what happened in a trial, but it is not the best *predictor* of what might happen if the trial is repeated. For this the alternative is the Best Linear Unbiased Predictor (**BLUP**) given by:

$$Ti = \sum_{j=1}^{N} \left(y_{ij} - \bar{y}_i\right) (n + \sigma_G^2/\sigma_E^2)$$

The effect of adding the variance ratio into the denominator is to shrink the genotype effect. If genotype variance (σ^2_G) approaches 0, the ratio (σ^2_G/σ^2_E) approaches infinity causing the predicted genotype effect to shrink to 0 (Gilmore, 2010).

2.7 Linear fixed effects model versus linear mixed effects model

2.7.1 Linear fixed effects model

The primary method used to analyse METs is based on the ANOVA, which is a fixed effects model and requires homogenous variance-covariance of data.

2.7.2 Linear mixed effects model

In the mixed model analysis for MET data, there has been two views as to the classification of genotype effects being random or fixed. Smith *et al.* (2005) and Piepho *et al.* (2008) stated that it depended on the objectives of the research.

The general form of a linear mixed model is:

$$Y=X\beta + Z\mu + e$$

where: Y is the response vector (data),
X and Z are known design matrices,
ß is a vector of fixed parameters,
u is random effects, and
e (error terms) are unobservable random vectors.

The E(u) and E(e) are assumed to be zero. Assumptions regarding the structure of G (the variance-covariance matrix of the random effects in u) and R (the variance-covariance matrix of the random effects in e) will be defined for a particular mixed model. Different models for the variance-covariance of the data, V = ZGZ' + R, are obtained by specifying the structure of Z, G and R. The simplest form for G and R is one that arises from the independence in random effects and error terms. Independence in the random term effects does not imply that the observations are independent. On the contrary, one sets up a common correlation among all observations having the same level of u. Laird and Ware (1982) considered the unstructured model for a covariance matrix, i.e. the more general case where all elements of the matrix are allowed to be different. Intermediate structures for G and R were more efficient in plant breeding. These allow for modelling correlations with a smaller number of covariance parameters than the unstructured one. In general, genetic correlations may be introduced into the model through G and experimental correlations among observations may be modelled by the off-diagonal elements of R (Balzarini, 2002).

Mixed model solutions can be written as:

$$\beta = (X'V^{-1}X)^{-1} X'V^{-1} y \text{ and } u = GZ'V^{-1}(y-XB)$$

2.7.3 Model selection

Several variance-covariance structures are available and were discussed in Hu and Spilke (2011) in more detail. To select an appropriate model, there are two main criteria to consider. The first criterion is:

$$\text{Akaike Information Criterion (AIC)} = -2LL + 2 \times q$$

where: LL denotes the log maximum likelihood of the related model, and
q is the number of parameters of the variance-covariance structure.

The calculation formulae of the information criteria are given in such a way that the model with the lower value of the information criterion is preferred.

The second criterion is:

$$\text{Schwarz Bayesian Information Criterion (BIC or SIC)} = -2LL + \log(N) \times q$$

where: LL denotes the log maximum likelihood of the related model
N is the total number of observations, and
q is the number of parameters in the variance-covariance matrix.

The best model is again the model with the lower value of the information criterion. The model with the smallest AIC and BIC in the METs is the best model (SAS Institute, 2012).

2.8 Linear regression

The popularity of estimating genetic advance (genetic improvement) over years using linear regression is escalating. Most studies reported genetic improvement from linear regression over years. Rodrigues *et al.* (2007) and Sener *et al.* (2009) reported genetic improvement from non-linear regression over years.

Two approaches, the year of release and the experimental years, using linear regression for estimating genetic improvement in the literature were found.

2.8.1 Year of release

This method consists of experiments where newer genotypes are compared to older genotypes in a given period of time. The research by Calderini *et al.* (1995), Giunta *et al*. (2007) and De Vita *et al.* (2010), among others, used this method.

2.8.2 Experimental years

This approach is based on a long time series of yield data mostly coming from long-term experiments compared to a historic check genotype. Within this category there are two methods:

i. This method compares the genotypes to one or more check genotypes that are consistent for the localities and years. A number of studies used this method, e.g. Graybosch and Peterson (2010), Green *et al*. (2012) and Sharma *et al.* (2012).

ii. Trethowan *et al.* (2002) proposed a method where the mean of the five best genotypes from a trial is expressed as a ratio of the trial mean (% TM). The % TM, the trial mean and the mean of the check or checks are then regressed against experimental years. The % TM combats the fluctuation in years.

2.9 Interpretation of the slope of linear regression

2.9.1 Definition of the slope

The slope in a regression model is often considered to be of great interest as it conveys how quickly the dependent variable (i.e. yield) changes in relation to the independent variable (i.e. experimental years), and as such determines whether a regression line is useful.

2.9.2 General interpretation of the slope to determine genetic improvement

Most studies reported genetic improvement from linear regression over years by dividing the slope (b) value by the number of experimental years and expressed genetic advance as a percentage.

2.9.3 General interpretation of the ratio to determine genetic improvement

The genetic advance is equal to the b-value of the % TM proposed by Trethowan *et al*. (2002).

2.10 Variance components models

Selection among genotypes is based on phenotypic variation, but the response to selection is a function of genetic variability. The prediction of genetic advance from selection depends on the proportion of phenotypic variance which is due to the genetic variance – this ratio is called heritability (Cooper and Hammer, 1996).

Restricted (or residual) maximum likelihood estimation (REML) has been the preferred method in estimation of variance components in MET data analyses especially with unbalanced data sets (Piepho and Mohring, 2007). Other approaches to estimate variance components were the Best Linear Unbiased Prediction (BLUP) (Cullis *et al.,* 2006; Piepho *et al*., 2008) and a bayesian estimation by Edwards and Jannink (2006). So (2009) concluded that the choice of the model and method depended on the data set. Similar conclusions were drawn by Hu and Spilke (2011).

The main purpose of estimating heritability and genetic parameters is to determine genetic gain from selection based on different selection strategies. It is preferable that heritability estimates are made from data collected in multiple localities and during multiple years (Holland *et al*., 2003).

The sources of variation would be partitioned in years, localities, replications within years and localities, genotypes and the interactions of genotypes, years and localities. The statistical model is:

$$Y_{ijkl} = \mu + Y_i + L_j + YL_{ij} + B(YL_{ijk}) + G_k + GY_{ik} + GL_{ik} + GYL_{ijk} + €_{ijkl}$$

where:
- Y_{ijkl} = observed yield or HLM or protein content value (depending on the elite or cultivar trials)
- μ = general mean
- Y_i = effect of the year
- L_j = effect of the locality
- YL_{ij} = interaction effect of the year and locality
- $B(YL_{ijk})$ = effect of block within year and locality

G_k = effect of genotype

GY_{ik} = interaction effect of the genotype and year effect

GL_{jk} = interaction effect of the genotype and locality

GYL_{ijk} = interaction effect of the genotype, year and locality

$€_{ijkl}$ = error or residual effect

$€_{ijkl} \sim NID(0,\sigma^2)$

2.10.1 Model proposed by Comstock and Moll (1963)

Comstock and Moll (1963) proposed a linear fixed effects model to optimise and determine the relationship of number of genotypes, blocks, localities and years in formulating genetic advance. In this model (**Table 2.1**) the optimum allocations of components can be determined from ratios of the genetic component of variance, the interaction components of genotypes x years, genotypes x localities and genotypes x years x localities to the error variance component.

Table 2.1 Sources of variation, calculated mean squares (MS) and their expected values

Source of variation	**MS**	**E(MS)**
Years (Y)		
Localities (L)		
Yx L		
B x Yx L		
Genotypes (G)	M1	$\sigma^2_E + r\sigma^2_{GLY} + rs\sigma^2_{GY} + rt\sigma^2_{GL} + rst\sigma^2_G$
G x Y	M2	$\sigma^2_E + r\sigma^2_{GLY} + rs\sigma^2_{GY}$
G x L	M3	$\sigma^2_E + r\sigma^2_{GLY} + rt\sigma^2_{GL}$
G x L x Y	M4	$\sigma^2_E + r\sigma^2_{GLY}$
Error	M5	σ^2_E
Corrected total		

The formula to determine genetic advance (ΔG) is as follows:

$$\Delta G = \frac{d\,\bar{x}m}{\sqrt{\frac{1}{rst} + \frac{e}{st} + \frac{by}{t} + \frac{bp}{s} + d}}$$

where: ΔG = genetic advance

d = M1 /M5

$\bar{x}m$ = standardised selection index for 1/(nr of genotypes)

e = M5

by = M2/M5

bp = M3/M5

r = replication of each s localities in t years (blocks)

s = number of localities

t = number of years

2.10.2 Model proposed by Allard (1960)

Assuming normally distributed residuals of the variable in question, genetic advance under selection can, according to Allard (1960), be calculated by :

$$\Delta G = ih^2\sigma_p$$

where: i = standard selection differential from a set of genotypes at 10% selection intensity

h^2 = broad sense heritability

σ_p = phenotypic standard deviation.

In terms of the model by Comstock and Moll (1963), broad sense heritability can be defined by:

$$h^2 = \sigma^2_G / \sigma^2_p$$

where:

$\sigma^2_p = (\sigma^2_G - \sigma^2_{GY} - \sigma^2_{GL} + \sigma^2_{GLY})/rst$

σ^2_G = variance of the genotypes

σ^2_{GY} = variance of the interaction of genotype by year

σ^2_{GL} = variance of the interaction between genotypes and localities

σ^2_{GLY} = variance of the interaction, among genotypes, years and localities

r = number of blocks

s = number of localities

t = number of years.

2.11 Interpretation of the Genetic Advance (ΔG) estimate

In the following chapters the estimates of Genetic Advance (GA) – more accurately described as the change in Genetic Advance (ΔG) – were given as percentage advance in the comparison tables of the research chapters (e.g. Table 3.3).

2.12 Correlation

The Pearson product-moment correlation coefficient measures the strength of the linear relationship between two variables. For response variables *X* and *Y*, it is denoted as r_{xy} and computed as:

$$r_{xy} = \frac{\sum_{i=1}^{n}\left(\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}}$$

where

r_{xy} = Pearson product-moment correlation coefficient between two variables x and y.

x_i = i th factor (genotype, locality or year) for variable x and $\bar{x}$ =mean of the factor, similar for variable y.

If there is an exact linear relationship between two variables, the correlation is 1 or –1, depending on whether the variables are positively or negatively related. If there is no linear relationship, the correlation tends toward zero.

2.13 Multivariate techniques

Multivariate methods have some advantages including deletion of noise from the data pattern, summarising the dataset, and revelation of the data structure (Crossa, 1990). In contrast with conventional (univariate parametric and non-parametric) statistical strategies, the function of multivariate analysis is to elucidate the internal structure of the data from which hypotheses can be produced and tested by statistical procedures (Gauch and Zobel, 1996).

Multivariate statistical methods are appropriate for analysing two-way layouts of genotypes and environments in multi-environment trials. The response of a special genotype in various test environments may be conceived as a pattern in multi-dimensional space, with the coordinates of an individual axis being that of yield or another trait. Additive main effects and multiplicative interaction model (AMMI), site regression also known as the genotype plus genotype-by-environment biplot (GGE) analyses, cluster analysis, principal component analysis and linear discriminant (canonical variate) analysis are the most commonly multivariate statistical methods used to investigate GEI.

A more comprehensive discussion of these techniques (except for the AMMI and GGE biplot analyses) is available in Rencher (2002).

2.13.1 Principal component analysis

Principal component analysis (PCA) is a multivariate technique statistical method to identify data patterns as well as similarities and dissimilarities among observations and variables. It uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables.

This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e. uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Rencher (2002) recommended that a correlation matrix should be used to standardise the data. PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualised as a set of coordinates in a high-dimensional data space, PCA can supply the user with a lower-dimensional picture, a "shadow" of this object when viewed from its most informative viewpoint. This is done by

using only the first few principal components so that the dimensionality of the transformed data is reduced.

2.13.2 AMMI analysis versus GGE biplot analysis

In breeding programmes, genotype-by-environment interaction (GEI) causes many difficulties, whereas the environmental factors such as temperature and drought stress affect the performance of genotypes. GEI reduces the genetic progress in plant breeding programs through minimizing the association between phenotypic and genotypic values (Comstock and Moll, 1963). Multi-environment trails (MET) are essential in estimation of GEI and identification of superior genotypes in the final selection cycles (Kaya *et al*., 2006; Mitrovic *et al*., 2012).

Accordingly, statistical methods for effective analysis of MET have received considerable development and discussion. Two frequently used models of statistical analyses have been the AMMI model and the GGE model.

2.13.2.1 AMMI analysis

The AMMI model is used to simplify the complicated GEI of multi-environmental trial analysis. The AMMI model combines regular analysis of variance for additive effects with PCA for multiplicative structure within the interaction. AMMI also provides a visual representation of patterns in the data through a biplot that makes use of the first interaction principal component axis (IPCA1) and the mean yields of both the genotypes and environments. The AMMI model is used in research to evaluate a number of genotypes established in a number of environments, identify stable and adaptable genotypes, determine the magnitude of GEI, and identify factors contributing to the GEI pattern. An ANOVA will show that the effects of environments, genotypes and GEI were highly significant ($p< 0.0001$) for a variable (i.e. yield). AMMI estimates ranked genotypes differently from unadjusted means producing sharper and more stratified rankings. The AMMI stability value (ASV), was developed by Purchase *et al*. (2000), and is a single value which ranks the genotypes or environments for stability.

2.13.2.2 GGE biplot analysis

The GGE refers to the genotype main effect (G) plus the GEI, which are the two sources of variation of the site regression (SREG) model. The term “GGE” emphasises that G and GEI are the two sources of variation that are relevant to genotype evaluation and must be considered simultaneously for appropriate genotype and test environment evaluation. GGE biplot analysis has evolved into a comprehensive analysis system whereby most questions that may be asked of a genotype by environment table can be graphically addressed.

2.13.2.3 Comparison of AMMI analysis and GGE biplot

A body of literature has been developed to demonstrate the effectiveness of the AMMI and the GGE biplot analyses. A comprehensive study of the two models is portrayed by Yan *et al.* (2007) and Gauch *et al.* (2008). The AMMI model is represented by:

$$\bar{y}_{ij.} = \mu + \tau_i + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij.}$$

The GGE model is given by:

$$\bar{y}_{ij.} = \mu + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij.}$$

where: $\bar{y}_{ij.}$ is the mean of the i[th] cultivar in the j[th] environments

μ is the overall mean

τ_i is the genotypic effect (separate effect in AMMI not in GGE)

δ_j is the environment effect

λ_k ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_t$) are scaling constants (singular values) that allow the imposition of ortho-normality constraints on the singular vectors for cultivars, $\boldsymbol{\alpha}_{ik}=(\alpha_{1k},\ldots,\alpha_{gk})$, and environments, $\boldsymbol{\gamma}_{jk}=(\gamma_{1k},\ldots,\gamma_{ek})$, such that

$\sum_i \alpha_{ik}^2 = \sum_j \gamma_{jk}^2 = 1$ and $\sum_i \alpha_{ik}\alpha_{ik'} = \sum_j \gamma_{jk}\gamma_{jk'} = 0$ for k≠k′;

α_{ik} and γ_{jk} for k=1,2,3,… are called “primary, secondary, tertiary” etc. effects of cultivars and sites, respectively

$\bar{\varepsilon}_{ij.}$ is the residual error assumed to be NID (0, σ^2/r) (where σ^2 is the pooled error variance and *r* is the number of replicates)

Least squares estimates of the multiplicative (bilinear) parameters in the k^{th} bilinear term are obtained as the k^{th} component of the deviations from the additive (linear) part of the model. In the AMMI model, only the GEI term is absorbed in the bilinear terms, whereas in the SREG model, the main effects of cultivars (G) plus the GEI are absorbed into the bilinear terms.

A number of multi-environment trial studies evaluate the AMMI and GGE biplot analyses. In maize, Kandus *et al*. (2010) found the AMMI model was the best model to describe the GEI. Stojaković *et al.* (2010) and Mitrovic *et al.* (2012) found the models provided similar results. Using data of bread wheat, Rad *et al*. (2013) indicated that both models performed equally on the objectives to be answered. Samonte *et al*. (2005) found the AMMI and GGE biplot analyses complemented one another.

2.13.3 Shifted Multiplicative model

Shifted multiplicative model (SHMM) is developed by Seyedsadr and Cornelius (1992). It is a tool to analyse the separability of genotypic effects from environment effects. The requirements for SHMM are non cross-over genotype-environment interaction (COI). Crossa *et al.* (1995) used the SHMM model for clustering five irrigation levels in two years (10 environments) and results were compared with the conventional cluster analysis using the Euclidean distance as the criterion. The SHMM clustering strategy formed more homogeneous non-COI subsets of sites than the conventional clustering.

2.13.4 Cluster analysis

Cluster analysis is a numerical classification technique that defines clusters of individuals, and may be defined as either hierarchical or non-hierarchical. In hierarchical methods the individuals are organised into a hierarchy where individuals or groups are fused one at a time to individuals or groups with the most similar patterns across all environments. In non-

hierarchical systems the individuals are organised into a set number of groups in the best possible manner (De Lacy *et al*., 1996). With hierarchical methods, the process usually starts with a dissimilarity matrix where two individuals with the smallest dissimilarity between them are fused into a group. The dissimilarities between this group and all remaining individuals is then calculated and added to the matrix of dissimilarities among the remaining individuals to form a new matrix. The procedure is then repeated continually with group-group dissimilarities being calculated thereafter. After intensive calculations, the structure of the groupings is usually represented by dendrogrammes that depict composition of groups and the degree of dissimilarity among groups. Cluster analysis, based on differences in genotype responses across test environments, is the most commonly used multivariate method. Abou-El-Fittouh *et al*. (1969) proposed cluster analysis as a technique to classify test environments for cotton. Leilah and Al-Khateeb (2005), Kaya *et al.* (2006) and Sabaghnia *et al.* (2012) performed cluster analysis to classify genotypes and test environments.

2.13.5 Linear discriminant analysis

Linear discriminant analysis or canonical variate analysis is a multivariate technique that shows differences among groups rather than among individuals. This approach allows the explanation of relationships in a data set among two or more large set of variables. The variables are reduced to a smaller set of variates that account for most of the variation in the data set. These new sets of variables are linear combinations of the original variables, and are thus indicated as vectors of loading for the original variables. A set of directions is obtained in such a way that the ratio of the between-group variability to the within-group variability in each direction is maximised. The horizontal separation (on the x-axis) represents the one set of canonical variates (CV1) and accounts for most of the variation. The vertical separation represents the second set of canonical variates (CV2) and explains the second largest variation in a data set.

References

Abou-El-Fittouh HA, Rawlings JO, Miller PA (1969) Classification of environments to control genotype by environment interactions with and application to cotton. Crop Science 9:135-140

Acquaah G (2007) Principles of plant genetics and plant breeding. Blackwell, Oxford

Allard RW (1960) Principles of plant breeding. Wiley, New York

Annicchiarico P (2002) Genotype x environment interactions: Challenges and opportunities for plant breeding and cultivar recommendations. FAO Plant production and protection papers. http://www.fao.org/DOCREP/005/Y4391E/y4391e05.htm. Accessed 13 June 2010

Balzarini M (2002) Applications of mixed models in plant breeding. In Kang M (ed) Quantitative genetics, genomics and plant breeding. CABI Publications, New York, pp 353-364

Bilgin O, Korkut KZ, Baser I, Daglioglu O, Ozturk I, Kahraman T, Balkan A (2011) Genetic variation and interrelationship of morpho-physiological traits in durum wheat (*Triticum durum* (L.) Desf). Pakistan Journal of Botany 43:253-260

Calderini DF, Dreccer MF, Slafer GA (1995) Genetic improvement in wheat yield and associated traits: A re-examination of previous results and the latest trends. Plant Breeding 114:108-112

Cochran WG (1937) Problems arising in the analysis of a series of similar experiments. Journal of the Royal Statistical Society 1:102-118

Comstock RE, Moll RH (1963) Genotype-environment interactions. In Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. National Academy of Science, Washington, pp 164-196

Cooper M, Hammer G (1996) Plant adaptation and crop improvement. CABI Publications, Oxford

Crossa J (1990) Statistical analysis of multilocation trials. Advances in Agronomy 44:55-85

Crossa J, Cornelius PL, Sayre K, Ortiz-Monasterio JI (1995) A Shifted multiplicative model fusion method for grouping environments without cultivar rank change. Crop Science 35:54-62

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. Journal of Agricultural, Biological and Environmental Statistics 11:381–393

DeLacy IH, Basford KE, Cooper M, Bull JK, McLaren CG (1996) Analysis of multi-environment trials – An historical perspective. In Cooper M and Hammer GL (eds) Plant Adaptation and crop improvement. CABI Publications, Wallingford, pp 39-124

De Vita P, Mastrangelo AM, Matteu L, Mazzucotelli E, Virzi N, Palumbo M, Lo Storto M, Rizzi F, Cattivelli L (2010) Genetic improvement effects on yield stability in durum wheat genotypes grown in Italy. Field Crops Research 119:68-77

Edwards JW (2007) Estimating genotype- and environment-specific heritabilities. In Illinois Corn Breeders School Proceedings. pp 228-241

Edwards JW, Jannink JL (2006) Bayesian modeling of heterogeneous error and genotype x environment variances. Crop Science 46:820-833

Falconer DS, MacKay TFC (1996) Introduction to quantitative genetics. Longman, London.

Finlay KW, Wilkinson GN (1963) The analysis of adaption in a plant breeding programme. Australian Journal of Agricultural Research 14:742-754

Frensham AB, Barr AR, Cullis BR, Pelham SD (1998) A mixed model analysis of 10 years of oat evaluation data: use of agronomic information to explain genotype by environment interaction. Euphytica 99:43-56

Gauch HG, Piepho HP, Annicchiarico P (2008) Statistical analysis of yield trials by AMMI and GGE: further considerations. Crop Science 48:866-889

Gauch HG, Zobel RW (1996) AMMI analysis of yield trials. In Kang MS, Gauch HG (eds) Genotype by environment interaction. Boca Raton CRC, New York, pp 85-122

Gilmore A (2010) Why use BLUPS? An introduction to plant breeders. http://mmontap.org/sites/default/files/workingPapers/Blup.pdf. Accessed 2 February 2012

Giunta F, Motzo R, Pruneddo G (2007) Trends since 1900 in yield potential of Italian-bred durum wheat cultivars. European Journal of Agronomy 27:12-24

Graybosch RA, Peterson CJ (2010) Genetic improvement in winter wheat yields in the Great Plains of North America, 1959-2008. Crop Science 50:1882-1890

Green AJ, Berger G, Griffey CA, Pitman R, Thomason W, Balota M, Ahmed A (2012) Genetic yield imrovement in soft red winter wheat in the eastern United States from 1919 to 2009. Crop Science 52:2097-2108

Holland JB, Nyquist WE, Cervante-Martinez CT (2003) Estimating and interpreting heritability for plant breeders: An update. Plant Breeding Reviews 22:9-112

Hu X, Spilke J (2011) Variance-covariance structure and its influence on variety assessment in regional crop trials. Field Crops Research 120:1-8

Kandus M, Almorza D, Boggio Ronceros R, Salerno JC (2010) Statistical methods for evaluating the genotype-environment interation in maize (*Zea mays* L.). PhyYld 39-46

Kaya Y, Akcura M, Taner S (2006) GGE biplot analysis of multi environment yield trials in bread wheat. Turkish Journal for Agriculture 30:325-337

Lange CE, Federizzi LC (2009) Estimation of soybean genetic progress in the south of Brazil using multi-environmental yield trials. Scientia Agricola (Piracicaba, Braz.) 66:309-316

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics 38:963-974

Leilah AA, Al-Khateeb SA (2005) Statistical analysis of wheat yield under drought conditions. Journal of Arid Environments 61:483-496

Mitrovic B, Stanisavljevi D, Treski S, Stojakovic M, Ivanovic M, Bekavac G, Rajkovic M (2012) Evaluation of experimental maize hybrids tested in multi-location trials using AMMI and GGE biplot analyses. Turkish Journal of Field Crops 17:35-40

Morris CF, Campbell KG, King GE (2004) Characterization of the end-use quality of soft wheat cultivars from the eastern and western US germplasm pools. Plant Genetic Resources 2:59-69

Piepho HP, Mohring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177:1881-1888

Piepho HP, Mohring J, Melchinger AE, Buchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209-228

Punto NG, Lantican RM (1983) Genotype x environment interaction for yield in field legumes: I. Peanut *Arachis hypogaea* (L.). Crop Science Society of the Phillipines 1:94-100

Purchase JL, Hatting H, Van Deventer CS (2000) Genotype x environment interaction of winter wheat (*Triticum aestivum* L.) in South Africa II: Stability analysis of performance. South African Journal of Plant and Soil 17:101-107

Rad NM, Kadir MA, Rafii MY, Jaafar HZ, Naghavi MR, Ahmadi F (2013) Genotype x environment interaction by AMMI and GGE biplot analysis in three consecutive generations of wheat (*Triticum aestivum*) under normal and drought stress conditions. Australian Journal of Crop Science 7:956-961

Rencher AC (2002) Methods of Multivariate Analysis. 2nd ed. John Wiley, New York

Rodrigues O, Lhamby JCB, Didonet AD, Marchese JA (2007) Fifty years of wheat breeding in Southern Brazil: yield improvement and associated changes**.** Pesquisa Agropecuária Brasileira 42:817-825

Sabaghnia N, Mohammadi M, Karimizadeh R (2012) Interpretation of genotype x environment interaction in multi environment trials of bread wheat using cluster analysis. Natura montenegrina, Podgorica 11:511-523

Samonte SOPB, Wilson LT, McClung AM, Medley JC (2005) Targeting cultivars onto rice growing environments using AMMI and SREG GGE biplot analyses. Crop Science 45:2414-2424

SAS Institute (2012). The Statistical Procedure manual. North Carolina 27513: SAS Campus Drive, Cary.

Sener O, Arslan M, Soysal Y, Erayman M (2009) Estimates of relative yield potential and genetic improvement of wheat cultivars in the Meditterean region. Journal of Agricultural Science 147:323-332

Seyedsadr M and Cornelius PL (1992) Shifted multiplicative model for nonadditive two-way tables.Communications in Statistics Part B Simulation and Computation 21:807-822

Sharma RC, Crossa J, Velu G, Huerta-Espino J, Verga M, Payne TS, Singh RP (2012) Genetic gains for grain yield in CIMMYT spring bread wheat across international environments. Crop Science 52:1522-1533

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. Journal of Agricultural Science 143:449-462

So Y (2009) Prediction of cultivar performance and heterogeneity of genotype variance, correlation and error variance in Iowa crop performance test in corn (*Zea mays* L.). Dissertation, Iowa State University

Stojaković M, Ivanović M, Jocković Đ, Bekavac G, Purar B, Nastasić A, Stanisavljević D, Mitrović B, Treskić S, Laišić R (2010) NS maize hybrids in production regions of Serbia. Field and Vegetable Crops Research 47:93-02

Trethowan RM, Van Ginkel M, Rajaram S (2002) Progress in breeding wheat for yield and adaptation iun global drought affected environments. Crop Science 42:1441-1446

Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007) GGE biplot analysis of genotype-by-environment data. Crop Science 47:643-655

Yates F, Cochran WG (1938) The analysis of groups of experiments. Journal of Agricultural Science 28:556-580

Chapter 3

THE DRYLAND WESTERN CAPE

Abstract

Spring bread wheat genotypes were evaluated for genetic improvement under dryland conditions over a 16 year period (1995-2010) in nine localities in the Rûens and Swartland wheat production regions of the Western Cape of South Africa (SA). The following statistical analyses and procedures were conducted to determine genetic improvement: (i) linear regression; (ii) a variance component method according to Comstock and Moll (iii) a variance component method according to Allard. The proposed method for determining genetic advance over time was Ratio (TRET). The estimated genetic gain in yield potential since 1995 was 1% per year for both regions of the elite trials. No yield increase was observed in the cultivar trials. Genetic advance of 1% per year was observed for protein content for both regions of the cultivar trials. The additive main effects and multiplicative interaction (AMMI) and genotype plus GEI (GGE biplot) analysis were compared for studying GEI. Application of AMMI and GGE biplots facilitated visual comparison and identification of superior genotypes for each target set of environments. The complementary nature of the models, as shown in the biplots, is a benefit. Principal component analysis (PCA) and discriminant analysis (DA) provided additional useful information.

3.1 Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most important grain crops of South Africa (SA). Approximately 2 million ton of wheat are produced on approximately 643 000 ha, of which 37% is derived from the dryland winter rainfall area of the Western Cape.

Dryland wheat production in the Western Cape of SA accounts for 37% of the mean national wheat crop of 1.8 million metric ton (Department of Agriculture, 2012). However, this contribution to wheat production in SA varies considerably due to the unpredictable nature of production (Tolmay, 2008).

Wheat yield has shown a notable increase during the 20th century. The global average increase during the past 50 years of 2.5-fold is attributed to the adoption of both better crop management practices and cultivars with higher yield. Nitrogen use efficiency was higher in new varieties (Sener *et al.,* 2009). However, they show lower protein concentration and reduced quality characteristics (De Vita *et al.*, 2010).

Investigations into the effect of the different factors on wheat grain yield focus especially on the determination of the intensity of the effect of values of certain yield components. Numerous yield structure elements determine the yield. According to Koekemoer (2003) spike number, kernel number per spike, thousand kernel weight and harvest index are generally the important ones.

Genetic improvement in wheat yield can be attributed to selection for improved plant characteristics concerning either higher yield potential or increased resistance to biotic and abiotic stresses or a combination of both (Smit *et al*., 2010). Among yield components Koekemoer (2003) suggested that kernels per square meter and kernel weight will be the most determining factors.

Yield gains for spring wheat were reported for a number of countries. DePauw *et al*. (2007) summarised the studies with the Canadian western spring wheat class. The genetic gains during 1984-2001 were much higher at 0.74% per year compared to 1908-1986 with 0.33% per year, which is attributed to more intensive breeding programmes. Morgounov *et al.* (2010) reported 0.70% increase per year in the course of 100 years in western Siberia. Research done on spring wheat in Finland demonstrated consistent genetic gains of 1.03% during 1970-2005 (Peltonen-Sainio *et al*., 2009). In Turkey, Sener *et al*. (2009) reported 2.6% gain per year over a 80 year period. A study with hard red spring wheat in north Dakota, USA, showed genetic improvement of 1.30% per year during 1968-2006 (Underdahl *et al.*, 2008). The elite spring wheat trials of the International Maize and Wheat Improvement (CIMMYT) showed an increase of 0.26% and 0.49% per year in Mexico (Trethowan *et al*., 2002).

In SA Pakendorf (1977) documented a yield increase of 2.6% per year over 26 years (1945-1971) in the Western Cape breeding trials. But is this the case for the past 16 years (1995-2011)?

To best estimate the effects of improved cultivars, is to have insight into genetic improvement and yield stability. The objectives of the present study were thus to:

i. evaluate the wheat yield improvements achieved over the last 16 years (1995-2010) through the breeding programmes by various biometrical techniques;
ii. determine the trends of yearly yield by regression methods and other biometrical techniques;
iii. demonstrate the direction of yield progress during the last 16 years by different biometrical/statistical techniques;
iv. compare the AMMI and the GGE analyses in assessing genotype-by-environment interaction for yield and two quality traits;
v. study the relationship between wheat grain yield and the quality traits by different statistical techniques.

In order to achieve this objective for the Mediterranean type of climate, breeding and cultivar results from the Agricultural Research Council Small Grain Institute (ARC-SGI) cultivar development programmes from the two distinct Western Cape production regions were used. These trials ranged from 1995-2011. The production regions were the Rûens and Swartland regions.

3.2 Materials

3.2.1 Elite field trials

Grain yield data for wheat genotypes of the ARC-SGI elite spring wheat yield trials (Western Cape), which comprised 225 genotypes tested in 11 localities in the southern Cape region of the Western Cape and nine localities during the 1995-2011 period, were collected. In this study three localities in the Rûens and three localities in the Swartland regions were chosen on the basis of continuity of the data for the period. Two trials were used: one sprayed with fungicide to control airborne leaf disease (B) and the other unsprayed (A) at Tygerhoek in the Rûens and Langgewens in the Swartland. These trials were perceived as different localities (**Table 3.1**).The geographical position of the localities is depicted in the map (**Figure 3.1**). The experimental lay-out was a randomised complete block design with four replications. Each trial included the same ARC-SGI check.

Table 3.1 Listings of elite and cultivar trial locations used in the study

Elite trials		Cultivar trials	
Rûens	**Swartland**	**Rûens**	**Swartland**
Bredasdorp	Langgewens A	Bredasdorp	Eendekuil
Tygerhoek A	Langgewens B	Roodebloem	Hopefield
Tygerhoek B	Philadelphia	Serjeants river	Langgewens
Voorstekop	Riebeek-West	Voorstekop	

3.2.2 Cultivar field trials

The data for wheat genotypes of the ARC-SGI cultivar spring wheat yield trials, comprising 31 genotypes tested in 16 localities in the Southern Cape region of the Western-Cape and 17 localities during 1998-2010, were collected (**Table 3.1**). In this study, four localities in the Rûens and three localities in the Swartland regions were chosen on the basis of continuity of the data for the period. The data of three variables, namely yield (ton per hectare), hectolitre mass (as kilograms per hectolitre) and protein content (on 12% moisture base) were investigated. The geographical positions of the localities are depicted on the map (**Figure 3.1**). The experimental lay-out was a randomised complete block design with four replications. Each trial included the same ARC-SGI check.

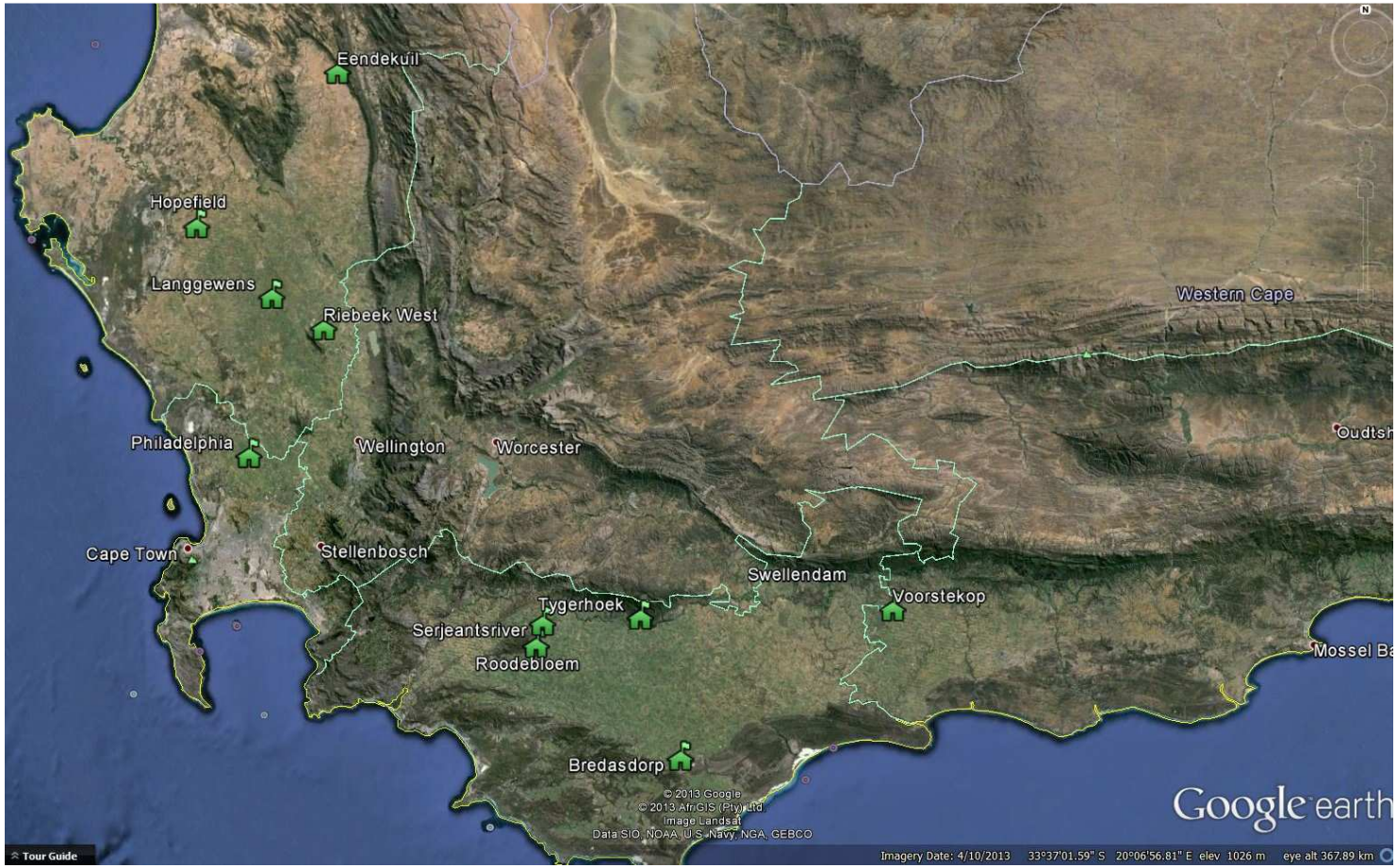


Figure 3.1 Google Map of the localities in the Western Cape.

3.3 Statistical techniques

Three statistical methods were compared to estimate genetic advance, namely:

i. Linear regression over years described by Trethowan *et al*. (2002);
ii. Variance component method described by Comstock and Moll (1963);
iii. Variance component method described by Allard (1960).

Several other statistical techniques were used to investigate the relationship among the different factors and/or variables:

i. Genotypes and environments – AMMI and GGE biplot analyses were performed to evaluate the GEI and to compare the two methods for a single locality. Years within the locality were used as environments;
ii. Pearson's product moment correlation was performed to determine the relationship among the three variables;
iii. PCA was performed to investigate the relationship between the factors (i.e. genotypes) and the variables;
iv. DA was used to determine whether the three variables could discriminate among the years.

3.3.1 The linear regression method over years (TRET)

The first method proposed, was a linear regression according to Trethowan *et al.* (2002). The yield data was analysed for each locality per year with the PROC MIXED procedure of the SAS software (SAS Institute, 2012) to generate genotype means of each locality and year. Genotypes were considered as fixed effects, and blocks (replications) as random effects. The mean of the five highest yielding genotypes from each site and year were calculated and expressed as a ratio of the trial means (TM). The ratio (mean of the five highest yielding genotypes)/TM, TM and the check genotype were regressed against each year to assess gains in yield over time using linear regression analysis (Trethowan *et al.*, 2002). This method will be called TRET in this study.

Two additional calculations were formulated to determine genetic progress with linear regression over time:

- M_T = mean of the five highest yielding genotypes minus the trial mean (TM)
- M_C = mean of the five highest yielding genotypes minus the check mean

The above-mentioned ratio, M_T, and M_C values were compared to determine best fit. The graphical presentations were done by Excel to provide a trend with R^2 values. The p values on the graphical presentations were from the regression generated by the PROC REG procedure of the SAS software (SAS Institute, 2012).

3.3.2 Linear regression with Best Linear Unbiased Predictors (BLUP)

Genotype means are usually calculated by:

$$\bar{y}i = 1/n \sum_{j=i}^{N} yij$$

The genotype effect is calculated by subtracting the overall mean y. This calculation

$$t\dot{\imath} = 1/n \sum_{j=i}^{N} (yij - \bar{y}\imath)$$

is called the Best Linear Unbiased Estimator (BLUE) and corresponds to the 'fixed effect' analysis. As such, it accurately represents what happened in a trial, but it is not the best *predictor* of what might happen if the trial is repeated. An alternative is the Best Linear Unbiased Predictor (BLUP) given by:

$$Ti = \sum_{j=i}^{N} (yij - \bar{y}\imath)\ (n + \sigma^2_{G/}\ \sigma^2_{E})$$

The effect of adding the variance ratio into the denominator is to shrink the genotype effect. If σ^2_G approaches 0, the ratio $\sigma^2_{G/}\ \sigma^2_E$ approaches infinity causing the predicted genotype effect to shrink to 0 (Gilmore, 2010).

3.3.3 Sources of variation, heritability and genetic advance

A third method to determine genetic advance is using sources of variation. This can either be a fixed effects model or a random effects model.

Selection among genotypes is based on phenotypic variation, but the response to selection is a function of genetic variability. The prediction of genetic advance from selection depends on the proportion of phenotypic variance which is due to the genetic variance — this ratio is called heritability (Cooper and Hammer, 1996).

The main purpose of estimating heritability and genetic parameters is to determine genetic advance from selection based on different selection strategies. It is preferable that heritability estimates are made from data collected in multiple localities and during multiple years (Holland *et al*., 2003).

Thus the sources of variation would be partitioned into years, localities, replications within years and localities, genotypes and the interactions of genotypes, years and localities. The statistical model is given by:

$$Y_{ijkl} = \mu + Y_i + L_j + YL_{ij} + B(YL_{ijk}) + G_k + GY_{ik} + GL_{ik} + GYL_{ijk} + €_{ijkl}$$

where: Y_{ijkl} = observed yield or hectolitre mass or protein content value (depending on the elite or cultivar trials)

μ = general mean
Y_i = effect of the year
L_j = effect of the locality
YL_{ij} = interaction effect of the year and locality
$B(YL_{ijk})$ = effect of block within year and locality
G_k = effect of the genotype
GY_{ik} = interaction effect of the genotype and year effect
GL_{jk} = interaction effect of the genotype and locality
GYL_{ijk} = interaction effect of the genotype, year and locality
$€_{ijkl}$ = error or residual effect

$$€_{ijkl} \sim NID(0,\sigma^2)$$

The primary method used to analyse multi-environment trials (MET) is based on the ANOVA (Model 1), which is a fixed effects model and requires homogenous variance-covariance of data. While other models are available, the problem of how the models should be assessed and which model is more suitable for a given trial's data has not been solved (Hu and Spilke, 2011).

3.3.4 Expectations of mean squares from a fixed model (Model 1)

The data was subjected to an ANOVA and analysed by the PROC GLM procedure of the SAS software (SAS Institute, 2012). Variance components were estimated using TYPE III expected mean squares of the ANOVA as pointed out in **Table 3.2**. It is assumed that the effects of genotype, the interactions of genotype with year and locality, and error are independently random, normally distributed and additive in their contribution to yield in the elite trials and yield, hectolitre mass (HLM) and protein content in the cultivar trials. Furthermore, the variances of the various effects are constant from one environment (meaning either year or locality or both) to another. Comstock and Moll (1963) emphasised that these assumptions need not be rigorously implemented because plot error variances differ from one experiment to the other.

Table 3.2 Sources of variation, calculated mean squares (MS) and their expected values

Source of variation	MS	E(MS)
Years (Y)		
Localities (L)		
Yx L		
B x Yx L		
Genotypes (G)	M1	$\sigma^2_E + r\,\sigma^2_{GLY} + rs\,\sigma^2_{GY} + rt\,\sigma^2_{GL} + rst\,\sigma^2_G$
G x Y	M2	$\sigma^2_E + r\,\sigma^2_{GLY} + rs\,\sigma^2_{GY}$
G x L	M3	$\sigma^2_E + r\,\sigma^2_{GLY} + rt\,\sigma^2_{GL}$
G x L x Y	M4	$\sigma^2_E + r\,\sigma^2_{GLY}$
Error	M5	σ^2_E
Corrected Total		

In the optimisation of years, localities, replications and genotypes in making maximum genetic progress, the method of determining the relationship of number of genotypes,

blocks, localities and years in formulating genetic advance was proposed by Comstock and Moll (1963). In this model, the optimum allocations of components can be determined from the ratios of the genetic component of variance, the interaction components of genotypes x years, genotypes x localities, and genotypes x years x localities to the error variance component. The same model (called model 1 in **Figure 3.2Error! Reference source not found.**) applies in this case and was tested as a fixed effects model with PROC GLM procedure of the SAS software (SAS Institute, 2012).The formula used was:

$$\Delta G = \frac{d\,\bar{x}m}{\sqrt{\frac{1}{rst} + \frac{e}{st} + \frac{by}{t} + \frac{bp}{s} + d}}$$

where: ΔG = genetic advance

d = M1 /M5

$\bar{x}m$ = standardised selection index for 1/(nr of genotypes)

e = M5

by = M2/M5

bp = M3/M5

r = replication of each s localities in t years (blocks)

s = number of localities

t = number of years

3.3.5 Linear mixed model framework to calculate variance components and genetic advance (model 2 to model 5)

The primary method used to analyse METs is based on ANOVA, which is a fixed effects model and requires a homogenous variance-covariance of data. While other models are available, the problem of how the models should be assessed and which model is more suitable for a given trial's data has not been solved (Hu and Spilke, 2011).

The data was analysed by the PROC MIXED procedure of the SAS software (SAS Institute, 2012). The restricted (or residual or reduced) maximum likelihood (REML) was chosen as the estimation method. The estimates of the variance components and their standard errors were calculated by the COVTEST option, e.g.

PROC MIXED data = my data method = REML COVTEST

3.3.6 Proposed linear mixed models

The data was analysed in its unbalanced form. Five models were used:

i. Model 1 - the combined analysis of variance with fixed effects.
ii. Model 2 - linear mixed model analysis for each year and each locality with genotype effects as fixed and blocks as random effects.
iii. Model 3 - linear mixed model analysis for each year with genotype and the interactions with genotype as random effects and locality, blocks within locality and interactions with locality as fixed effects.
iv. Model 4 - linear mixed model analysis for full model without taking heterogeneity of variances in account. Genotype and genotype interactions with year and locality as random effects and year, locality, the interaction of year and locality and block within year and locality as fixed.
v. Model 5 - linear mixed model analysis for full model taking heterogeneity of variances in account. Genotype and genotype interactions with year and locality as random effects and year, locality, the interaction of year and locality and block within year and locality as fixed. SAS mixed procedure with subject = year x locality with random statement, was used to compensate for the heterogeneity in year and locality.

The results of the models will not be given formally. The model with the smallest AIC and BIC was taken in METs as the best model. Model 5 should have been the model with the smallest AIC and BIC values but the large data sets and unbalanced data prevented the fitting of the model. Runtime errors were frequently experienced.

3.3.7 Model selection

Several variance-covariance structures are available and discussed in Hu and Spilke (2011) in more detail. To select an appropriate model, there are a number of criteria to be considered:

$$\textit{Akaike Information Criterion (AIC)} = -2LL + 2 \times q$$

where: LL denotes the log maximum likelihood of the related model
q is the number of parameters of the variance-covariance structure

The calculation formulae of the information criteria are given in such a way that the model with the lower value of the information criterion is preferred.

$$\textit{Schwarz Bayesian Information Criterion (BIC or SIC)} = -2LL + \log(N) \times q$$

where: LL denotes the log maximum likelihood of the related model

N is the total number of observations

q is the number of parameters in the variance-covariance matrix

The best model is again the model with the lower value of the information criterion. The proposed methods and models for estimating genetic advance in this study are presented in **Figure 3.2.**

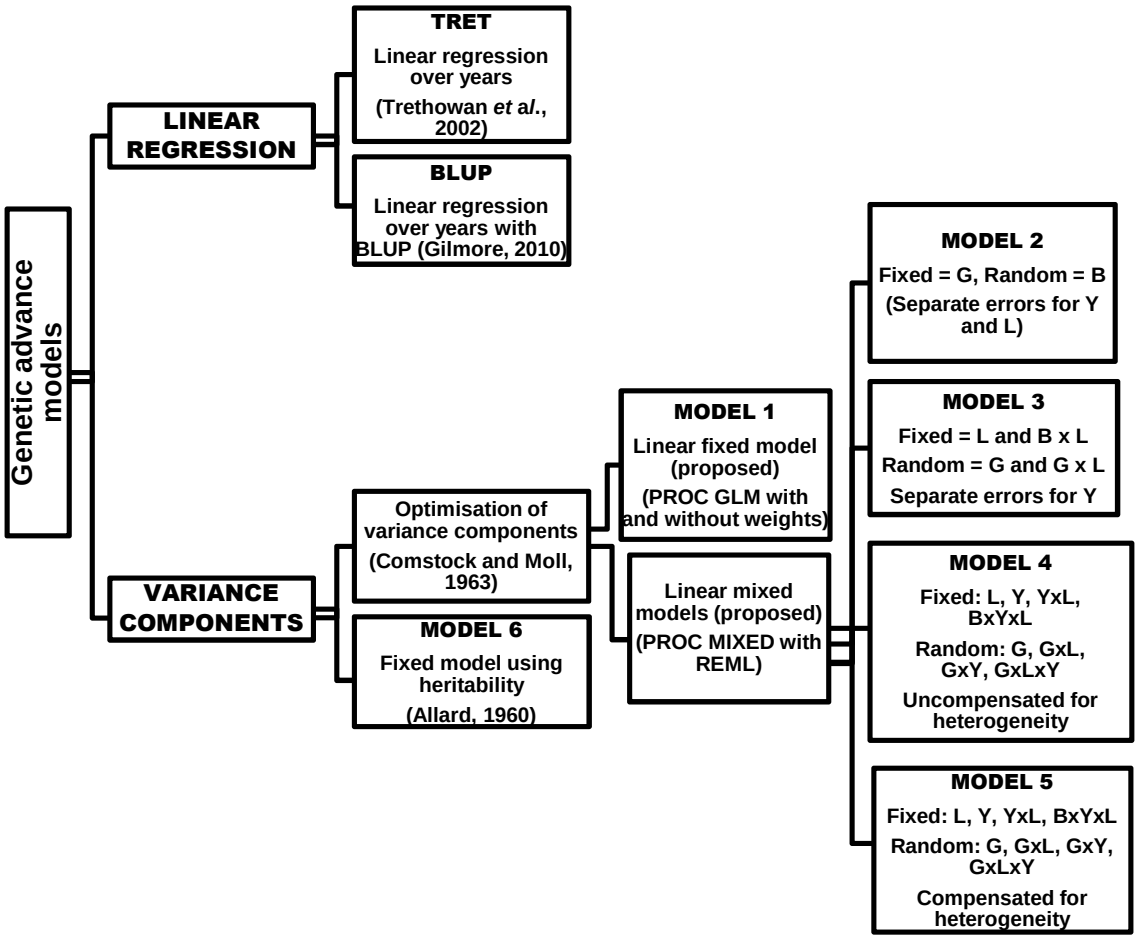


Figure 3.2 Schematic presentation of models to estimate genetic advance. (B=blocks, G=genotypes, L= localities, Y=years)

3.3.8 Model 6 – the model proposed by Allard (1960)

Assuming normally distributed residuals of the variable in question, genetic advance under selection can, according to Allard (1960), be calculated by:

$$G_s = ih^2\sigma_p$$

where: i = standard selection differential from a set of genotypes at 10% selection intensity

h^2 = broad sense heritability and

σ_p = phenotypic standard deviation.

In terms of the model by Comstock and Moll (1963), broad sense heritability can be defined by:

$$h^2 = \sigma^2_G / \sigma^2_p$$

where:

$\sigma^2_p = (\sigma^2_G - \sigma^2_{GY} - \sigma^2_{GL} + \sigma^2_{GLY})/rst$

σ^2_G = variance of the genotypes

σ^2_{GY} = variance of the interaction of genotype by year

σ^2_{GL} = variance of the interaction between genotypes and localities

σ^2_{GLY} = variance of the interaction, among genotypes, years and localities

r = number of blocks

s = number of localities

t = number of years.

3.3.9 AMMI versus GGE biplot

Targeting genotype selection onto its growing environments is the prime interest of any plant-breeding programme. To understand this, breeding programmes usually undertake a rigorous genotypes performance evaluation across environments (localities, years or both) mostly at the final stage of variety development. However, GEI reduces the genetic progress in plant breeding programmes through minimizing the association between phenotypic and genotypic values (Comstock and Moll, 1963). Hence, GEI must be either exploited by

selecting superior genotypes for each specific target environment or avoided by selecting widely adapted and stable genotypes across wide range of environments.

Two frequently used models to investigate GEI are the AMMI model (Gauch *et al*., 2008) and the GGE model (Yan *et al*., 2007).

3.3.9.1 The AMMI analysis

The AMMI model estimates the magnitude and significance of G x E interaction effects of each genotype‘s or cultivar’s response. It essentially combines ANOVA techniques and PCA into a single model where ANOVA allows studying the main effects of genotypes and environments, while the GEI is explored by multivariate PCA (Gauch and Zobel, 1996).

3.3.9.2 The GGE analysis

The GGE model shows genotype response to specific environments and is very similar to AMMI, except for the fact that this model allows grouping environments with similar performance and graphically identifying which genotype has the greatest potential within each subgroup of environments. That is, it groups genotypes and environments without any crossed interaction. Each genotype and environment has a corresponding principal component (PC) with a value to which a vector is assigned. Given this association between genotypes and environments, it is possible to determine the adaptation and/or magnitude of the GEI defined by the linear dependence and association between them (Castillo *et al*., 2012).

The AMMI model is represented by $\bar{y}_{ij.} = \mu + \tau_i + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij.}$

The GGE model is given by $\bar{y}_{ij.} = \mu + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij.}$

where: $\bar{y}_{ij.}$ is the mean of the i[th] cultivar in the j[th] environments

μ is the overall mean

τ_i is the genotypic effect (separate effect in AMMI not in GGE)

δ_j is the environment effect

λ_k ($\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_t$) are scaling constants (singular values) that allow the imposition of ortho-normality constraints on the singular vectors for cultivars, $\boldsymbol{\alpha}_{ik}=(\alpha_{1k},\ldots,\alpha_{gk})$, and environments, $\boldsymbol{\gamma}_{jk}=(\gamma_{1k},\ldots,\gamma_{ek})$, such that

$\sum_i \alpha_{ik}^2 = \sum_j \gamma_{jk}^2 = 1$ and $\sum_i \alpha_{ik}\alpha_{ik'} = \sum_j \gamma_{jk}\gamma_{jk'} = 0$ for k≠k′;

α_{ik} and γ_{jk} for k=1,2,3,… are called "primary, secondary, tertiary" etc. effects of cultivars and sites, respectively

$\bar{\varepsilon}_{ij.}$ is the residual error assumed to be NID (0, σ^2/r) (where σ^2 is the pooled error variance and *r* is the number of replicates)

Least squares estimates of the multiplicative (bilinear) parameters in the kth bilinear term are obtained as the kth component of the deviations from the additive (linear) part of the model. In the AMMI model, only the GEI term is absorbed in the bilinear terms, whereas in the SREG model, the main effects of cultivars (G) plus the GEI are absorbed into the bilinear terms.

The GGE biplot method and the AMMI method were applied for visual examination of the GEI in the cultivar trial data using GenStat 15th edition (Payne *et al*., 2012).

3.3.10 Other statistical techniques

The Pearson's product moment correlation matrix of the pairwise correlations among the three dependent variables was built to show their linear relationships, which were then graphically represented using two major components of PCA. Hierarchical cluster analysis of the target environments was performed to identify the similarities among the environments. The dendrogram was constructed using Ward's method with the automatic clustering option. Linear discriminant analysis was performed to see whether the variables could separate the years and if a trend was evident in the separation of the years. These methods were performed using XLSTAT software (XLStat, 2012).

3.4 Results and discussion

3.4.1 Preliminary analyses

When data sets are unbalanced, the arithmetically simpler, sequential (Type I) Sum of Squares (SS) can be biased for some effects (with bias proportional to the degree of imbalance). Corrected SS (usually Type III) should be adopted for this situation (Miliken and Johnson, 1984).

Preliminary analysis of variance with PROC GLM of the software SAS (2012) using Error Type III was carried out for individual trials (each locality and year) to assess the residuals and predicted values. The residuals were tested by the Shapiro and Wilk test for normality (Shapiro and Wilk, 1965). Outliers (any residual value above the absolute value of three ($ABS(r_{ijkl}) > 3$) were replaced by the predicted values of the above-mentioned model (Section 3.3.3).

Homogeneity of variances was tested by Levene's test in the PROC GLM in the SAS programmes (Levene, 1960). Homogeneity of variances was also tested in the linear mixed model environment by PROC MIXED programmes suited for this assumption. Heterogeneity in year variances and the interactions with year, genotype x year, genotype x locations, and genotype x year x locations were found.

3.4.2 Heterogeneity, unbalanced data and possible remedies

For the fixed analysis of variance model, weighted analysis of variance was considered. The weight was the reciprocal of year variances:

$$(W_{ijkl} = \frac{1}{\sigma i jkl})$$

This is in accordance with studies by Annicchiarico (2002), Morris *et al.* (2004) and So (2009). De Vita *et al.* (2010) and Sener *et al*. (2009) used natural logarithm transformation to resolve heterogeneity of variances.

In the linear mixed model scenario, when data is unbalanced or heteroscedastic (heterogeneous variances) the mean might be taken to be the adjusted mean, i.e. the best

linear unbiased estimator (BLUE) assuming fixed genotypic effects. Furthermore, with unbalanced data and a large number of genotypes it is desirable to estimate genotypic effects by best linear unbiased prediction (BLUP) rather than BLUE and in computation of BLUP, an adjusted genotype, is not directly involved (Piepho and Mohring, 2007). To accommodate the imbalance of the dataset, Van Eeuwijk *et al*. (2001) and Smith *et al*. (2005) proposed that the data should be analysed as mixed models with separate residual terms for the different trials.

These proposed solutions were tested. The outcome is discussed in the appropriate sections.

3.4.3 Linear regression over years using TRET

Most research studies report progress in mean yield over time from two approaches: either from historical genotypes, *inter alia* De Vita *et al.* (2010) and Rodrigues *et al*. (2007), or using mean yield for genotypes in the experimental years to express genetic progress, e.g. Morgounov *et al.* (2010) and Sener *et al.* (2009). Measuring rates of progress in dryland environments (rain fed environments) by TM or the check genotype can be misleading as yield is greatly affected by rainfall patterns in any given year.

In this research, TRET, the method of Trethowan *et al.* (2002), was chosen to regress mean yield of the five highest yielding genotypes from each site and each year as a ratio of the trial mean (Ratio) and compared with the trial mean (TM) and check mean (Check) against experimental years. The majority of genotypes in the elite trials are not expected to be adapted to a specific environment. Up to five new genotypes could enter the trials each year (Pakendorf personal communication). It is thus appropriate to calculate the mean of the five highest yielding genotypes. The reason for regressing the ratio, rather than %TM as Trethowan *et al*. (2002) proposed, lies in the graphical presentation of the results on one Y-axis.

Although the same check was used throughout the period and for all localities in both the elite and cultivar trials, the modifications described in Section 3.3.1, namely M_T (mean of five

highest yielding genotypes minus trial mean), M_C (mean of five highest yielding genotypes minus check mean) would be relevant. These modifications did not add value to the study.

The data of the localities were tested by Levene's test for homogeneity of locality variances (Levene, 1960). To enable comparison among the different methods proposed to determine genetic advance, it was decided to perform the linear regression over localities in a region.

Results generated with BLUE were similar to the results of TRET. This was expected due to genotype being a fixed effect in the linear regression for each year and each locality, while blocks are the random effect. If BLUP is used, the genotype and blocks would have been the random effect with no fixed effect.

3.4.3.1 Elite trials

In the Rûens region (**Figure 3.3**) the trial means and check mean showed no significant trends, respectively (R^2 = 0.07, p = 0.48 and R^2 = 0.02, p = 0.64). The ratio showed a significant advance in yield over time with a slope (b = 0.01 and R^2 = 0.28, p = 0.05). Thus there was nearly 1% progress per year over 15 years.

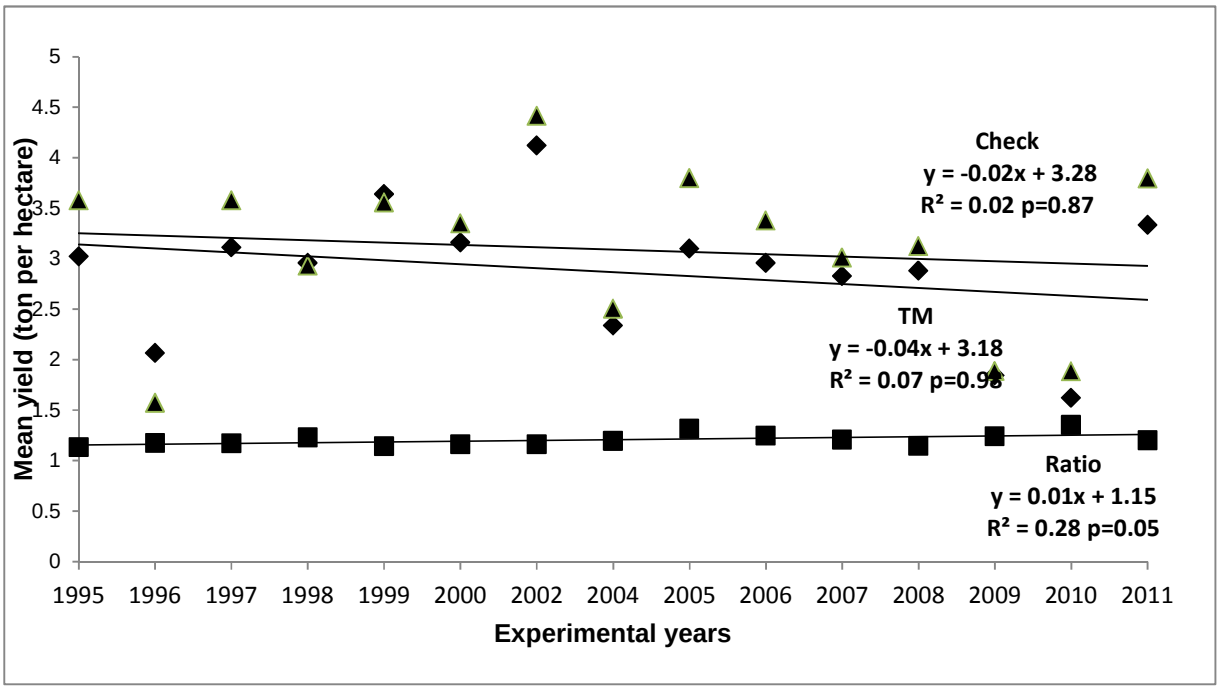


Figure 3.3 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the elite trials in the Rûens region. ♦ = Trial means (TM), ▲ = Check, ■ = Ratio.

In the Swartland region (**Figure 3.4**) the TM and Check, on the one hand, showed significant negative trends respectively (b = -0.13, R^2 = 0.26, p = 0.01 and b = -0.14, R^2 = 0.28, p = 0.01). Ratio, on the other hand showed a significant advance in yield over time with a slope (b = 0.01 and R^2 = 0.24, p = 0.05). Thus 1% progress was made per year over 16 years.

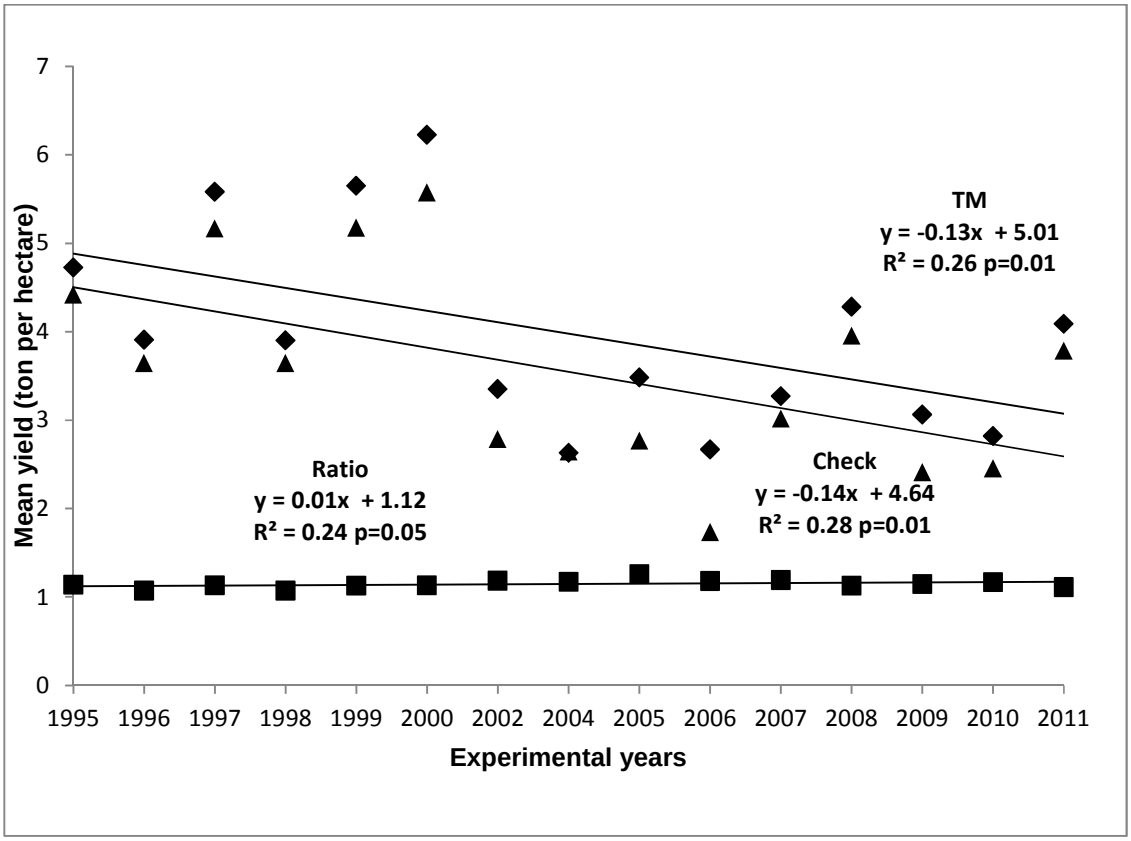


Figure 3.4 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the elite trials in the Swartland region. ♦ = Trial means (TM), ▲ = Check, ■ = Ratio.

3.4.3.2 Cultivar trials

In this part of the study we are looking at yield as well as two quality traits, HLM and protein content. In the Rûens region (**Figure 3.5**) the Ratio showed no significant advance in yield over time with a slope (b = 0.00 and R^2 = 0.02, p = 0.60). No significant trends were evident for TM (R^2 = 0.15, p = 0.36) and Check (R^2 = 0.16, p = 0.33).

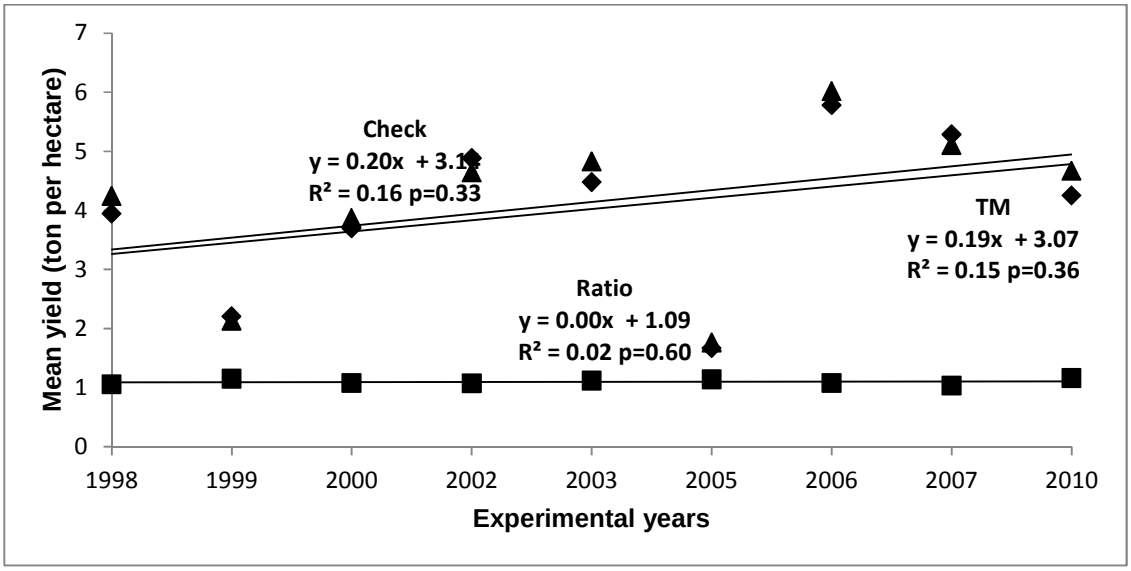


Figure 3.5 Mean yield of five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trials in the Rûens region. ♦ = Trial means (TM), ▲ = Check, ■ = Ratio.

In the Swartland region (**Figure 3.6**) the Ratio showed no significant advance in yield over time with a slope (b = 0.00 and R^2 = 0.00, p = 0.79). TM showed no significant trend either (R^2 = 0.17, p = 0.27) Check showed a significant trend (b = 0.13 with R^2 = 0.38, p = 0.05). This is misleading because the trend of Check might be influenced by climatic influences.

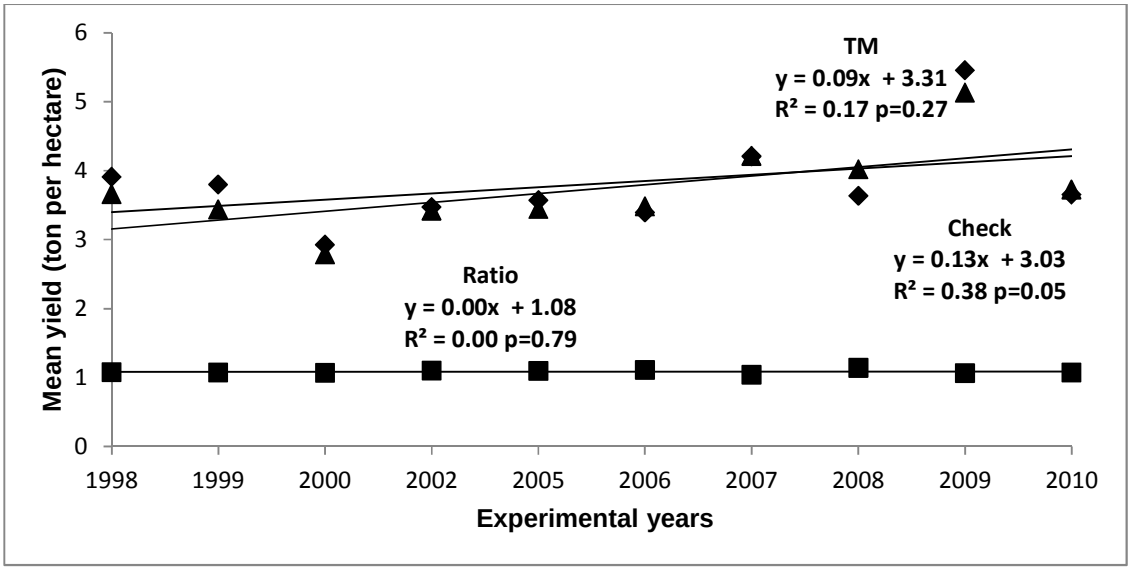


Figure 3.6 Mean yield of five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trials in the Swartland region. ♦ = Trial means (TM), ▲ = Check, ■ = Ratio.

Although no trends were evident in the Ratio for both the regions, it is possible that the Ratio absorbed the climatic effect better than TM and Check. This was precisely why this method was chosen.

The Ratio for HLM in the Rûens region (**Figure 3.7**) showed no significant trend (b = 0.00, R^2 = 0.23, p = 0.19). TM (b = -0.17, R^2 = 0.15, p = 0.30) and Check (b = -0.08, R^2 = 0.03, p = 0.69) showed no trend either.

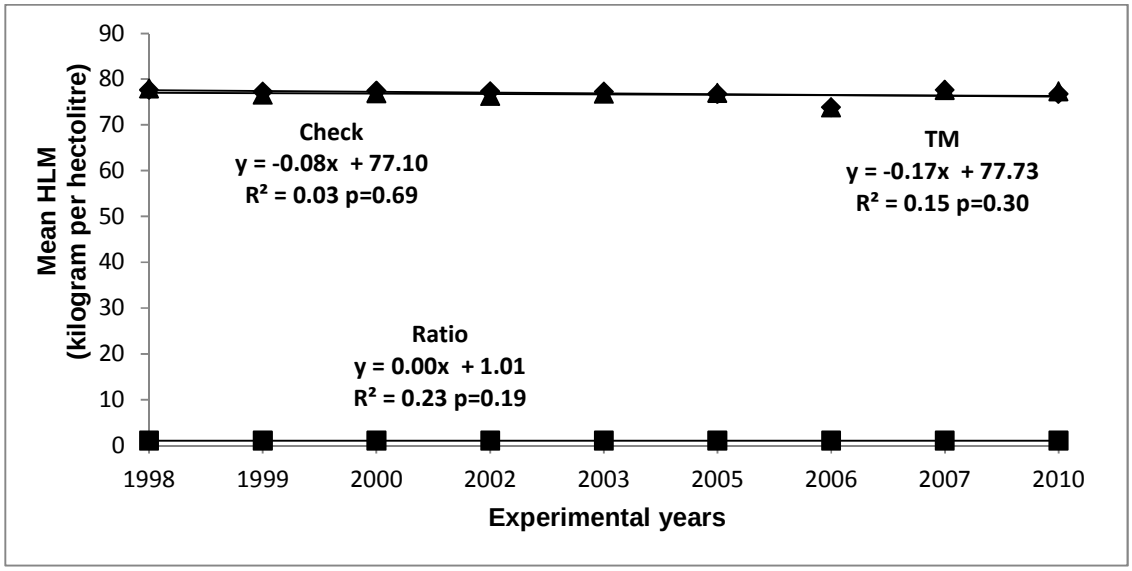


Figure 3.7 Mean hectolitre mass (HLM) of five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trials in the Rûens region. ♦ = Trial means (TM), ▲ = Check, ■ = Ratio.

In the Swartland, the Ratio for HLM (**Figure 3.8**) showed no significant trend (b = 0.00, R^2 = 0.24, p = 0.19). TM (b = 0.17, R^2 = 0.10, p = 0.45) and Check (b = 0.27, R^2 = 0.18, p = 0.23) showed no trend either.

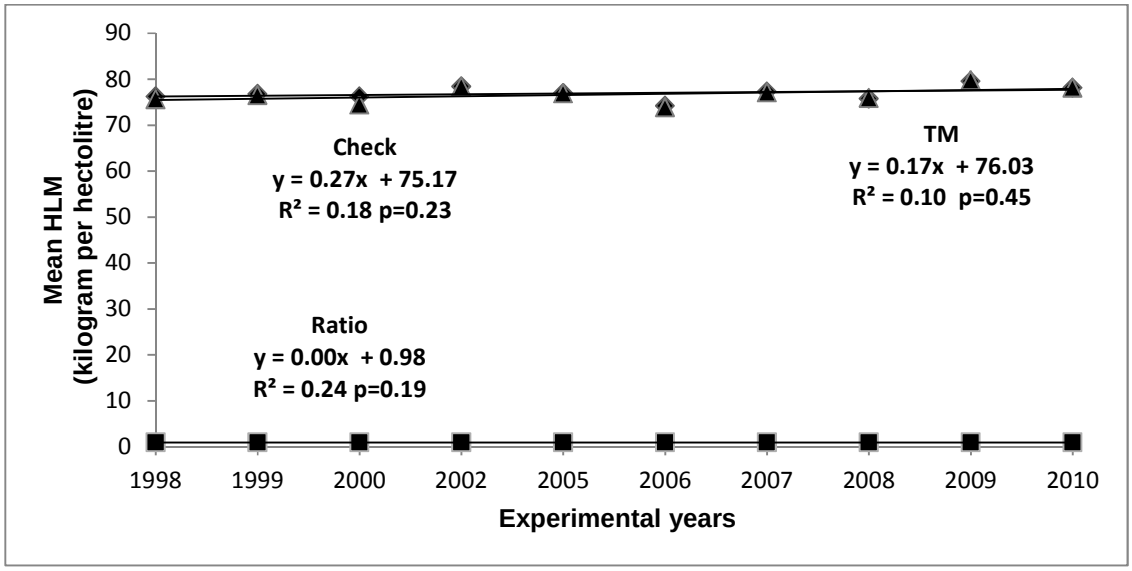


Figure 3.8 Mean hectolitre mass (HLM) of five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trials in the Swartland region. ♦ =Trial means (TM), ▲ = Check, ■ = Ratio.

Ratio for protein content in the Rûens region (**Figure 3.9**) showed a significant trend (b = 0.01, R^2 = 0.33, p = 0.11). However, TM (b = -0.04, R^2 = 0.02, p = 0.79) and Check (b = -0.04, R^2 = 0.01, p = 0.80) showed no trend.

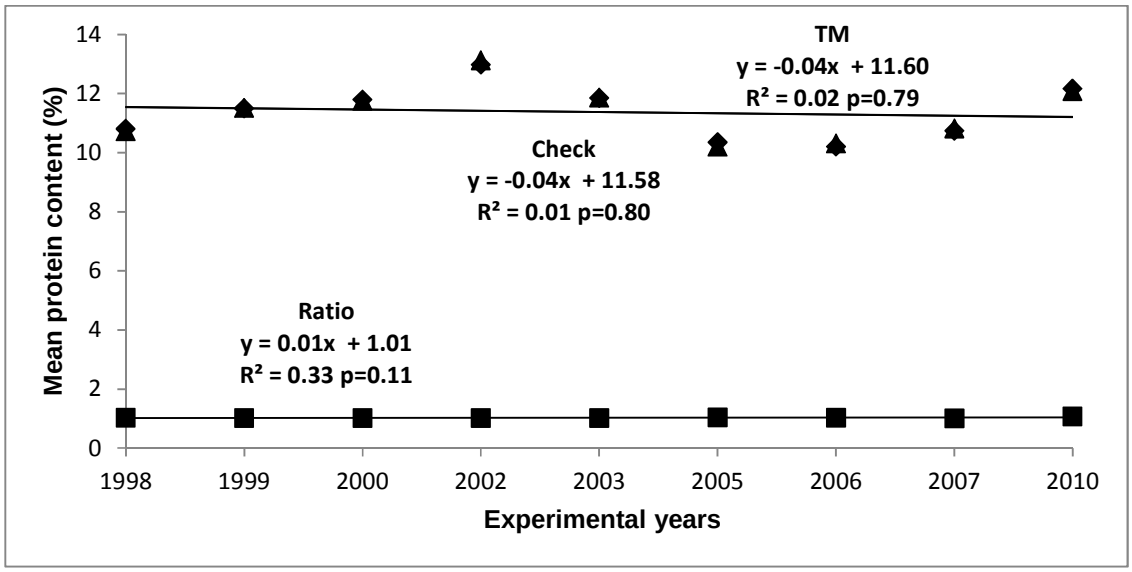


Figure 3.9 Mean protein content of five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trials in the Rûens region. ♦ = Trial means (TM), ▲ = Check, ■ = Ratio.

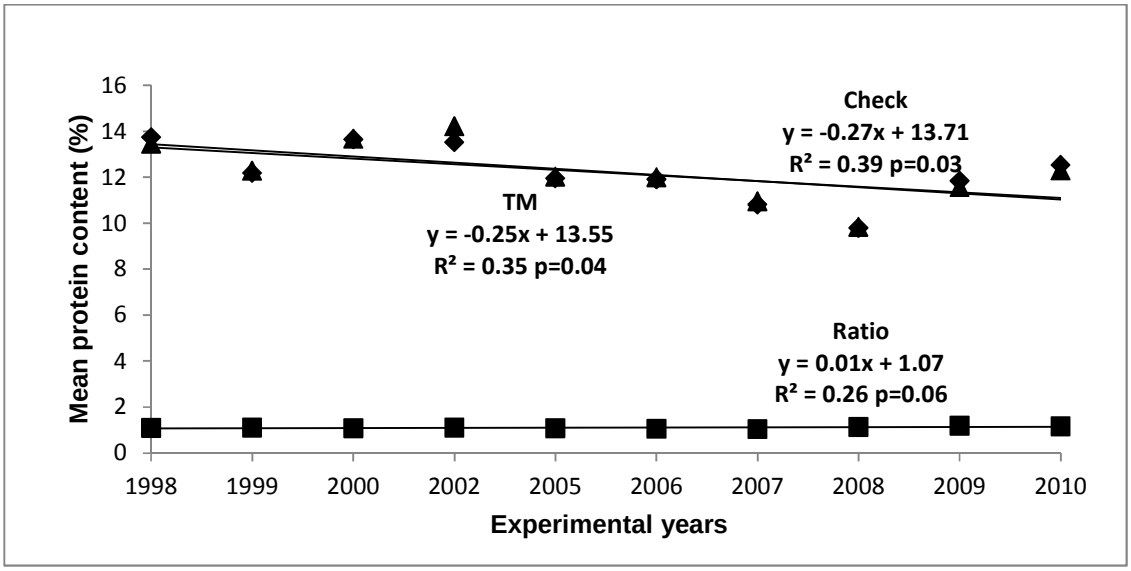


Figure 3.10 Mean protein content of five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trials in the Swartland region. ♦ = Trial means (TM), ▲ = Check, ■ = Ratio.

In the Swartland, the Ratio for protein content (**Figure 3.10**) showed a significant trend at 16% (b = 0.01, R^2 = 0.26, p = 0.06). TM (b = -0.25, R^2 = 0.35, p = 0.04) and Check showed a significant trend as well (b = -0.27, R^2 = 0.39, p = 0.03). This is in accordance with Sener *et al*. (2009) and Mut *et al*. (2010).

The present study examined a set of approximately 285 bread wheat genotypes in the breeding trials and 30 bread wheat genotypes in the cultivar trials. The method to combat environmental fluctuations, namely Ratio (TRET), successfully showed progress or a lack thereof, in the elite trials. Genetic improvement in yield determined by linear regression of 1% per year could only be shown in both regions of the elite trials. This is consistent with the results reported by Morgounov *et al*. (2010) of a genetic gain of nearly 1% per year and the report on genetic advance in Canada of 0.74% per year (DePauw *et al.,* 2007). Trethowan *et al*. (2002) reported increases of 0.09 to 0.34% per year. De Vita *et al.* (2010) and Underdahl *et al*. (2008) found 1.3% per year.

3.4.4 Sources of variation, heritability and genetic advance

The mean squares obtained from the combined analysis of variance showed all effects were significant ($p<0.0001$) for yield (Yld), HLM, protein content (Prot), as well as for the logarithmic transformed variables yield (LYld), HLM (LHLM) protein content (LProt) and those of the weighted analyses (WYld, WHLM and WProt respectively) from the elite trials and cultivar trials of the different regions for both planting dates (Appendix A **Tables A1** to **Table A4)**.

3.4.5 Comparison of linear fixed models and linear mixed models

With reference to mixed models 2 to 5 (**Figure 3.2**), the model with the smallest AIC (Akaike Information Criterion) and BIC (Schwarz Bayesian Information Criterion*)* should be selected for analysis of multi-environment trials as the best model. Theoretically, model 5 should have been the model with the smallest AIC and BIC values but the large data sets and unbalanced nature of the data prevented the model to be fitted and runtime errors were frequently encountered. According to the criteria stipulated, model 3 was the second best model but its pooled error was similar to the single error of model 4 (results not shown). Therefore model 4 was chosen due to computer efficiency. In **Table 3.3** the linear regression coefficient b (TRET), genetic advance (ΔG) of the fixed model (model 1), mixed model (model 4) and model 6 [the model according to Allard (1960)] for yield of the elite and cultivar trials are shown.

Table 3.3 Estimates of phenotypic coefficient of variation (%PCV), genotypic coefficient of variation (%GCV), % broad sense heritability (H^2) and genetic advance (% ΔG) from the different models for the yield (Yld) of the elite and the cultivar trials (LYld = logarithmic transformed yield, WYld = weighted analysis of yield)

						% ΔG			
Trial	**Region**	**Variable**	**%PCV**	**%GCV**	**H2**	**Model 1**	**Model 4**	**Model 6**	**TRET**
Elite	Rûens	Yld	17.56	8.00	0.02	1.73	4.47	0.13	1.00
	Rûens	L(Yld)	5.98	3.14	0.03	2.07	4.40	0.07	
	Rûens	WYld	52.07	7.64	0.02	1.67		0.33	
	Swartland	Yld	14.95	4.54	0.01	1.00	4.10	0.07	1.00
	Swartland	L(Yld)	5.06	1.37	0.01	0.73	3.73	0.00	
	Swartland	WYld	39.94	3.80	0.01	0.87		0.13	
Cultivar	Rûens	Yld	12.53	1.79	0.02	2.33	4.60	0.13	0.00
	Rûens	L(Yld)	12.67	2.29	0.03	4.11	5.00	0.07	
	Rûens	WYld	46.91	7.67	0.03	3.11		0.40	
	Swartland	Yld	15.40	2.13	0.02	1.90	2.93	0.13	0.00
	Swartland	L(Yld)	13.11	1.70	0.17	1.80	2.53	0.00	
	Swartland	WYld	34.62	6.46	0.04	3.10		0.47	

Phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) is not only useful for comparing the relative amount of phenotypic and genotypic variations but are also very useful to estimate the range for improvement by selection. The reliability of a variable to be selected for a breeding programme, among other factors, depends on the magnitude of its coefficient of variation (CV) especially the GCV. However, the differences between genotypic and phenotypic coefficient of variability indicate the environmental influence (Bello *et al*., 2012).

A lower value of CV generally depicts low variability within a breeding population. A high proportion GCV to the PCV is desirable in breeding trials. Results given in **Table 3.3** show that PCV was more than two times higher than GCV for the yield and its transformations (in **Table 3.3** called variables). This indicates that environments (years and localities) in the expression of yield in both the Rûens and Swartland regions are the dominant factors. The low heritability (H^2) values and the low yield improvement (GA) estimates indicate large environmental influences. Eid (2009) found similar results. Inheritance of yield is complex and is greatly influenced by several genes interacting with various environmental conditions (Koekemoer, 2003).

The yield GA estimates within a region varied extensively among the models and within a model for yield and its transformed variables (i.e. logarithmic and weighted). In the Rûens region, estimates of genetic improvement from weighted analysis were similar to those of the original data (model 1). The estimates of the genetic advance values from the logarithm transformed data are similar to those of the original data in model 4. The yield GA estimates of model 1 and model 4 from the original data and logarithmic transformed data in the Swartland region are similar. The GA estimates from model 6 in both regions are totally different.

The comparison of the genetic advance values of the different models is difficult. There is no statistic to compare the values, such as a standard error or a test due to the “singleness” of the value. In the field of psychometrics, reliability (also called agreement) is tested by Pearson’s product moment correlation and/or Cronbach’s alpha (Gregory, 2000). These statistics evaluate consistency of different tests (in layman’s terms: do the tests test the some construct?). In this study the construct was genetic advance from model 1 (the linear fixed effects model), model 4 (linear mixed model), model 6 (method according to Allard)

and TRET (the slope from the linear regression). Pearson's product moment correlation was used in this study to test agreement between two models and Cronbach's alpha test agreement among the models simultaneously. The tests were performed by combining the genetic advance values of the cultivar and elite trials to ensure more reliable results. The agreement between genetic advance of yield (Yld) for model 1 and model 4 was moderate and significant at 10% (r=0.69, p=0.06) for the Rûens region. The Cronbach alpha did not validate these results. For the Swartland region the agreement between model 1 and model 4 was significant at p=0.001 with r=0.94 and model 1 and model 6 with r=0.70 and p=0.01. Cronbach alpha = 0.78 with model 1, model 4, model 6 and TRET and Cronbach alpha=0.88 without model 6. Thus for the elite trials the annual yield genetic improvement (genetic advance) on the original data in the Rûens region were 1.73% (model 1), 4.47% (model 4) and TRET estimated 1% per year. The yearly yield genetic advance estimates for the Swartland region were 1 % (model1), 4.10% (model 4) and 1% for TRET.

Table 3.4 Estimates of phenotypic coefficient of variation (%PCV), genotypic coefficient of variation (%GCV), broad sense heritability (H^2) and genetic advance (%ΔG) from the different models for HLM of the cultivar trials (LHLM = logarithmic transformed HLM, WHLM = weighted analysis of HLM)

					% ΔG			
Region	**Variable**	**% PCV**	**% GCV**	**H^2**	**Model 1**	**Model 4**	**Model 6**	**TRET**
Rûens	HLM	1.21	0.24	0.04	4.89	20.11	0.67	0.00
Rûens	L(HLM)	0.28	0.06	0.04	4.89	20.00	0.00	
Rûens	WHLM	2.34	0.43	0.03	4.44		1.11	
Swartland	HLM	2.05	0.38	0.03	3.11	8.44	1.00	0.00
Swartland	L(HLM)	0.48	0.09	0.03	3.11	8.44	0.00	
Swartland	WHLM	1.52	0.33	0.04	4.22		1.11	

Estimates of GA for the quality traits (**Table 3.4** to **Table 3.5**) calculated by model 4, is much higher than the values estimated by model 1 in HLM and protein content and their transformed variables (called variable in the tables). The GA values from model 6 are much lower.

Table 3.5 Estimates of phenotypic coefficient of variation (%PCV), genotypic coefficient of variation (%GCV), broad sense heritability (H^2) and genetic advance (ΔG) from the different models for protein content (Prot) of the cultivar trials (LProt = logarithmic transformed protein content, WProt = weighted analysis of protein content)

					% ΔG			
Region	**Variable**	**% PCV**	**% GCV**	**H^2**	**Model 1**	**Model 4**	**Model 6**	**TRET**
Rûens	Prot	3.94	0.92	0.05	4.33	18.00	0.44	1.00
Rûens	L(Prot)	1.62	0.36	0.05	4.11	16.44	0.00	
Rûens	WProt	10.18	2.02	0.04	3.44		0.89	
Swartland	Prot	5.15	1.67	0.10	6.33	16.77	1.22	1.00
Swartland	L(Prot)	2.04	0.64	0.09	6.11	15.78	0.11	
Swartland	WProt	9.85	3.23	0.10	6.67		2.33	

Barnard *et al.* (2002) and Bilgin *et al*. (2011) reported moderately high H^2 (>0.50) values for HLM and protein content which indicated fair repeatability.

In this study very low broad sense H^2 (<0.10) values were found for yield, HLM and protein content which indicated weak repeatability. Low repeatability means no response to selection and thus no genetic advance.

Although logarithmic transformation and weighted analyses were performed to combat year differences, the “verdict” is given on the original data. The similarities and dissimilarities among the original data (yield, HLM and protein content) and its transformations (LYld and WYld etc.) were discussed earlier.

In the elite trials the annual yield genetic improvement (genetic advance) in the Rûens region was 1.73% (model 1), 4.47% (model 4) and TRET estimated 1% per year. The yearly yield genetic advance estimates for the Swartland region were 1% (model 1), 4.1% (model 4) and 1% for TRET. Annual HLM rates in the Rûens region were 4.89% (model 1), 20% (model 4) and TRET estimated 0% per year. The yearly HLM rates in the Swartland region were 3.11% (model1), 8.4% (model 4) and 0% for TRET. In contrast, TRET showed genetic improvement in both the regions for protein content of 1%, 4.33% (model 1), 18% (model 4) in the Rûens region and in the Swartland region 6.33% (model 1) and 16.77% in model 4.

3.4.6 AMMI versus GGE

It is important for wheat breeders to identify genotypes adapted or stable to different environments, thereby achieving genetic gain through screening genotypes for high adaptation and stability under different environmental conditions prior to their release as cultivars. The AMMI and GGE biplots are the latest methods to investigate GEI. The objective of this study was to compare these methods and not to give an interpretation of the results. An in depth explanation of the above models may be found in Gauch *et al.* (2008) and Yan *et al*. (2007).

The prerequisites for AMMI and GGE analyses are significant GEI and balanced data. To illustrate the similarities and dissimilarities of the two methods, six genotypes over two years and four localities (eight environments) in the two separate regions were submitted to these methods. The combined ANOVA with balanced data indicated a highly significant environment, genotype and GEI between genotypes and environmental effects (results not shown).

3.4.6.1 Rûens

The AMMI analysis of variance for yield, HLM and protein content respectively (**Table 3.6**) showed that the total sum of squares attributed to genotypes was 4.77%, 7.30%, 2.25%, and the environments were 82.24%, 78.47%, 85.05%. GEI interaction accounted for 4.92%,8.35% and 4.48% of sum of squares.

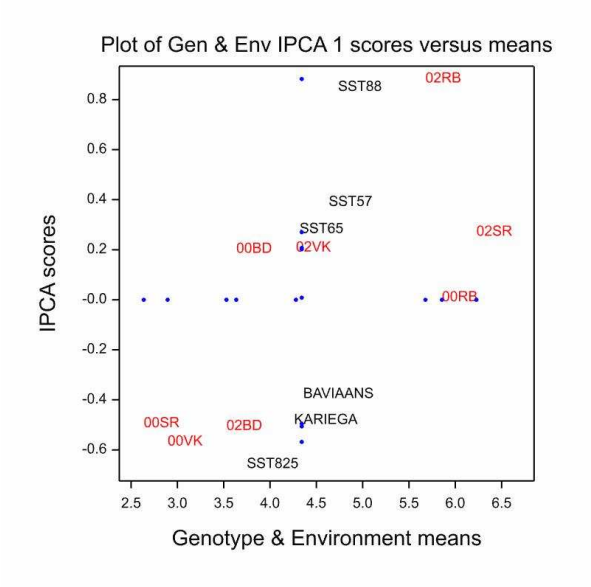
Table 3.6 Sources of variation from the AMMI model of the Rûens cultivar trials

		Yield		Hectolitre mass		Protein content		
Source	**DF**	**SS**	**% SS**	**SS**	**% SS**	**SS**	**% SS**	**P**
Block(Environments)	24	18.90	4.69	26.50	2.25	12.80	3.28	<0.00
Treatments	47	370.10	91.90	1109.10	94.10	358.50	91.76	<0.00
Genotypes	5	19.20	4.77	86.00	7.30	8.80	2.25	<0.00
Environments	7	331.20	82.24	924.80	78.47	332.30	85.05	<0.00
GxE interaction	35	19.80	4.92	98.40	8.35	17.50	4.48	<0.00
IPCA	11	12.40	3.08	61.10	5.18	9.10	2.33	<0.00
Residuals	24	7.40	1.84	37.30	3.16	8.50	2.18	
Error	120	13.70	3.40	43.00	3.65	19.30	4.94	
Total	191	402.70		1178.60	100.00	390.70	100.00	

The AMMI biplot and the GGE biplot of yield (**Figure 3.11**) showed that the eight environments (2 years x 4 localities) formed three groups. The low stability, low yielding

environments (00SR and 00VK) with genotypes Baviaans, Kariega and SST825 associated with them are shown in the AMMI analysis (**Figure 3.11 A**). In the GGE biplot genotype, SST825 is associated with no environment (**Figure 3.11B**).

A



B

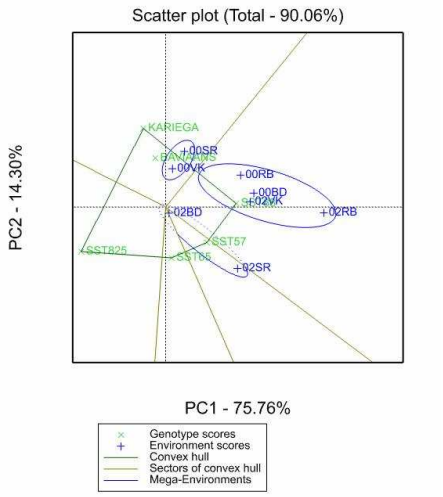
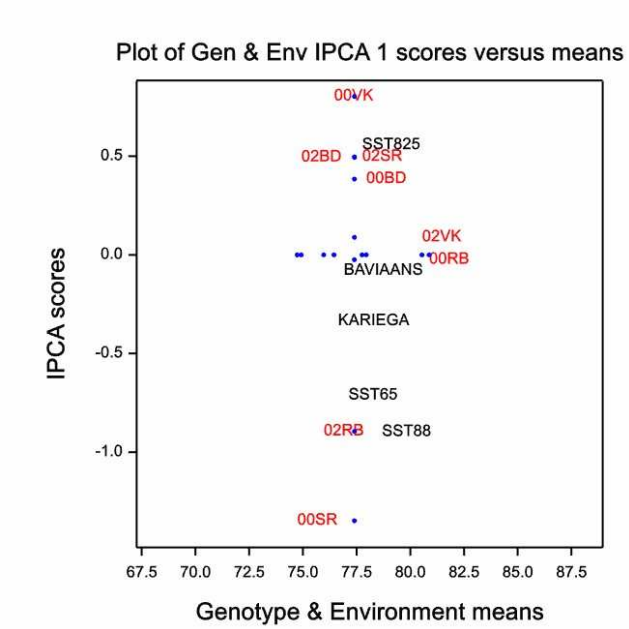


Figure 3.11 Genotype-by-environment biplot of the Rûens region for yield (A) Principal Component I versus mean yield (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

Intermediate yielding to high yielding and stable environments were 00BD, 02VK, 00RB associated with genotypes SST57 and SST65. In the AMMI analysis, environment 02RB was shown as highest yielding associated with genotype SST88. In the GGE biplot it grouped with the higher yielding environments. In the “which-one-wins-where” scenario the eight environments used in this study formed three mega-environments with different “winning” genotypes. This tendency was identified using the scatter plot with the polygon bisects and confidence ellipses (**Figure 3.11B**). Mega-environments are test environments with different winning genotypes located at the vertices on the polygon. Mega-environment I included environments 02BD and 02SR with SST57 as the winning genotype. Mega-environment II included environments 00BD, 00RB, 02RB and 02VK with winning genotype SST88. Mega-environment III was comprised of environments 00SR and 00VK with Baviaans as the winning genotype. The AMMI analysis provided this information in a table called “first four AMMI selections per environment” (data not shown). In this analysis two mega-environments were identified. Mega-environment I included environments 02RB, 02SR, 02VK, 00BD and 00RB with winning genotype, SST88. The second mega-environment consisted of environments 00SR, 02BD and 00VK yield with winning genotype, Baviaans. In retrospect the two analyses were similar.

The AMMI biplot and the GGE biplot of the variable HLM (**Figure 3.12**) showed that the eight environments (2 years x 4 localities) formed two groups, but only one mega-environment. It is interesting to see that the AMMI biplot confirmed the GGE biplot and vice versa. The genotype SST88 is the winning genotype in two environments (00SR and 02RB) and SST825 is winning in three environments (00VK,00BD and 02BD).

A



B

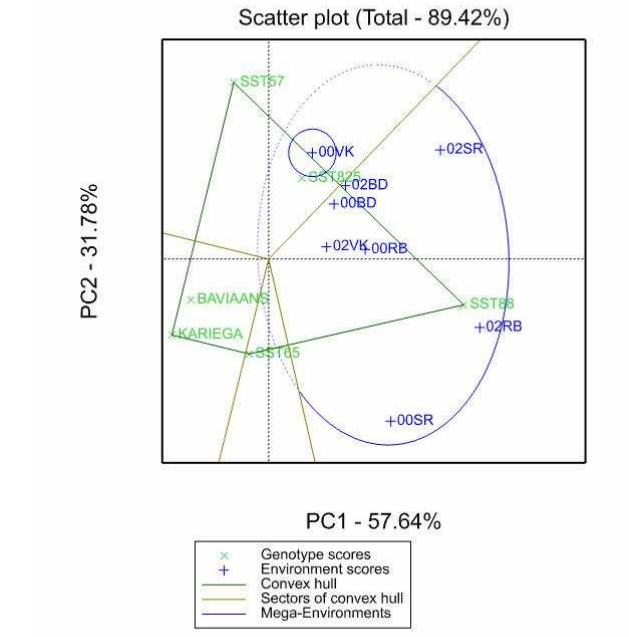
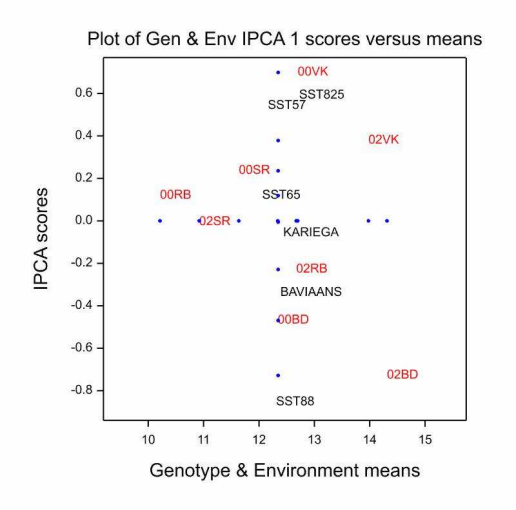


Figure 3.12 Genotype by environment biplot of the Rûens region for hectolitre mass (HLM) (A) Principal Component I versus mean HLM (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

The AMMI and GGE biplots for protein content in the Rûens region (**Figure 3.13**) again showed two groups in one mega-environment. SST825 is the winning genotype in environments 00VK and 02VK. The genotype, SST88, is winning in the environments 00BD and 02BD.

A



B

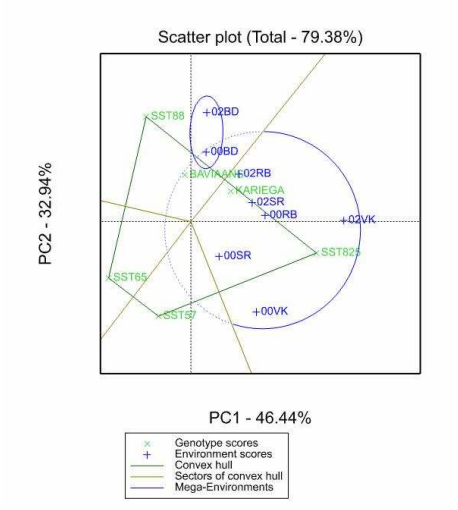


Figure 3.13 Genotype by environment biplot of the Rûens region for protein content (A) Principal Component I versus mean protein content (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

3.4.6.2 Swartland

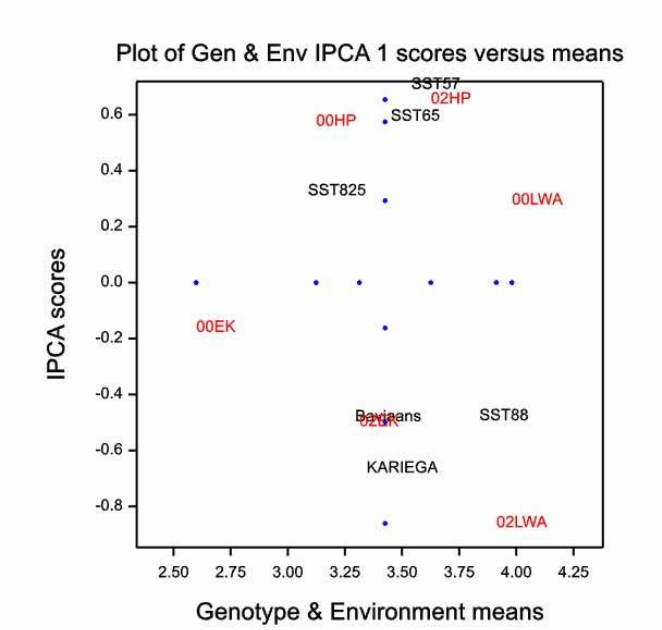
The AMMI analysis of variance for yield, HLM (hectolitre mass), and protein content (Table 3.7) showed that the sum of squares attributed to genotypes were 8.77%, 2.52% and 10.97% respectively. The contribution of environments was 37.39%, 68.96%, and 40.14% respectively. Variance of GEI contributed 23.76%, 9.79% and 12.91% respectively, to sum of squares.

Table 3.7 Sources of variation from the AMMI model of the Swartland cultivar trials

		Yield		HLM		Protein content		
Source	**df**	**SS**	**% SS**	**SS**	**% SS**	**SS**	**% SS**	**P**
Block(Environments)	18	14.78	16.79	124.70	8.45	17.64	12.46	<0.00
Treatments	35	61.55	69.94	1198.90	81.27	90.66	64.02	<0.00
Genotypes	5	7.72	8.77	37.20	2.52	15.53	10.97	<0.00
Environments	5	32.91	37.39	1017.30	68.96	56.85	40.14	<0.00
GxE interaction	25	20.91	23.76	144.40	9.79	18.28	12.91	<0.00
IPCA	9	13.81	15.69	101.70	6.89	8.48	5.99	<0.00
Residuals	16	7.11	8.08	42.70	2.90	9.80	6.92	
Error	90	11.69	13.28	151.60	10.28	33.31	23.52	
Total	143	88.01	100.00	1475.20	100.00	141.62	100.00	

The GGE biplot of yield in the Swartland region (**Figure 3.14**) showed two mega-environments. In mega-environment I (consisting of environments 00HP and 02HP) the winning genotype was SST57. Mega-environment II comprised of environments 00LWA, 02LWA, 00EK, and 02EK and the winning genotypes were Baviaans, Kariega and SST88. The AMMI biplot showed similar results. **Figure 3.15** involves the AMMI and GGE biplots of HLM in the Swartland region. Three mega-environments can be distinguished, namely the environments 02HP and 00EK form mega-environment I, with genotypes SST57 and SST826 competing to be the winner. The second mega-environment comprised of 02LWA and 02EK with Kariega and SST88 as the winning genotypes. The third mega-environment with environments 00LWA and 00HP produced only one winner, genotype SST65. The AMMI and GGE biplots for protein content in the Swartland region (**Figure 3.16**) showed two mega-environments. 00EK and 02EK form mega-environment I with SST825 as the winning genotype. Environments 00LWA, 02LWA, 00HP and 02HP classified as the second mega-environment with three genotypes performing outstandingly.

A



B

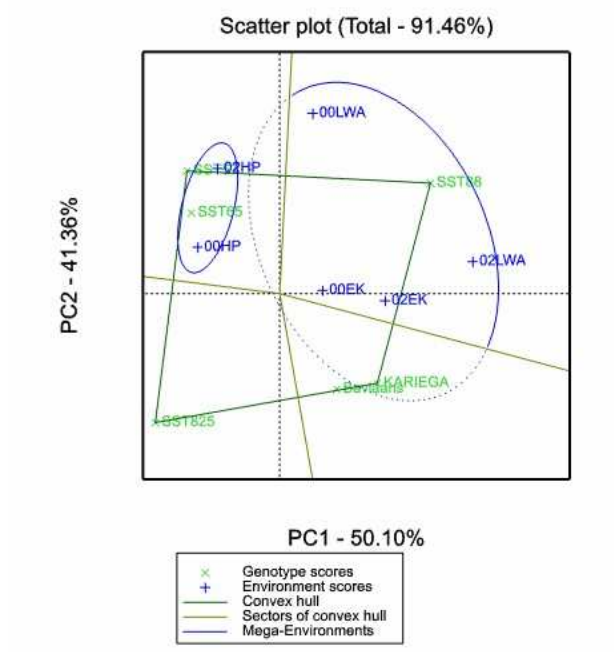
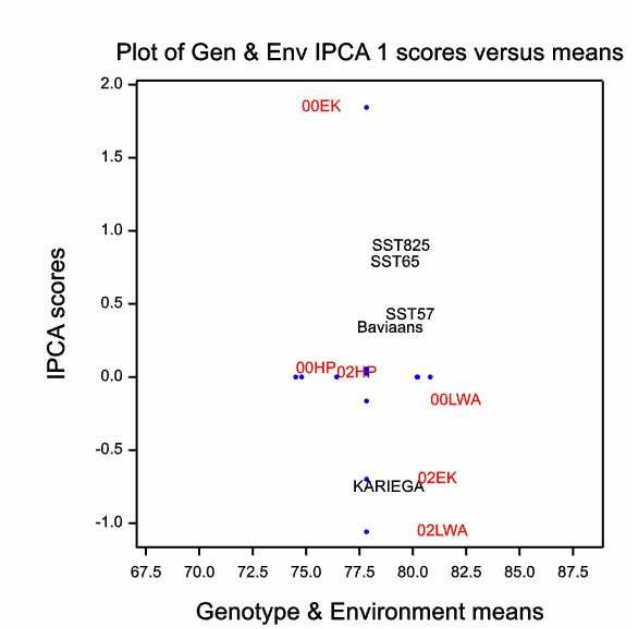


Figure 3.14 Genotype by environment biplot of the Swartland region for yield (A) Principal Component I versus mean yield (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

A



B

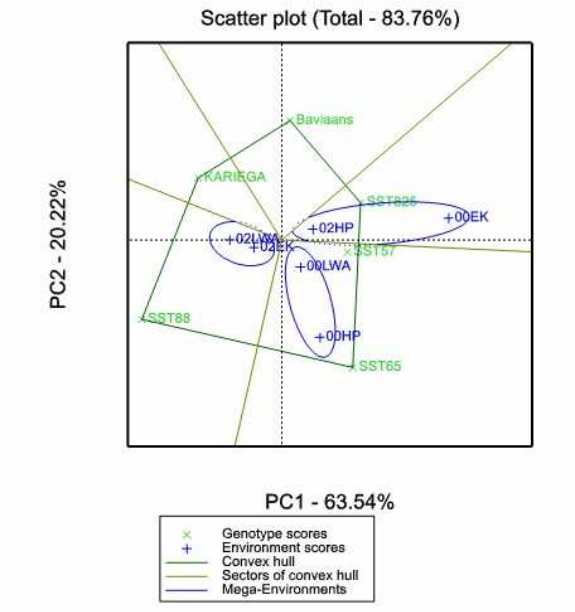
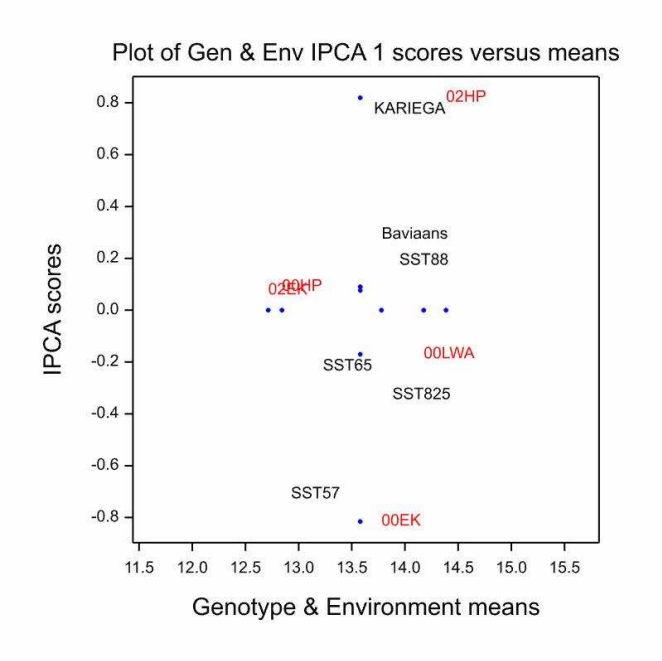


Figure 3.15 Genotype by environment biplot of the Swartland region of hectolitre mass (HLM) (A) Principal Component I versus mean HLM (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

A



B

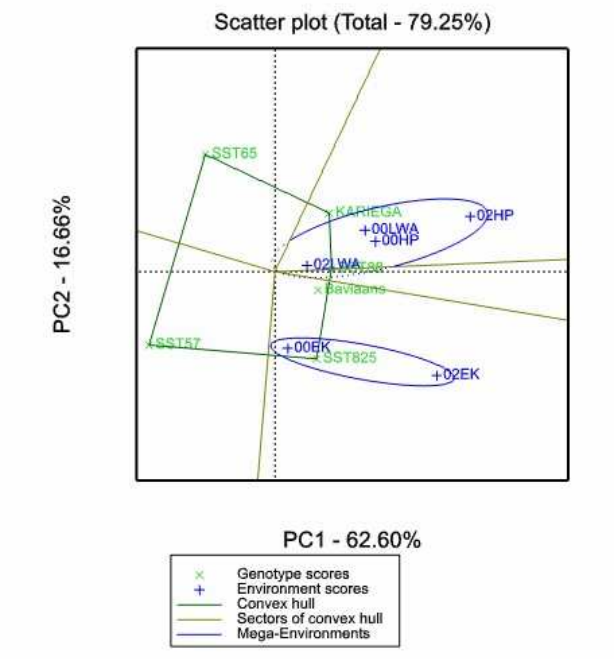


Figure 3.16 Genotype by environment biplot of the Swartland region for protein content (A) Principal Component I versus mean protein content (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

3.4.7 Other statistical techniques

Other statistical techniques were performed on seven years and 30 genotypes in the Rûens region and nine years with 31 genotypes in the Swartland region of the cultivar trials. These techniques were used to provide auxiliary information to the genetic aim of the study. Grain protein is an important factor affecting grain quality and thus a key determinant of both end use and market value in bread wheat. However, it is known that grain protein content is negatively correlated with grain yield in wheat.

Table 3.8 Phenotypic correlation of yield and quality traits for the regions

	Rûens		**Swartland**	
	Yield	**HLM**	**Yield**	**HLM**
Yield				
HLM	-0.40 *		0.45 *	
Protein content	0.09	0.18	-0.29	0.14

*** P≤0.05 HLM = hectolitre mass**.

The Pearson's product moment correlation matrix (**Table 3.8**) showed that yield correlated significantly negative with HLM (r = -0.40, p = 0.05). No correlation between yield and protein content (protein in the table) was found (r = 0.09). The correlation between HLM and protein content was very weak and not significant (r = 0.18) in the Rûens region. The correlations between yield and HLM in the Swartland region was significant and positive (r = 0.45, p = 0.05). The correlation between yield and protein content was negative but not significant (r = -0.29). The correlation between HLM and protein content was weak and not significant (r = 0.14).

It should have been the ultimate to explore the coinciding relationship of the three variables and genotypes with localities, but the biplots of the PCA were cluttered. Thus the results are given on the relationship among the variables and the genotypes**.**

Table 3.9 Loadings of the variables onto the first two principle components for the genotypes

	Rûens		Swartland	
	PC1	**PC2**	**PC1**	**PC2**
Yield	**0.63**	0.15	**0.82**	0.02
HLM	**0.74**	0.02	**0.56**	0.30
Protein	0.04	**0.87**	0.11	**0.81**

Values in bold correspond for each variable to the principal component (PC) for which the loading is the largest

Table 3.9 presents the correlation (loading) of each variable with the first two principal components (PC) for the regions respectively. For the Rûens region yield and HLM were strongly correlated to PC1, which accounts for 47.10% of the total variation of 82.69%. Protein content correlated strongly to PC2, which accounts for 35.59% of the variation. The relationship among the variables and genotypes is shown in **Figure 3.17**.

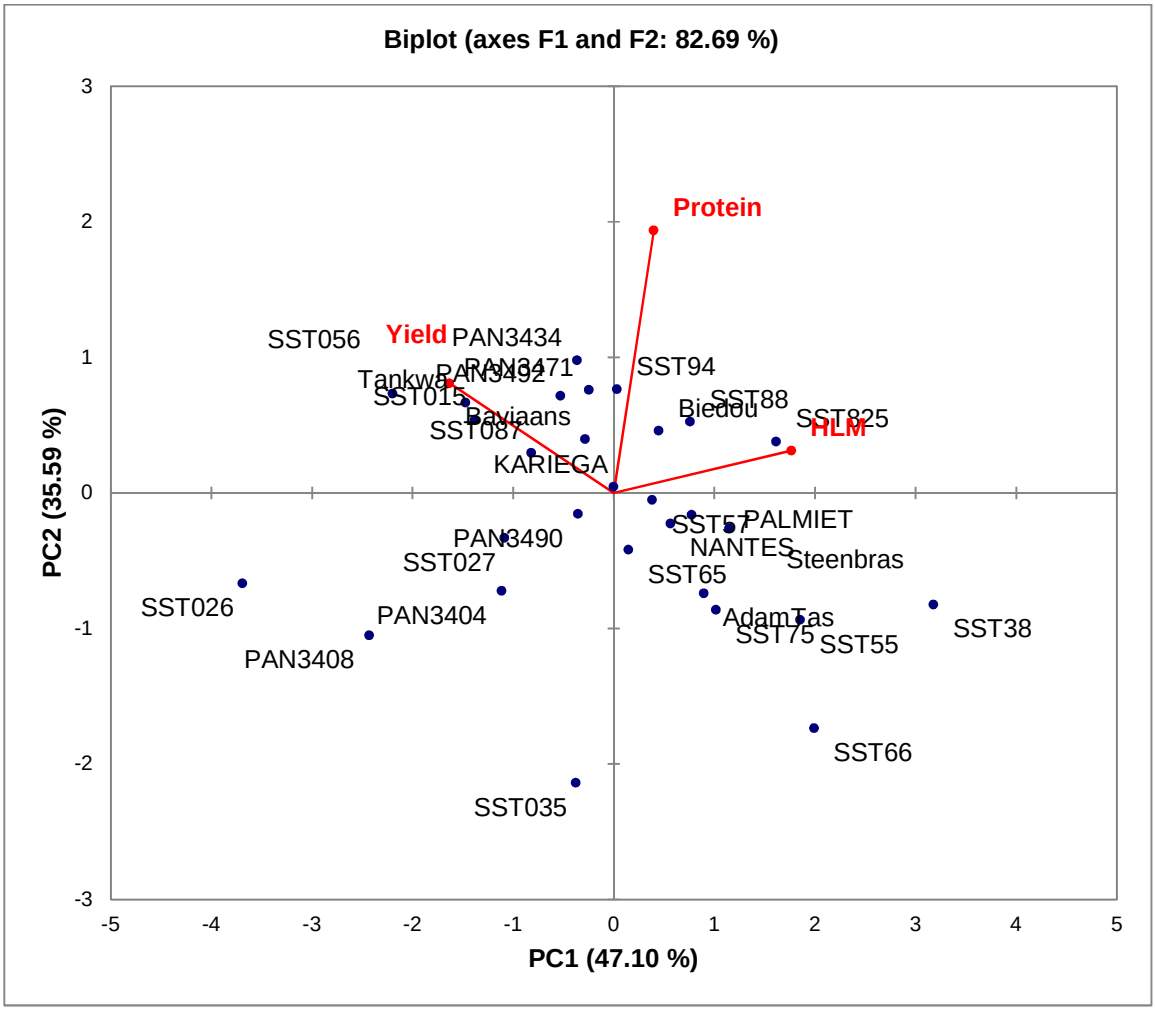


Figure 3.17 PCA biplot of relationship between all variables and the 30 genotypes for the Rûens region.

The genotypes, Kariega and Palmiet showed strong correlation with PC1. They are ranked as intermediate performers due to the negative correlation between yield and HLM. SST026 had the highest yield but showed low HLM values. Genotype SST66 was the best performer for HLM but had the lowest yield. PAN3434 showed the second highest yield but a relatively low HLM value. Genotype SST94 had the highest protein content value. This is shown in the PCA biplot.

In the Swartland region yield and HLM were strongly correlated to PC1 (**Table 3.9**) which accounts for 49.50% of the total variation of 87.04%. Protein content was strongly correlated to PC2. PC2 accounts for 37.54% of the variation (**Figure 3.18**).

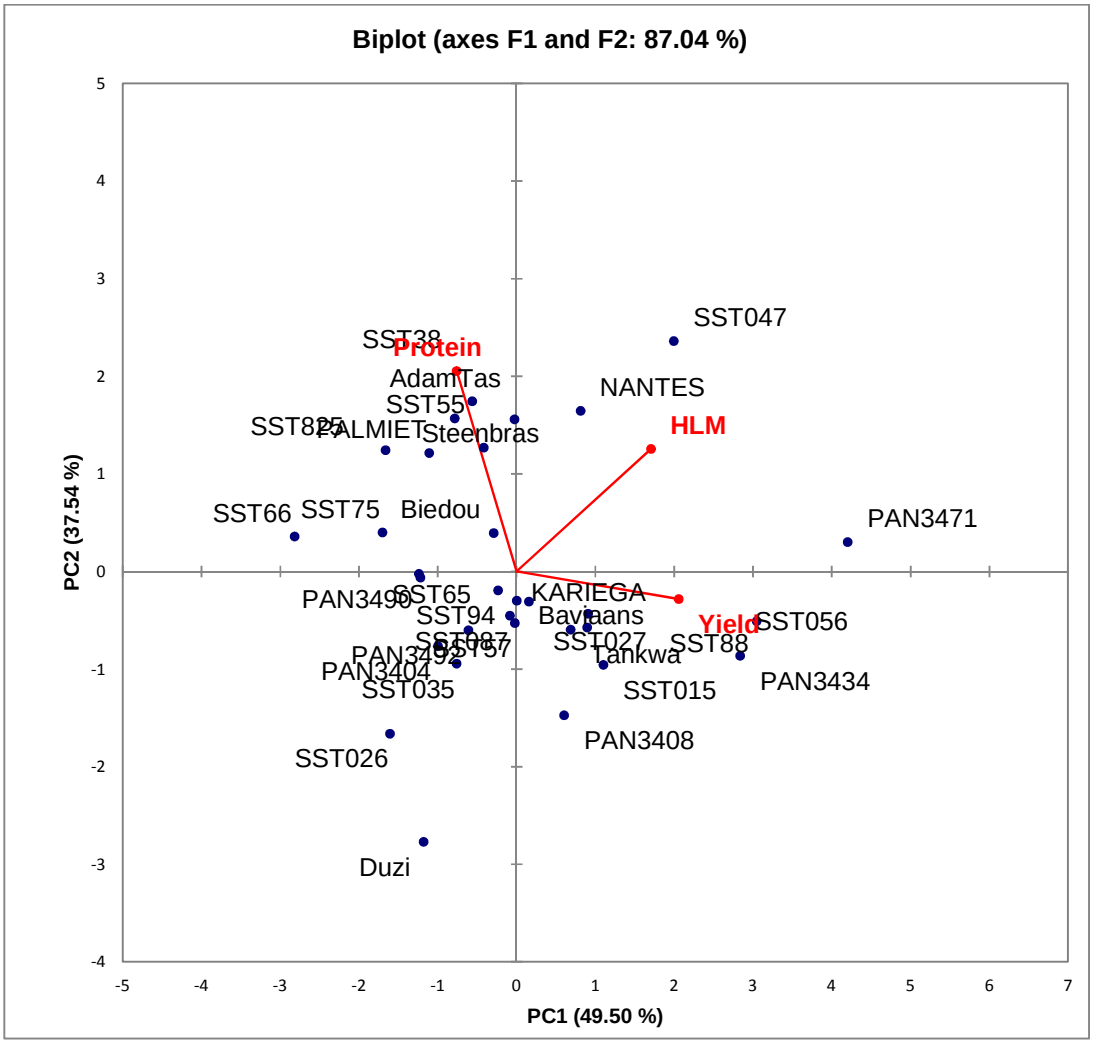


Figure 3.18 PCA biplot of relationship between all variables and the 30 genotypes for the Swartland region.

The relationship of the variables in the PCA of the Swartland region (**Figure 3.18**) differed from the Rûens region. The correlation between yield and HLM in the Swartland was positive and significant, and between yield and protein content negative.

The genotypes SST38, AdamTas and SST55 showed strong association with protein content. This was verified by looking at the means of the genotypes (means not shown). The correlation between yield and HLM was positive and significant. Genotype SST056 was the highest yielder and had a relatively high HLM. Genotype PAN3471 showed the highest HLM but moderate yield. Genotype SST047 was the second best performer for HLM and

performed well in protein content. Genotypes Duzi, SST026 and SST94 had high protein content, but low HLM and yield.

Linear Discriminant Analysis (DA) was used to determine whether the variables could provide additional information on genetic advance in this study. The same genotypes within a year were planted at the localities within a region. The objective was to determine whether the three variables can discriminate among years.

The first two canonical variates (called factors) accounted for 95.16% and 91.42% of the variation among the years in the two regions respectively. For the Rûens region, the plot of the second canonical variate (F2) versus the first factor (F1) is shown in **Figure 3.19**. In this plot the genotypes of a year were enclosed by 95% confidence ellipse. Overlapping of the ellipses of the years indicates no differences among years for the three variables. This plot indicates two distinct groups of years. Year 2005 was depicted as a group on its own. The separation was on F1, which accounted for 70.49% of the variation in the data. Yield correlated strongly negative (-0.97) with F1 and is thus most likely responsible for the grouping. The genotype means of 2005 was much lower than the other years (data not shown). There were two genotypes not included in the eclipses: one near 2005 (SST015) and one near 2010 (SST88). These genotypes did not perform similarly to the group but were not seen as outliers in the diagnostic tests.

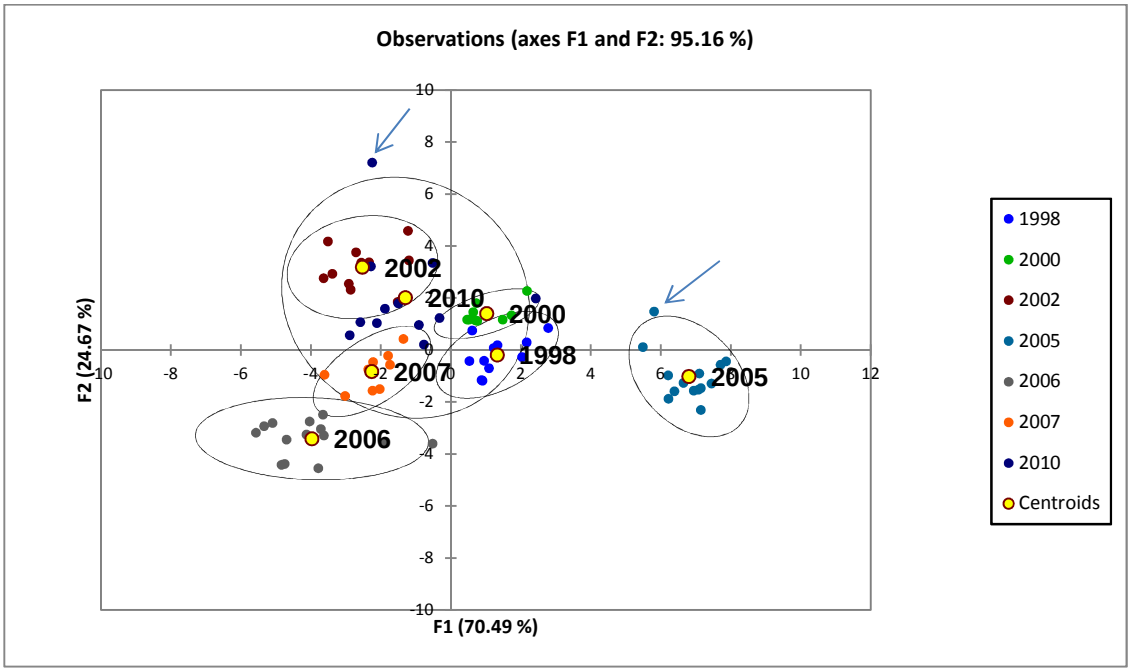


Figure 3.19 Linear discriminant biplot of the years of the Rûens region.

Three groups were visible in the DA biplot for the Swartland region (**Figure 3.20**), namely year 2008, year 2009, the years 1998-2007, and 2010. The second factor (F2) was responsible for the separation of the years into three groups. The correlation of yield (0.84) and protein content (-0.71) with F2 was responsible for the separation. Genotypes of 2009 showed the highest yield of all years but moderate protein content. The mean yield and protein content of the genotypes in 2008 were moderate (data not shown). Genotype SST015 is probably the “outcast” marked with the arrow on the biplot with the highest yield mean and low protein content mean.

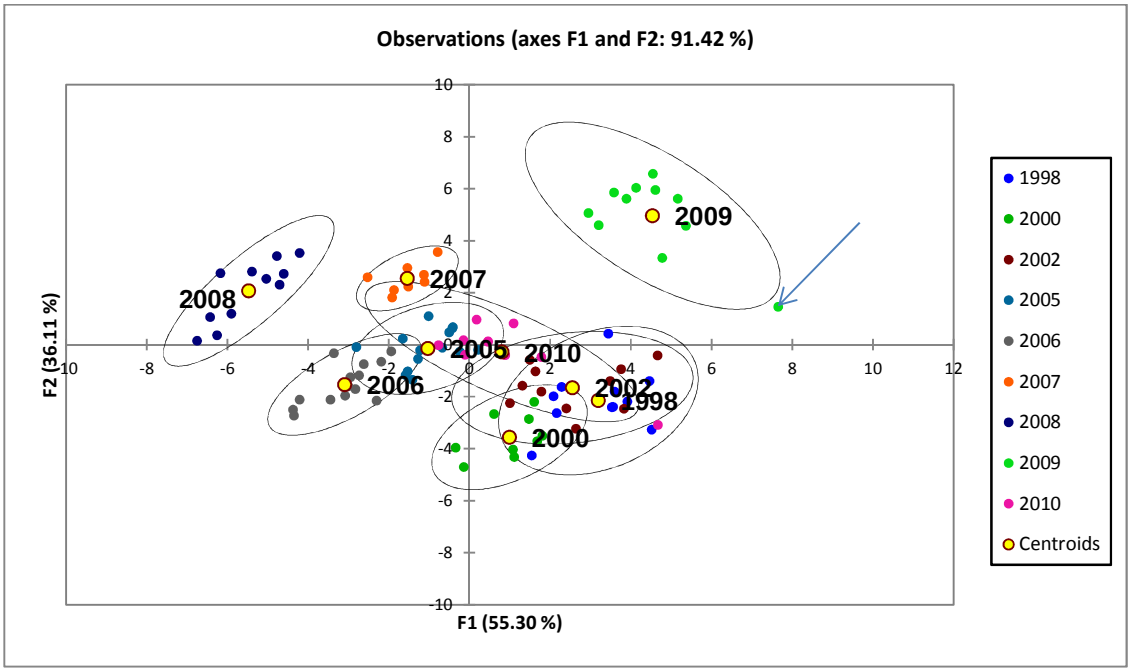


Figure 3.20 Linear discriminant biplot of the years of the Swartland region.

3.5 Conclusions and recommendations

The genetic advance estimates from the various models differed largely. The genetic advance and heritability estimates of model 4 are inflated.

For the elite trials (both regions) TRET provided a significant regression coefficient and the yield progress was 1% per year for the elite trials in both regions. This is in accordance with the world-wide trend. Genetic progress could not be predicted by TRET for HLM. In contrast TRET predicted a genetic advance in protein content of 1% for cultivar trials in both regions.

The complementary nature of the AMMI and GGE models as shown in the biplots is an advantage. The principal component analysis and discriminant analysis provided additional information. It depends on the objectives of the study which methods should be used for determining genetic advance and modelling GEI. Using multivariate techniques like PCA or DA provides auxiliary information in the holistic picture of genotype selection.

References

Allard RW (1960) Principles of plant breeding. Wiley, New York

Annicchiarico P (2002) Genotype x environment interactions: Challenges and opportunities for plant breeding and cultivar recommendations. FAO Plant production and protection papers. http://www.fao.org/DOCREP/005/Y4391E/y4391e05.htm. Accessed 13 June 2010

Barnard AD, Labuschagne, MT, van Niekerk HA (2002) Heritability estimates of bread wheat quality traits in the Western Cape province of South Africa. Euphytica 127:115-122

Bello OB, Ige SA, Azeez MA, Afolabi MS, Abdulmaliq SY, and Mahamood J (2012) Heritability and genetic advance for grain yield and its component characters in maize (*Zea mays* L.). International Journal of Plant Research 2:138-145

Bilgin O, Korkut K, Baser I, Daglioglu O, Ozturk I, Kahraman T, Balkan A (2011) Genetic variation and interrelationship of morpho-physiological traits in durum wheat (*Triticum durum* (L.) Desf.). Pakistan Journal of Botany 43:253-260

Castillo D, Matus I, Del Pozo A, Madariaga R, Mellado M (2012) Adaptability and genotype x environment interaction of spring wheat cultivars in Chile using regression analysis, AMMI, and SREG. Chilean Journal of Agricultural Research 72:167-174

Comstock RE, Moll RH (1963) Genotype-environment interactions. In Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. National Academy of Science, Washington, pp 164-196

Cooper M, Hammer G (1996) Plant adaptation and crop improvement. CABI Publications, Oxford

Department of Agriculture (2012) Trends in the agriculture sector in 2012. In Agricultural statistics. Pretoria, pp 1-111

DePauw RM, Knox RE, Clarke FR, Wang H, Fernandez MR, Clarke JM, McCaig TN (2007) Shifting undesirable correlations. Euphytica 157:409-415

De Vita P, Mastrangelo AM, Matteu L, Mazzucotelli E, Virzi N, Palumbo M, Lo Storto M, Rizzi F, Cattivelli L (2010) Genetic improvement effects on yield stability in durum wheat genotypes grown in Italy. Field Crops Research 119:68-77

Eid MH (2009) Estimation of heritability and genetic advance of yield traits in wheat (*Triticum aestivum* L.) under drought conditions. International Journal of Genetics and Molecular Biology 1:115-120

Gauch HG, Piepho HP, Annicchiarico P (2008) Statistical analysis of yield trials by AMMI and GGE: further considerations. Crop Science 48:866-889

Gauch HG, Zobel RW (1996) AMMI analysis of yield trials. In Kang MS, Gauch HG (eds) Genotype by environment interaction. Boca Raton CRC, New York, pp 85-122

Gilmore A (2010) Why use BLUPS? An introduction to plant breeders. http://mmontap.org/sites/default/files/workingPapers/Blup.pdf. Accessed 2 February 2012

Gregory RJ (2000) Psychological testing: History, principles and applications, 3rd ed. Allyn & Bacon, Boston

Holland JB, Nyquist WE, Cervante-Martinez CT (2003) Estimating and interpreting heritability for plant breeders: An update. Plant Breeding Reviews 22:9-112

Hu X, Spilke J (2011) Variance-covariance structure and its influence on variety assessment in regional crop trials. Field Crops Research 120:1-8

Koekemoer FP (2003) Genetic variability of South African spring wheat for milling and baking characteristics. Dissertation, University of the Free State

Levene H (1960) Robust test in the equality of variance. In Olkin I (ed) Contributions to probability and statistics: Essays in honor of Harold Hotelling. Stanford UP, Stanford, pp 278-292

Miliken GA, Johnson DE (1984) Analysis of Messy Data Volume 1: Designed Experiments. Lifetime Learning Publications, Belmont

Morgounov A, Zylkin V, Belan I, Roseeva L, Zelensky Y, Gomez-Becerra HF, Budak H, Bekes F (2010) Genetic improvements for grain yield in high latitude spring wheat grown in western Siberia in 1900-2008. Field Crops Research 117:101-111

Morris CF, Campbell KG, King GE (2004) Characterization of the end-use quality of soft wheat cultivars from the eastern and western US germplasm pools. Plant Genetic Resources 2:59-69

Mut Z, Aydin N, Bayramoglu, HO, Ozcan H (2010) Stability of some quality traits in bread wheat (*Triticum aestivum*) genotypes. Journal of Environmental Biology 31:489-495

Pakendorf KW (1977) A study on the efficiency of current methods of breeding and testing for wheat improvement in the Western Cape Province. Dissertation, Stellenbosch University, pp 1-190

Payne R, Murray DA, Harding SA, Baird DB, Soutar D (2012) GenStat for Windows: Introduction, 15th edition. VSN International, Hemel Hempstead

Peltonen-Sainio P, Jauhiainen L, Laurila IP (2009) Cereal yield trends in northern European conditions: Changes in yield potential and its realisation. Field Crops Research 110:85-90

Piepho HP, Mohring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177:1881-1888

Rodrigues O, Lhamby JCB, Didonet AD, Marchese JA (2007) Fifty years of wheat breeding in Southern Brazil: yield improvement and associated changes**.** Pesquisa Agropecuária Brasileira 42:817-825

SAS Institute (2012). The Statistical Procedure manual. North Carolina 27513: SAS Campus Drive, Cary

Sener O, Arslan M, Soysal Y, Erayman M (2009) Estimates of relative yield potential and genetic improvement of wheat cultivars in the Meditterean region. Journal of Agricultural Science 147:323-332

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52:591-611

Smit HA, Tolmay VL, Barnard A, Jordaan JP, Koekemoer FP, Otto WM, Pretotius ZA, Purchase JL, Tolmay JPC (2010) An overview of the context and scope of wheat (*Triticum aestivum*) research in South Africa from 1983 to 2008. South African Journal of Plant and Soil 27:81-96

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. Journal of Agricultural Science 143:449-462

So Y (2009) Prediction of cultivar performance and heterogeneity of genotype variance, correlation and error variance in Iowa crop performance test in corn (*Zea mays*, L.). Dissertation, Iowa State University.

Tolmay JPC (2008) Morphological and physiological responses of spring wheat (*Triticum aestivum* L.) to spatial arrangements. Dissertation, Stellenbosch University, pp 1-264

Trethowan RM, Van Ginkel M, Rajaram S (2002) Progress in breeding wheat for yield and adaptation iun global drought affected environments. Crop Science 42:1441-1446

Underdahl JL, Mergoum M, Ransom JK, Schatz BG (2008) Agronomic traits improvement and associations in hard red spring wheat cultivars released in north Dakota from 1968 to 2006. Crop Science 48:158-166

Van Eeuwijk FA, Cooper M, DeLacy IH, Ceccarelli S, Grando S (2001) Some vocabulary and grammar for the analysis of multi-environment trials, as applied to the analysis of FPB and PPB trials. Euphytica 122:477-490

Yan W, Cornelius PL, Hunt LA (2001) Two types of GGE biplots for analysing multi-environment trial data. Crop Science 41:643-663

Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007) GGE biplot analysis of genotype-by-environment data. Crop Science 47:643-655

XLStat (2012) XLStat. Addinsoft, Paris

Chapter 4
THE DRYLAND FREE STATE

Abstract

Winter and intermediate type bread wheat cultivars were evaluated for genetic improvement under dryland conditions over a 16 year period (1995-2010) in 15 localities in the eastern, central and western Free State wheat production areas of South Africa. The following statistical analyses and procedures were conducted to determine genetic improvement: (i) linear regression; (ii) a variance component method according to Comstock and Moll; (iii) a variance component method according to Allard. The proposed method for determining genetic advance over time was Trethowan (TRET), models 1 and 6. No significant trend was observed in the elite trials with TRET. Yield showed 0.5% and 0.6% improvement for the two planting dates in the cultivar trials of the eastern region. Yield progress in the central Free State was 0.3% for both planting dates. The effects of GEI on yield and quality traits were studied by comparing the additive main effects and multiplicative interaction (AMMI) and genotype plus genotype-by-environment interaction (GGE) biplot analysis. Application of AMMI and GGE biplots facilitated visual comparison and identification of superior genotypes for each target set of environments. The complementary nature of the models, as shown in the biplots, is an advantage. Cluster analysis (CA), principal component analysis (PCA) and linear discriminant analysis (DA) provided additional useful information.

4.1 Introduction

Dryland wheat production in the Free State province of South Africa (SA) accounts for 29% of the mean national wheat crop of 1.8 million metric ton (Department of Agriculture, 2012). However, this contribution to wheat production in SA varies substantially due to the unpredictable nature of production. Intermediate and winter genotypes are planted in autumn (April/May) and winter (June/July) on residual soil water conserved during the summer rainfall months. Wheat in the Free State is dependent on highly variable spring rainfall in September and October to ensure economic yield. The considerable variation in soil and climate has resulted in significant variation in yield performance of wheat cultivars

(Purchase *et al*., 2000). It is thus essential that yield performance measured by yield improvement be determined in order to make specific recommendations to producers.

Improvement in grain yield is the primary breeding goal of all wheat breeding programmes in SA and elsewhere. Wheat breeding efforts in the Free State began at the beginning of the twentieth century. These efforts have allowed wheat producers in the region to achieve remarkable increases in grain productivity.

Periodic evaluation of the rate of genetic gain in grain yield is important, as world population growth remains unchecked; in SA the population surpassed the 50 million mark in 2012 (Department of Agriculture, 2012). An assessment of recent trends in genetic improvement of wheat yields is vital to determine whether increased consumption demands of an ever-growing SA population will be met.

Genetic yield gains have historically been estimated using two principal approaches. Firstly, analysis of yields in uniform regional nurseries can be used to evaluate breeding progress as these nurseries contain the best material in the breeding programmes at the time of their testing. Relative genetic gains over time can be estimated by comparing yield of long-term checks with new experimental lines (Graybosch and Peterson, 2010). These authors reported a wheat yield gain of 1.3% per year in the Southern Plains Regional Nursery and 0.79% per year in the Northern Plains Regional Nursery. Austin *et al*. (1989) documented genetic gain of 0.40% per year from 1912 to 1980 and 1% per year for that period in the winter wheat trials of England.

The second approach is based on a long-time series of yield data mostly coming from long-term experiments compared to a historic check variety. Cox *et al*. (1988) used this procedure to demonstrate a 1% per year genetic gain for grain yield in hard red winter wheat, relative to the keystone variety Turkey, over the time period 1919 to 1987. Donmez *et al.* (2001) estimated genetic gain for yield at only 0.44% per year in hard winter wheat, also relative to Turkey, over the time period 1873-1995. Fufa *et al.* (2005) used a similar approach with Nebraska-adapted winter wheat and arrived at a comparable estimate for genetic gain in grain yield of 0.48% per year.

In SA Van Lill and Purchase (1995) used data from the dryland winter regional trials conducted from 1930 to 1980 to demonstrate a 1.36% genetic improvement in grain yield of the advanced breeding lines. Van Niekerk (2001) recorded an incremental yield increase of 1.31% per year from 1979-2001 without the loss of quality or quantity of protein.

The objectives of the present investigation were to:

i. evaluate the wheat yield improvements achieved over the last 16 years (1995-2010) through the breeding programmes by various biometrical techniques;
ii. determine the trends of yearly yield by regression methods and other biometrical techniques;
iii. demonstrate the direction of yield progress during the last 16 years by different biometrical/statistical techniques;
iv. compare the AMMI and the GGE analyses in assessing GEI for yield and the two quality traits;
v. study the relationship between wheat grain yield and quality traits by different statistical techniques.

In order to achieve these objectives for the dryland Free State, breeding and cultivar data from the ARC-SGI cultivar development programmes from four distinct Free State production regions were used. These trials took place from 1995-2010. The production regions were eastern, central, western and southern Free State.

4.2 Materials

Winter wheat is planted in the Free State under dryland conditions on stored moisture that has accumulated during the preceding summer and autumn. Suitable production areas depend on soil type, soil water status and onset of the spring and summer rains. Planting date varies over a period of four months, leaving room for varietal development including a range of maturities and adaptation characteristics.

Varieties differ in adaptation due to different factors, including planting dates and production potential. Varietal recommendations take this into account. The Free State is divided into four relatively homogeneous wheat production areas, namely:

i. The north western Free State with low mean rainfall and deep yellow sandy loam soils. A part of the North West Province adjacent to this region is included in this area.
ii. The south western area with warmer spring temperatures, low mean rainfall and deep red sandy loam soils.
iii. The eastern area with cooler temperatures, as well as higher and more stable rainfall.
iv. The central area with moderate temperatures and rainfall, and a lower evaporation requirement (ARC-SGI Report, 2010).

4.2.1 Elite field trials

For this study, data for genotypes from the ARC-SGI elite intermediate and winter wheat yield trials in the Free State were used. These trials were conducted in three of the four geographical regions, namely eastern, central and western Free State. Data was screened for continuity over 15 years (1995-2010). Due to the lack of consistent trials in localities and years consequently only data of the intermediate eastern trials could be used. The data of all three regions were valid in the winter wheat trials. The geographical position of the localities is depicted on the map in **Figure 4.1**.

In the eastern, central and western regions (**Table 4.1**) there were a total of 295, 304 and 395 genotypes inspected over the 15 years respectively and each trial included the same ARC-SGI check. However, during 1995-1997 a different check was used and these years were subsequently excluded from the study. Standard cultivation practices were followed except for Bethlehem where two trials, namely sprayed against aphids and an unsprayed trial, were conducted. The experimental layout was a randomised complete block design (RCBD) with four replications.

Table 4.1 Listings of the ARC-SGI Elite Intermediate and Winter Wheat Yield Trials in the Free State

Eastern Free State	Central Free State	Western Free State
Bethlehem (Unsprayed)	Bloemfontein	Bultfontein
Bethlehem (Sprayed)	Henneman	Wesselsbron
Clarens		
Clocolan		
Kransfontein		
Meets		
Petrus Steyn		
Reitz		
Tweespruit		

4.2.2 Cultivar field trials

Data for genotypes from the ARC-SGI cultivar wheat yield trials in the Free State were used in this study. These trials were subdivided in four geographical regions namely central, eastern, north western and southern Free State (**Table 4.2**). The geographical position of the localities is depicted on the map (**Figure 4.1**). Each region was further divided according to different planting dates and the data was screened for continuity over 12 years (1998-2010). Consequently only two planting dates could be incorporated in this study. Data of these two planting dates were analysed separately. Although the same cultivars were used in all the localities, not all cultivars were equally represented and each trial included the same ARC-SGI check. Standard cultivation practices were followed in all of the regions. The experimental layout was a RCBD with four replications.

Table 4.2 Listings of the ARC-SGI Cultivar Trials in the Free State

Central Free State	Eastern Free State	Western Free State	Southern Free state
Arlington	Bethlehem	Bultfontein	Hebron
Excelsior	Clarens	Henneman	Petrusburg
Senekal	Clocolan	Wesselsbron	Tweespruit
	Ficksburg		
	Harrismith		
	Ladybrand		
	Reitz		

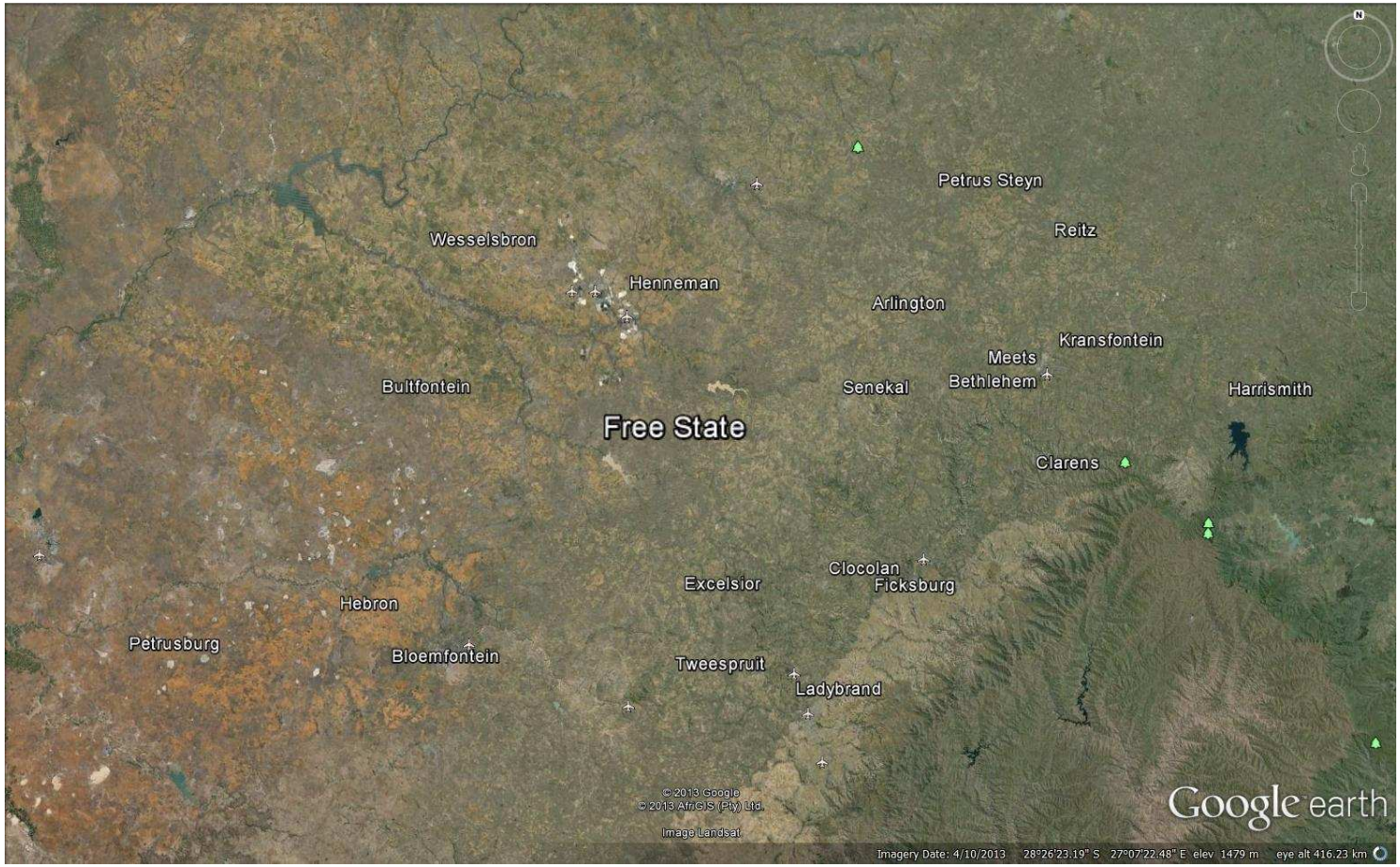


Figure 4.1 Google Map of the localities in the Free State.

4.3 Statistical techniques

Three statistical methods were compared to estimate genetic advance namely:

i. Linear regression over years described by Trethowan *et al*. (2002);
ii. Variance component method described by Comstock and Moll (1963);
iii. Variance component method described by Allard (1960).

Several other statistical techniques were used to investigate the relationship among the different factors and/or variables:

iv. Genotypes and environments ─ AMMI and GGE biplot analyses were performed to evaluate the GEI and to compare the two methods for a single locality. Years within the locality were used as environments;
v. Pearson's product moment correlation was performed to determine the relationship among the three variables;
vi. PCA was performed to investigate the relationship between the factors (i.e. genotypes) and the variables;
vii. CA was used to find grouping among the environments;
viii. DA was used to determine whether the three variables could discriminate among the years.

4.3.1 The linear regression method over years (TRET)

The first method to determine genetic improvement was a linear regression method described by Trethowan *et al*. (2002). The statistical procedure of this method was described in Section 3.3.1. The graphic presentations were done with Excel to provide a trend with R^2 values. The p values on the graphical presentations were provided by the regression generated by the PROC REG procedure of the software SAS (SAS Institute, 2012).

4.3.2 Sources of variation, heritability and genetic advance

Other methods to determine genetic advance are those using sources of variation. This can either be fixed effects models or random effects models. In **Figure 3.2** the layout of the

different methods to calculate genetic advance is shown. The statistical procedure of this method was described in Section 3.3.3.

4.3.2.1 Expectations of mean squares from a fixed model (Model 1)

The data was subjected to an ANOVA and analysed by the PROC GLM procedure of the SAS software (SAS Institute, 2012). Variance components were estimated using TYPE III expected mean squares of the ANOVA as pointed out in **Table 3.2** and discussed in Section 3.3.4

.

4.3.2.2 Linear mixed model framework to calculate variance components and genetic advance (Model 2 to Model 5)

The statistical procedure of this method was described in Section 3.3.5.

4.3.3 Model 6 – the model proposed by Allard (1960)

The statistical procedure of this method was described in Section 3.3.8.

4.3.4 AMMI versus GGE biplot

Usually a large number of genotypes are evaluated across a number of sites and years, and it is often difficult to determine the pattern of genotypic response across sites (localities) without the help of graphical display of the data (Yan *et al*., 2001). Biplot analysis provides a solution to the above-mentioned problem as it displays GEI data and allows visualisation of the interrelationship among environments, genotypes and the interactions between genotypes and environments. Two types of biplots, the AMMI biplot (Gauch *et al*., 2008) and the GGE biplot (Yan *et al*., 2007), have been widely used to visualise genotype x environment interaction.

4.3.4.1 AMMI analysis

This analysis integrates ANOVA and PCA into a united approach. The significant feature of this analysis is that adjustment is carried out using information from other locations to refine the estimates within a given location and it removes residual or noise variation from GEI (Crossa *et al.,* 1990).

4.3.4.2 GGE biplot analysis

This analysis is a visual representation of the genotype main effect plus the GEI. GGE biplot analysis provides a range of plots that are useful for assessing the performance of genotypes in different environments. The observed phenotypic variation (P) of genotypes across environments is made up of environment variations (E), genotype variations (G) and G x E interaction (GE): i.e. P = E + G + GEI. Usually E is the dominant source of variation, while G and GE are relatively small. Thus it is typical to remove the environmental main effect and focus only on G and GEI (Yan *et al*., 2007).

The AMMI and GGE biplots were constructed using GenStat, 15th edition (Payne *et al*., 2012). AMMI analysis was based on the model by Gauch and Zobel (1996) and GGE was based on the model for two principal components according to Yan and Kang (2003).

4.3.5 Other statistical techniques

The Pearson's product moment correlation matrix of the pairwise correlations among the three dependent variables was built to show their linear relationships, which were then graphically represented using two major components of PCA. Hierarchical cluster analysis of the target environments was performed to identify the similarities among the environments. The dendrogram was constructed using Ward's method with the automatic clustering option. Linear discriminant analysis was performed to see whether the variables could separate the years and if a trend was evident in the separation of the years. These methods were performed using XLSTAT software (XLStat, 2012).

4.4 Results and discussion

4.4.1 Preliminary analyses

Preliminary ANOVA with PROC GLM of the SAS software (2012) using Error Type III was carried out for individual trials (each locality and year) to assess the residuals and predicted values. The residuals were tested by the Shapiro and Wilk (1965) test for normality. Outliers (any residual value above the absolute value of the absolute value of three — ABS(rijkl) > 3) were replaced by the predicted values.

Homogeneity of variances was tested by the Levene's test in the PROC GLM in the SAS programme (Levene, 1960). Homogeneity of variances was also tested in the linear mixed models by PROC MIXED programmes suited for this assumption. Heterogeneity in year variances and the interactions with year, genotype x year, genotype x locations, as well as genotype x year x locations were found.

4.4.2 Linear regression over years using TRET

In this study, mean yield for genotypes in experimental years were used to express genetic progress. A linear regression over time for each locality was performed (results not shown), but to compare methods to determine genetic advance proposed in this study, it was decided to perform the linear regression over localities in a region.

Homogeneity of variances for localities within a region was found. The data of the elite trials and the cultivar trials were subjected to linear regression analysis for each region over time according to Trethowan *et al.* (2002). In both trials non-recurrent genotypes occurred, leading to unbalanced data. The method proposed by Trethowan *et al.* (2002) renders a solution to the unbalanced nature of the data. The mean of the best five genotypes was expressed as a ratio (from now on called the Ratio) of the trial mean (TM). The same check was used throughout the period and for all localities in both the elite and cultivar trials. The modifications described in Section 3.3.1 namely M_T (mean of five highest yielding genotypes minus trial mean), M_C (mean of five highest yielding genotypes minus check mean), would be relevant and was therefore tested.

4.4.2.1 Elite trials

The trial means (TM), the check means (Check) and Ratio across all environments within the regions are presented graphically. It was found that Mc and M_T added no extra value to the study; thus it will not be discussed further.

Ratio in the eastern region for the intermediate trials indicated no significant progress over time (**Figure 4.2**).

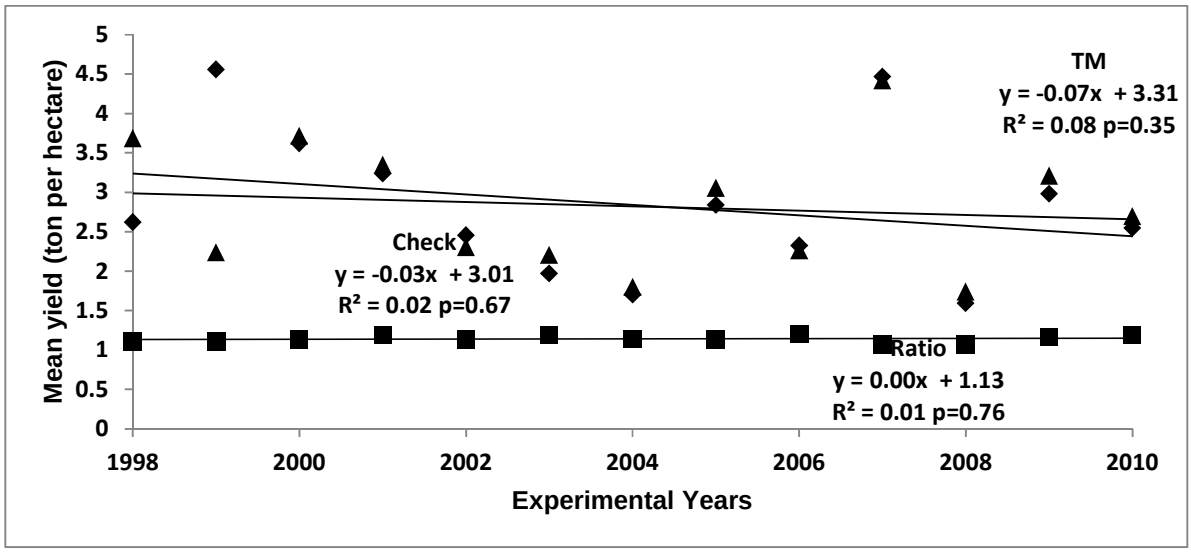


Figure 4.2: Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the intermediate elite trial in the eastern Free State region. ♦ = Trial means (TM); ▲ = Check; ■ = Ratio.

In the winter trial of the eastern region the trial means and check means showed no significant progress (**Figure 4.3**). The ratio showed a significant progress in yield over time (R^2=0.32, p=0.00, b=-0.01). A significant negative trend was observed. In the winter trial of the central region the trial means and ratio showed no significant progress over time (**Figure 4.4**). The check means showed a significant negative trend over time (R^2=0.26, p=0.04, b=-0.07).

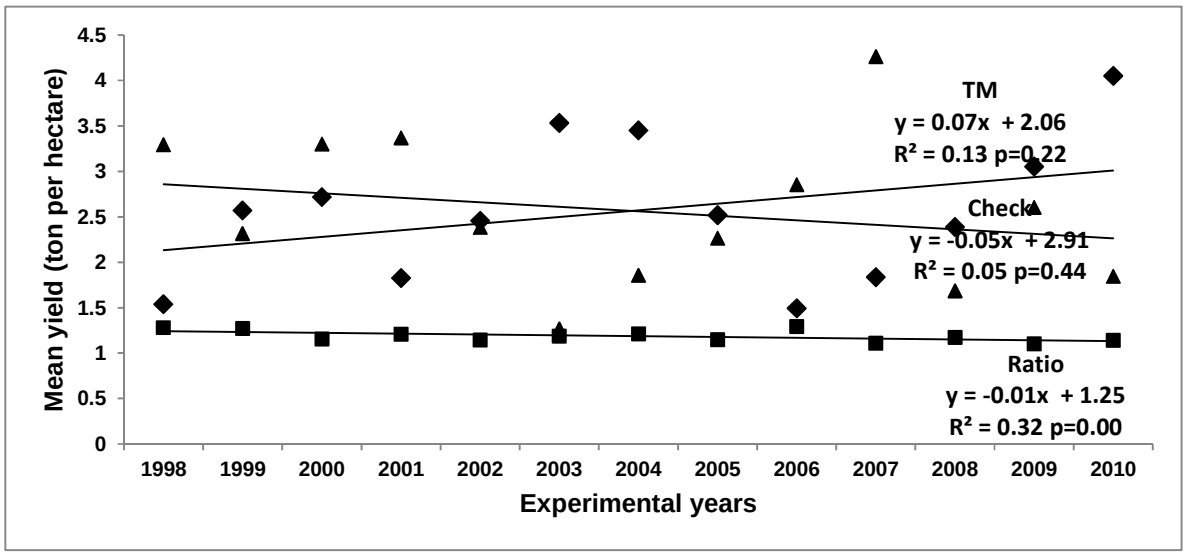


Figure 4.3 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the winter elite trial in the eastern Free State region. ♦ = Trial means (TM); ▲ = Check; ■ = Ratio.

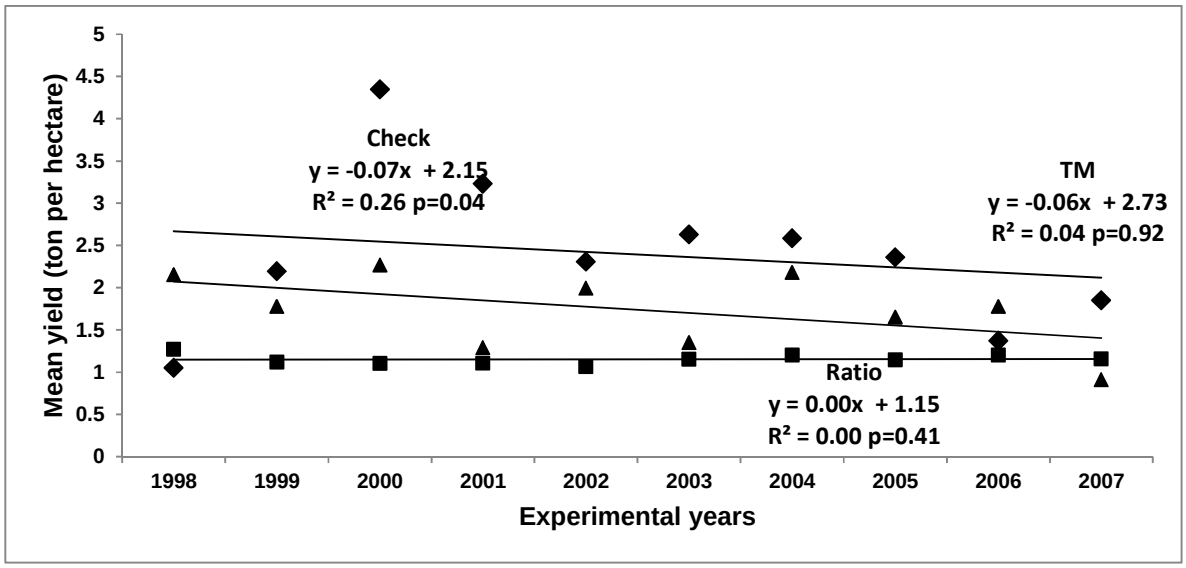


Figure 4.4 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the winter elite trial in the central Free State region. ♦ = Trial means (TM); ▲ = Check; ▪ = Ratio.

Neither ratio, the check or the trial means in the winter trial of the western region showed a significant progress over time (**Figure 4.5**).

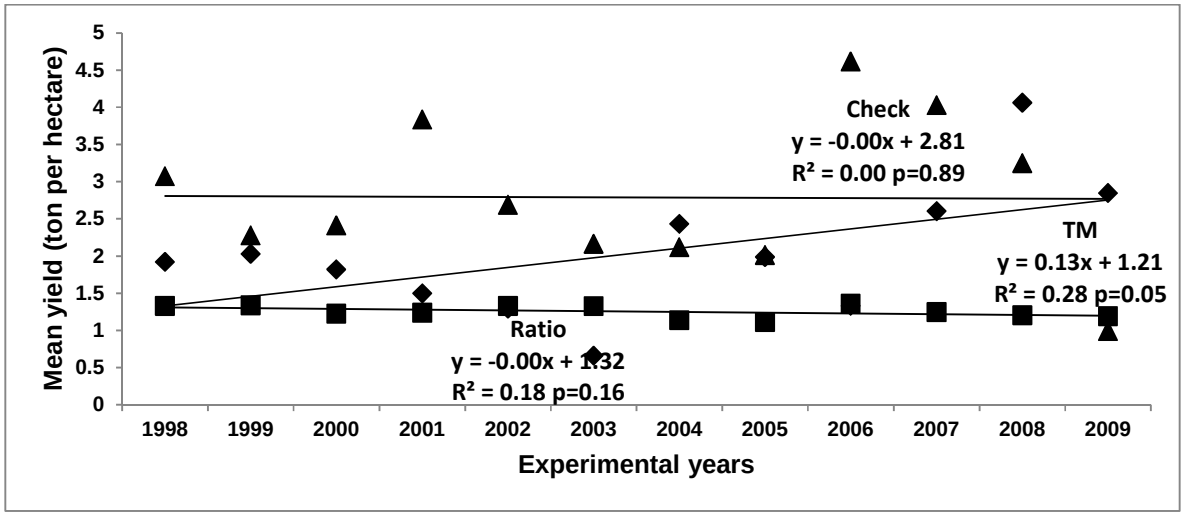


Figure 4.5 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the winter elite trial in the western Free State region. ♦ = Trial means (TM); ▲ = Check; ■ = Ratio.

Although genetic improvement could not be shown in the elite trials, these results were not unique to the research of the breeding trials of the Free State. Graybosch and Peterson (2010) found no significant genetic improvement for the period 1984-2008 in the Southern Regional Performance Nurseries of the Great Plains in North America. Yasin *et al.* (2011) found similar results in their study of wheat yield improvement between 969-2005 in India.

4.4.2.2 Cultivar trials

Using the trial means (TM) and the check mean to measure genetic progress can be misleading as yield is greatly affected by weather patterns in a given year. Thus only Ratio and Mc were graphically presented. Although Ratio showed no significant trend over time, Mc showed a significant trend over time (R^2=0.35, p=0.00) in the eastern Free State at planting date 1 (**Figure 4.6 A**). The slope, b-value = 0.07 (progress =0.07/13 *100) of Mc, indicates a 0.5% progress per year over 13 years.

Similar results were found for the second planting time (**Figure 4.6 B**) with significant genetic progress of 8% over 13 years (R^2=0.32, p=0.04) predicting a 0.6% progress per year.

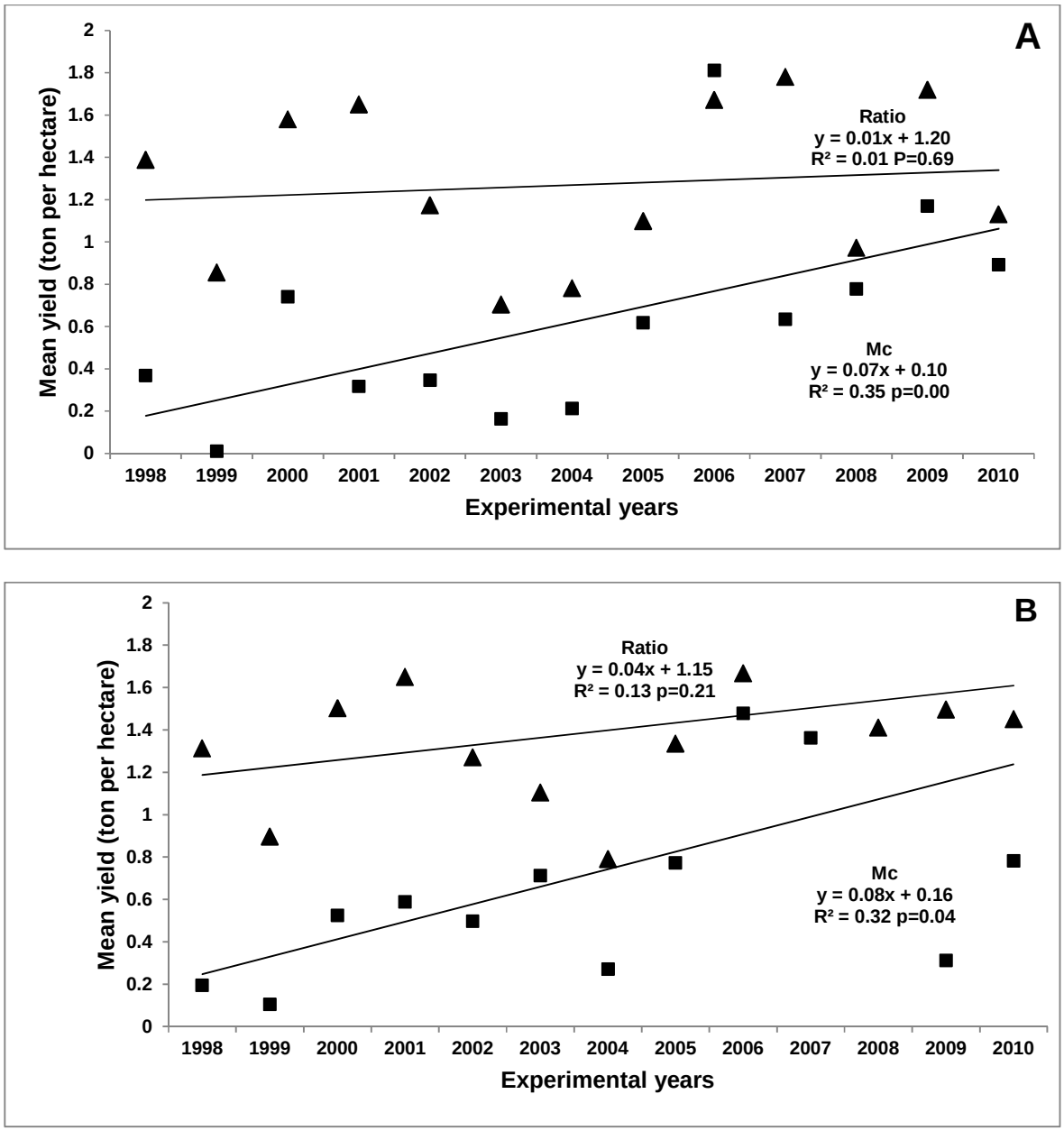


Figure 4.6 The mean yield of Ratio and Mc over sites regressed against experimental years of the cultivar trials in eastern Free State region for (A), the first planting date, and (B), for the second planting date. ▲ = Ratio; ■ = Mc

The progress in mean yield for Ratio and Mc for both planting dates in the north western Free State (**Figure 4.7 A and B**) was not significant.

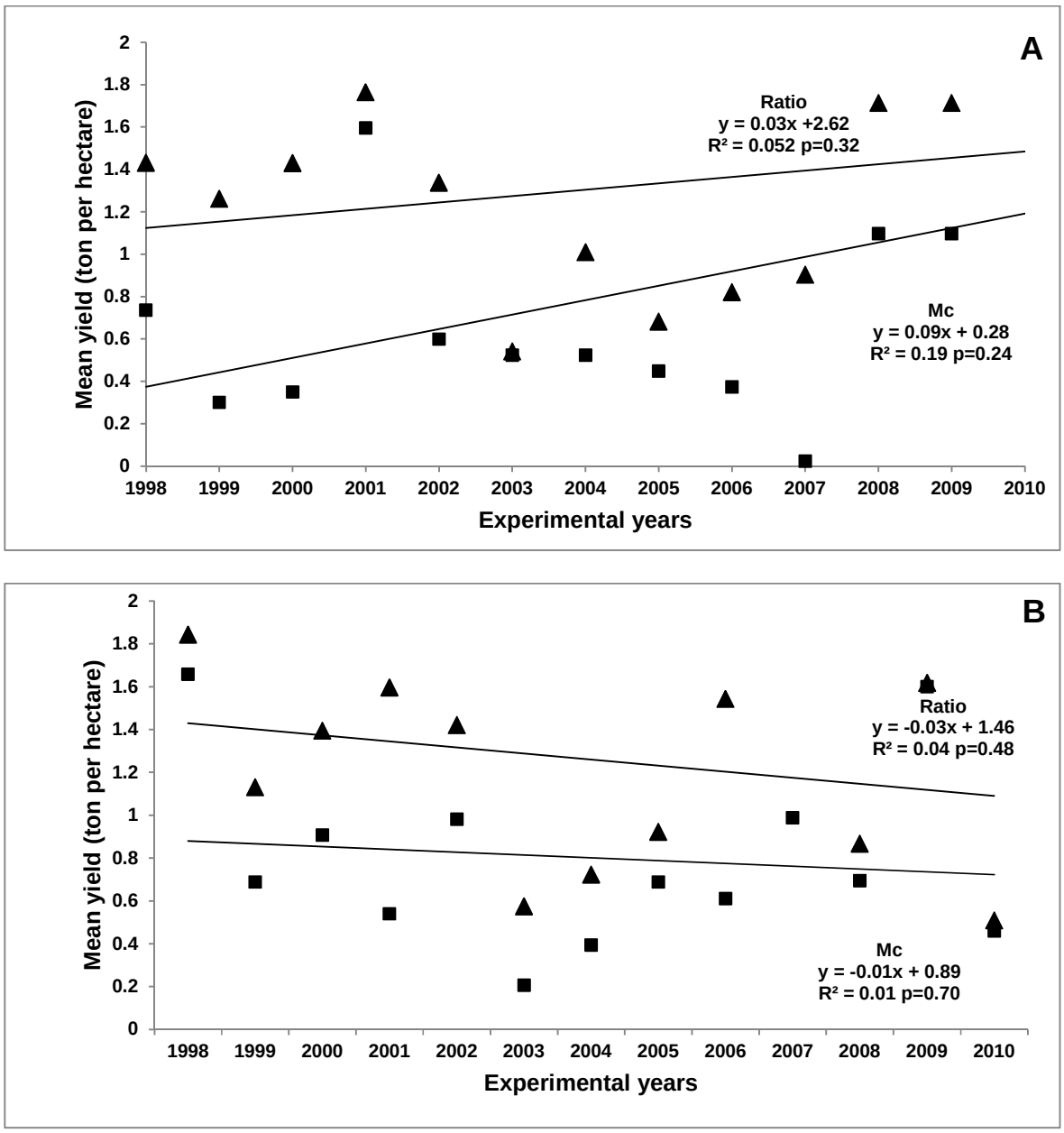


Figure 4.7 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial means of the cultivar trial in the north western Free State region for (A), the first planting date, and (B), for the second planting date. ▲ = Ratio; ■ = Mc.

In the central Free State (**Figure 4.8 A**) a significant improvement of mean yield for Mc, planting date 1 with R^2=0.39, p=0.02, was found. The slope, b=0.04, of Mc indicated a 4% genetic advance per year over 13 years. Mc for planting date 2 (**Figure 4.8 B**) showed a significant trend at 10% (with R^2=0.28, p=0.06).

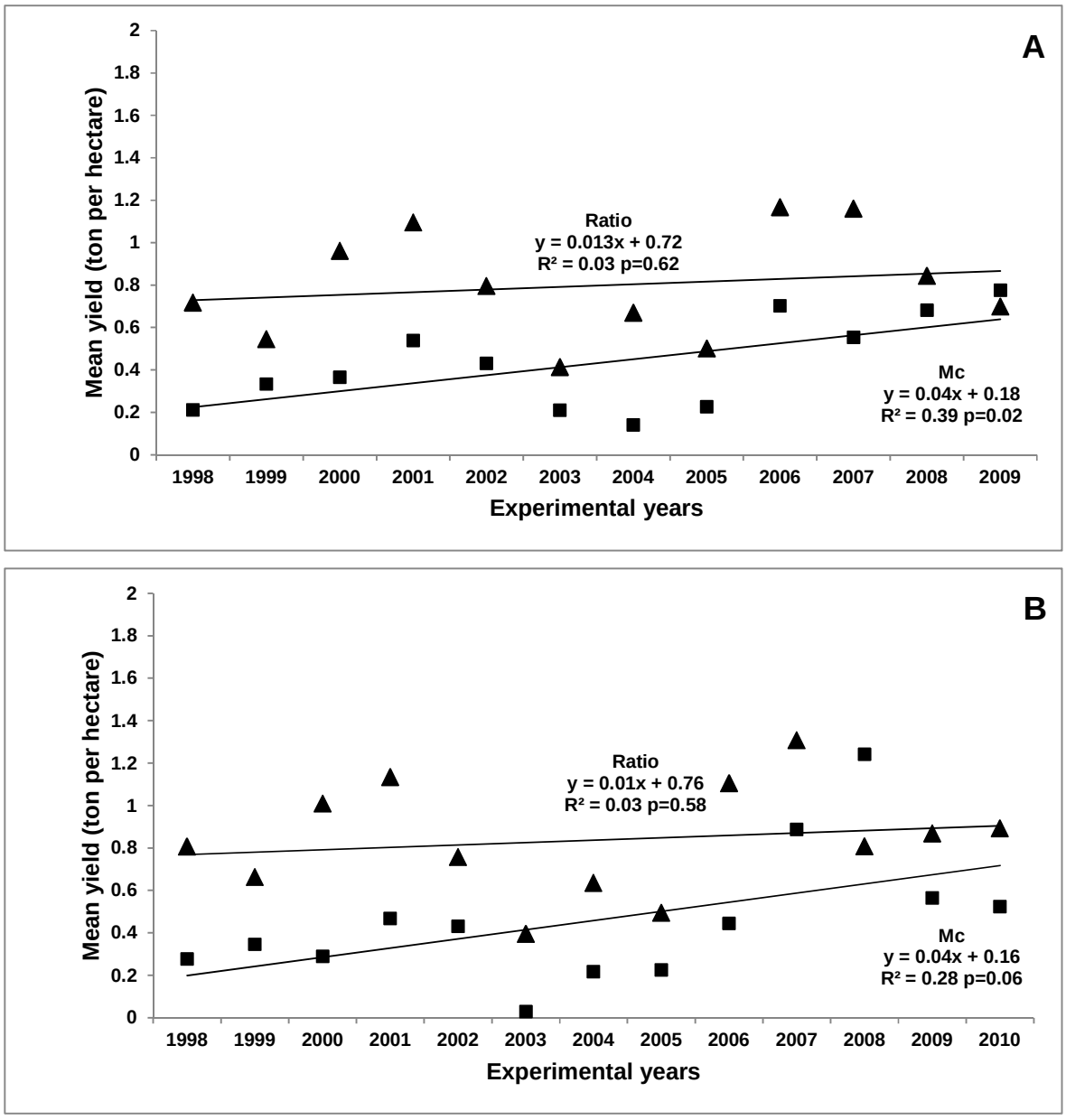


Figure 4.8 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial means of the cultivar trial of the central Free State region for (A), the first planting date, and (B), the second planting date. ▲ = Ratio; ■ = Mc

The genetic improvement measured by the slope, b=0.04, indicated a 0.3 % (= 0.04/13*100) yield progress per year. Mc should be divided by number of years (see Section 2.9 for interpretation of the slope). In the southern Free State no significant progress over time was evident for both planting dates for Ratio and Mc (**Figure 4.9 A and B**).

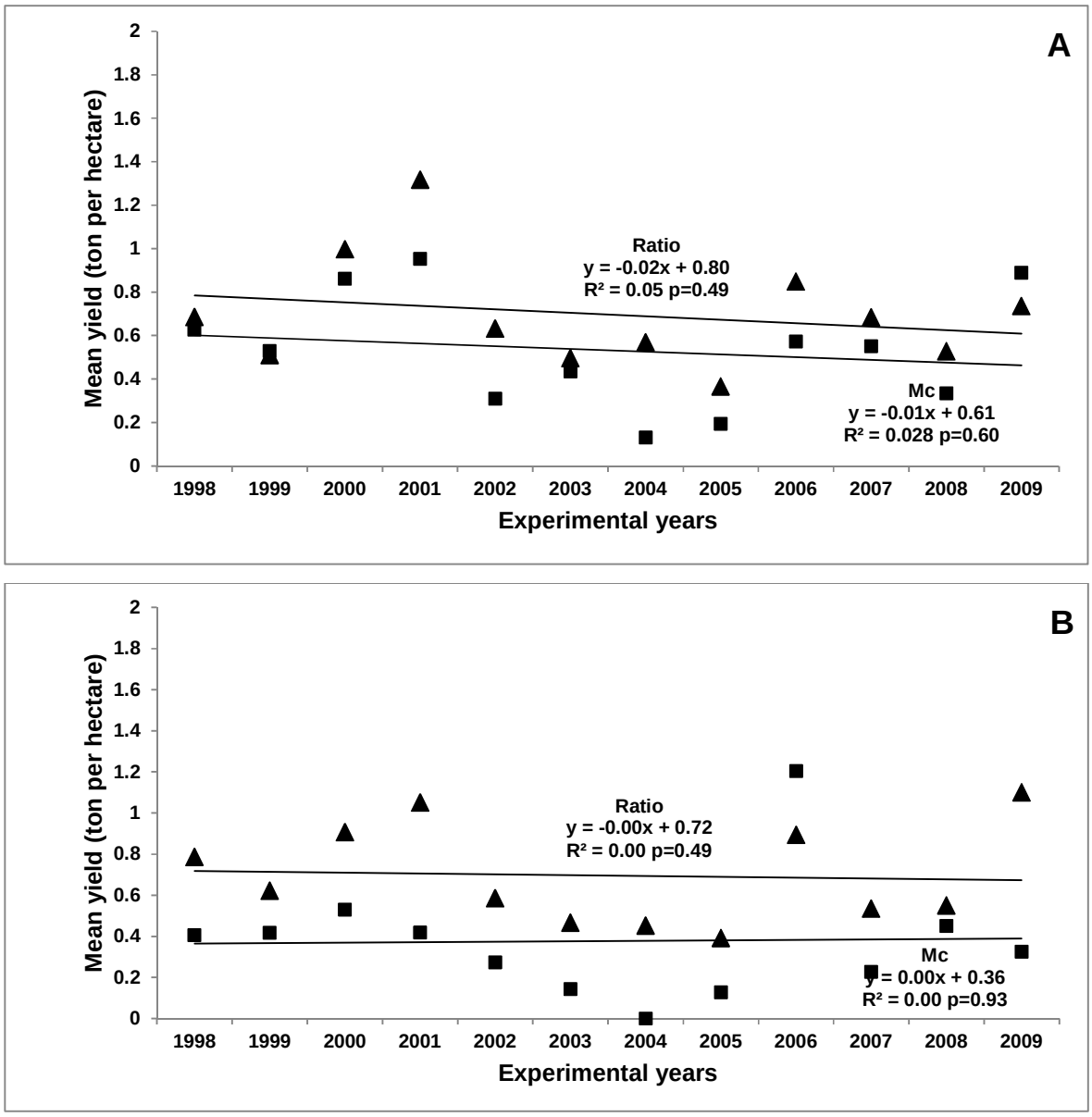


Figure 4.9 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial means of the cultivar trial of the southern Free State region for (A), the first planting date, and (B), the second planting date. ▲ = Ratio; ■ = Mc.

In the cultivar trials, genetic improvement was found for both planting dates in the eastern and central regions with Mc. The evaluation of historical series of genotypes is a common approach used in several countries to quantify the achievements of past breeding efforts (Austin *et al.,* 1989; Donmez *et al.,* 2001). The present study examined a set of approximately 250 bread wheat genotypes in the breeding trials and 50 bread wheat genotypes in the cultivar trials. The methods to combat environmental fluctuations used, namely Ratio (TRET) and Mc, showed contradictory results. For the breeding trials Ratio showed the progress or rather lack thereof successfully. In the cultivar trials Mc was a far better predictor of the progress than Ratio.

4.4.3 Sources of variation, heritability and genetic advance

The mean squares obtained from the combined analysis of variance showed all effects were significant ($p<0.0001$) for yield (Yld), hectolitre mass (HLM), protein content (Protein), logarithm transformed variables yield (LYld), HLM (LHLM) and protein content (LProt). The mean squares of yield, HLM and protein content from the weighted analyses (WYld, WHLM and WProtein respectively) from the elite trials and cultivar trials of the different regions for both planting dates (Appendix B Tables B1 to Table B14).

4.4.4 Comparison of linear fixed models and linear mixed models

In **Table 4.3** and **Table 4.4** the linear regression coefficient, b (TRET), genetic advance (ΔG) of the linear fixed model (model 1), linear mixed model (model 4) and the model using heritability (model 6) for yield are shown. The broad sense heritability (H^2) values for yield and the variables from the transformations (LYld and WYld) were low. Thus the repeatability values of the trials were low and this indicated no genetic progress was made. This is also confirmed by no progress measured by TRET (**Table 4.3**).

Table 4.3 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) calculated from the various proposed models for the yield (Yld) of the elite trials (LYld = logarithmic transformed yield, WYld = weighted analysis of yield)

				% ΔG			
Planting date	**Region**	**Variable**	**H^2**	**Model 1**	**Model 4**	**Model 6**	**TRET**
Intermediate	Eastern	Yld	0.02	1.97	8.13	0.10	0.00
Intermediate	Eastern	LYld	0.02	2.39	9.50	0.06	
Intermediate	Eastern	WYld	0.02	2.28	0.00	0.43	
Winter	Central	Yld	0.01	0.70	5.48	0.02	0.00
Winter	Central	LYld	0.00	0.30	1.83	0.01	
Winter	Central	WYld	0.00	0.46	0.00	0.07	
Winter	Eastern	Yld	0.02	2.37	0.92	0.13	-1.00
Winter	Eastern	LYld	0.02	2.61	1.15	0.07	
Winter	Eastern	WYld	0.02	2.25	0.00	0.42	
Winter	Western	Yld	0.02	1.40	7.92	0.09	0.00
Winter	Western	LYld	0.01	1.05	5.90	0.03	
Winter	Western	WYld	0.01	1.36	0.00	0.32	

In the cultivar trials the broad sense heritability values for yield were low (**Table 4.4**) and the genetic advance (GA) values differed largely amongst the models. TRET normally indicated Ratio but in these trials it referred to the modification Mc (see Section 4.3.1 for a definition).

Table 4.4 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) calculated from the various proposed models for the yield (Yld) of the cultivar trials (LYld = logarithmic transformed yield, WYld = weighted analysis of yield)

Planting date	Region	Variable	H^2	% ΔG			
				Model 1	Model 4	Model 6	TRET(Mc)
1	Central	Yld	0.02	2.08	10.23	0.08	0.30
1	Central	LYld	0.01	1.69	8.46	0.00	
1	Central	WYld	0.06	3.69	0.00	1.00	
1	Eastern	Yld	0.04	3.38	11.46	0.31	0.54
1	Eastern	LYld	0.03	2.54	12.23	0.08	
1	Eastern	WYld	0.03	3.62	0.00	0.92	
1	North West	Yld	0.05	3.08	7.08	0.31	Ns
1	North West	LYld	0.05	3.46	6.23	0.15	
1	North West	WYld	0.04	2.77	0.00	0.69	
1	Southern	Yld	0.03	2.46	12.92	0.00	Ns
1	Southern	LYld	0.01	1.31	6.54	0.15	
1	Southern	WYld	0.04	3.00	0.00	0.77	
2	Central	Yld	0.02	2.00	8.92	0.08	0.30
2	Central	LYld	0.01	1.62	9.77	0.00	
2	Central	WYld	0.00	3.69	0.00	1.00	Ns
2	Eastern	Yld	0.04	3.23	10.15	0.31	
2	Eastern	LYld	0.03	2.46	8.38	0.08	
2	Eastern	WYld	0.00	3.62	0.00	0.92	0.60
2	North West	Yld	0.04	2.85	7.38	0.31	
2	North West	LYld	0.05	3.23	8.69	0.08	
2	North West	WYld	0.00	2.77	0.00	0.69	Ns
2	Southern	Yld	0.04	2.77	7.00	0.15	
2	Southern	LYld	0.02	1.54	4.31	0.08	
2	Southern	WYld	0.00	3.00	0.00	0.77	

* Ns = no significant trend

The many negative genetic advance values portrayed by model 1 and model 6 for HLM of the cultivar trials showed that these models (**Table 4.5**) could not predict genetic advance.

The broad sense heritability (H^2) values varied in a range from 0 to 0.15. Unfortunately this was not large enough to indicate repeatability and a fair response to selection.

Table 4.5 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) calculated from the various proposed models for the HLM of the cultivar trials (LHLM = logarithmic transformed HLM, WHLM = weighted analysis of HLM)

Planting date	Region	Variable	H^2	% ΔG		
				Model 1	Model 4	Model 6
1	Central	HLM	0.00	-9.54	15.77	-1.08
1	Central	LHLM	0.02	1.62	15.23	0.00
1	Central	WHLM	0.09	6.77	0.00	2.31
1	Eastern	HLM	0.02	2.23	20.15	0.62
1	Eastern	LHLM	0.09	7.23	20.23	0.00
1	Eastern	WHLM	0.11	8.77	0.00	3.23
1	North West	HLM	0.15	8.54	14.31	2.62
1	North West	LHLM	0.15	8.62	14.00	0.00
1	North West	WHLM	0.11	6.69	0.00	2.54
1	Southern	HLM	0.06	4.77	2.85	1.15
1	Southern	LHLM	0.06	4.62	22.85	0.00
1	Southern	WHLM	0.01	2.23	0.00	0.62
2	Central	HLM	0.00	-9.23	9.31	-1.00
2	Central	LHLM	0.02	1.54	6.85	0.00
2	Central	WHLM	0.00	-11.69	0.00	-0.92
2	Eastern	HLM	0.02	2.15	0.85	0.62
2	Eastern	LHLM	0.09	6.85	18.54	0.00
2	Eastern	WHLM	0.00	-23.54	0.00	-4.85
2	North West	HLM	0.13	7.92	22.85	2.23
2	North West	LHLM	0.13	8.08	23.00	0.00
2	North West	WHLM	0.12	8.00	0.00	2.92
2	Southern	HLM	0.08	5.46	12.85	1.46
2	Southern	LHLM	0.08	5.31	12.62	0.00
2	Southern	WHLM	0.10	6.54	0.00	2.38

In **Table 4.6** the linear regression coefficient b (TRET), genetic advance (ΔG) of the fixed model (model 1), mixed model (model 4) and the model using heritability estimates (model 6) for protein content of the cultivar trials were compared. In the central region broad sense heritability value for protein content was 0.46. This indicates repeatability, which in its turn indicates response to selection indicating genetic progress. This was confirmed by the progress measured by TRET Mc for yield of 0.3% per year **(Table 4.4)**.

Table 4.6 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) from the various models for protein content (Prot) of the cultivar trials (LProt = logarithmic transformed protein content, WProt = weighted analysis of protein content)

Planting date	Region	Variable	H^2	% ΔG		
				Model 1	Model 4	Model 6
1	Central	Prot	0.46	1.69	13.02	0.38
1	Central	LProt	0.45	0.00		0.38
1	Central	WProt	0.40	3.23		0.38
1	Eastern	Prot	0.07	4.46	34.32	0.69
1	Eastern	LProt	0.07	4.31	33.14	0.08
1	Eastern	WProt	0.05	3.46		0.85
1	North West	Prot	0.06	1.46	27.81	0.54
1	North West	LProt	0.05	3.38	26.04	0.00
1	North West	WProt	0.02	3.62		0.31
1	Southern	Prot	0.03	2.77	21.30	0.38
1	Southern	LProt	0.03	2.62	20.12	0.00
1	Southern	WProt	0.03	2.23		0.46
2	Central	Prot	0.05	3.46	26.63	0.38
2	Central	LProt	0.05	3.38		0.00
2	Central	WProt	0.06	3.85		1.08
2	Eastern	Prot	0.07	4.23	32.54	0.69
2	Eastern	LProt	0.07	4.08	31.36	0.08
2	Eastern	WProt	0.06	4.00		1.15
2	North West	Prot	0.05	3.38	26.04	0.46
2	North West	LProt	0.04	3.15	24.26	0.00
2	North West	WProt	0.04	3.23		0.77
2	Southern	Prot	0.05	3.08	23.67	0.54
2	Southern	LProt	0.04	2.92	22.49	0.08
2	Southern	WProt	0.05	3.23		0.77

The genetic advance (more precisely change in genetic advance) in **Table 4.3** to **Table 4.6** for yield, HLM and protein content portray a certain picture. The genetic advance values of model 6 and TRET were similar (for yield). The values of model 4 were very different to the other models, forming “groups” and the values of model 1 were somewhere in the middle. Statistically it tells us nothing. The comparison of the genetic advance values of the different models is difficult. Performing a Pearson’s product moment correlation test as described in Section 3.4.5 is a way to validate the genetic advance values of the different models.

The agreement between yield (Yld) for model 1 and model 6 was strong and significant (r=0.80, p=0.005). The correlation coefficient between model 1 and TRET was r=0.65 with p=0.04. The agreement between model 4 and TRET was r=0.81 with p=0.00. Cronbach's alpha=0.90 when omitting model 4 in the comparison of the models, which is a good indication that model 4 was not consistent with the other models. No significant agreement could be found among the models of the logarithmic transformed yield (LYld). A significant agreement (r=0.98, p<0.00) between model 1 and model 6 was found for weighted yield (WYld). Cronbach's alpha=0.85 when model 4 was omitted. These results are an indication that model 1, model 6 and TRET are testing the same construct, namely genetic advance.

Similar results were found for the variables HLM and protein content. TRET was not performed for HLM and protein content, therefore only the variance component methods (models 1, 4 and 6) were compared for the quality traits. For HLM the agreement between model 1 and model 6 was strong (r=0.97, p<0.00). Cronbach's alpha =0.98 when model 4 was omitted. A significant agreement was found between model 1 and model 6 (r=0.94, p<0.00) for the weighted HLM (WHLM). Cronbach's alpha =0.99 confirmed the result of the Pearson's product moment correlation.

The agreement of model 1 and model 6 for protein content was r=0.79 with p=0.02. No significant agreement was found among the models of the genetic advance values for the logarithmic transformed protein content or weighted protein content. The Cronbach's alpha=0.88 for protein content was consistent with the Pearson's product moment correlation.

4.4.5 AMMI versus GGE

The goal of cultivar trials is to determine the adaptability and stability of a superior genotype selected from elite trials. Various methods are available in the literature to determine stability, of which AMMI and GGE biplots are the latest. The objective of this study was to compare these techniques and not to give an interpretation of the results.

A high incidence of unbalanced data resulted in only two years analysed for five localities (10 environments), and 16 and 11 genotypes of the trials in the eastern Free State for the two planting dates respectively. The first step in a genotype by environment analysis is to look at a combined ANOVA if GEI is present and determining how large the variation is. The ANOVA for yield, HLM and protein content indicated that the effects of genotype, environment and their interaction on yield, HLM and protein content were significant (results not shown). For planting date 1, the proportion of the total treatment variation for genotype was 8.28%, 23.83% and 8.42% for yield, HLM and protein respectively. The proportion of variation for environment respectively for yield, HLM and protein was 84.04%, 56.71% and 84.74%. The proportion of variation for the interaction was 7.67%, 19.46% and 6.84% respectively.

Table 4.7 Analysis of variance from AMMI analysis of yield, HLM and protein of eastern Free State cultivar trials from planting date 1 during 2009 and 2010

		Yield		HLM		Protein	
Source	**DF**	**MS**	**%SS**	**MS**	**%SS**	**MS**	**%SS**
Block (Env.)	30	0.76		1.53		2.40	
Treatments	**159**	**5.39**		**16.16**		**14.02**	
Environments	**9**	**79.94**	**84.04**	**161.89**	**56.71**	**209.89**	**84.74**
Genotypes	**15**	**4.73**	**8.28**	**40.82**	**23.83**	**12.51**	**8.42**
Interactions	**135**	**0.49**	**7.67**	**3.70**	**19.46**	**1.13**	**6.84**
IPCA1	23	0.84		10.37		2.37	
IPCA2	21	0.68		3.75		1.60	
Residuals	91	0.36		2.01		0.71	
Error	450	0.14		0.53		0.36	
Total	639	1.47		4.47		3.86	

The percentage variation, accounted for by the genotypes in the three variables, was larger than the percentage variation accounted for by the interaction.

For planting date 2 the proportion of the total treatment variation for genotype was 9.24%, 16.15% and 6.11% for yield, HLM and protein respectively. The proportion of environment variation respectively for yield, HLM and protein was 74.29% 63.29% and 88.49%. The proportion of variation for the interaction was 16.48%, 20.56% and 5.41% respectively (**Table 4.8**).

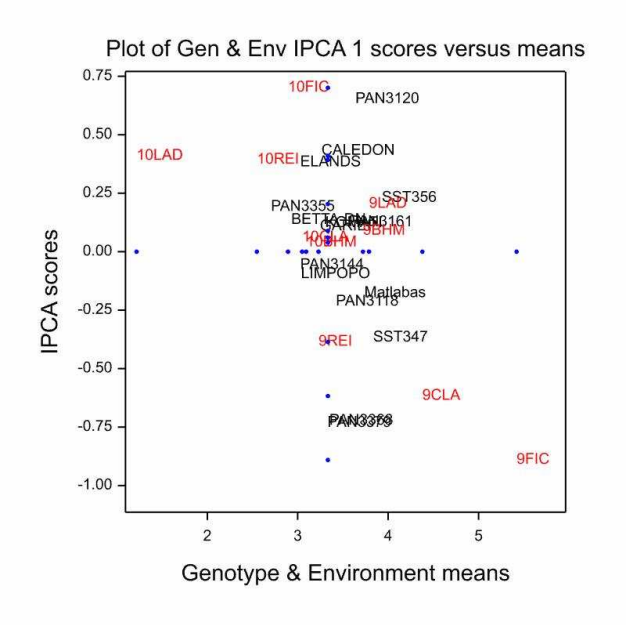
This is in accordance with Gauch and Zobel (1996) who postulated that, in multi-environmental trials, 80% of the total sum squares of treatment was due to environment effect, 10% due to the effects of genotypes, and 10% due to the interaction effect. A large variation in yield, HLM and protein explained by environments indicates the environments were diverse.

Table 4.8 Analysis of variance from AMMI analysis of yield, hectolitre mass (HLM) and protein of eastern Free State cultivar trials from planting date 2 during 2009 and 2010

		Yield		HLM		Protein	
Source	**DF**	**MS**	**%SS**	**MS**	**%SS**	**MS**	**%SS**
Block (Env.)	30	0.41		2.38		1.87	
Treatments	**109**	**2.80**		**24.97**		**12.20**	
Environments	**9**	**25.20**	**74.29**	**191.44**	**63.29**	**130.74**	**88.49**
Genotypes	**10**	**2.82**	**9.24**	**43.96**	**16.15**	**8.12**	**6.11**
Interactions	**90**	**0.56**	**16.48**	**6.22**	**20.56**	**0.80**	**5.41**
IPCA1	18	1.33		20.19		1.77	
IPCA2	16	0.78		4.83		1.21	
Residuals	56	0.25		2.12		0.37	
Error	300	0.11		0.52		0.53	
Total	439	0.80		6.72		3.52	

The AMMI biplot and the GGE biplot in **Figure 4.10** for yield of the first planting date showed similar results. The genotypes were all high yielding, but interacted with different environments. Genotype PAN3120 had the highest yield in mega environment 10FIC.

A



B

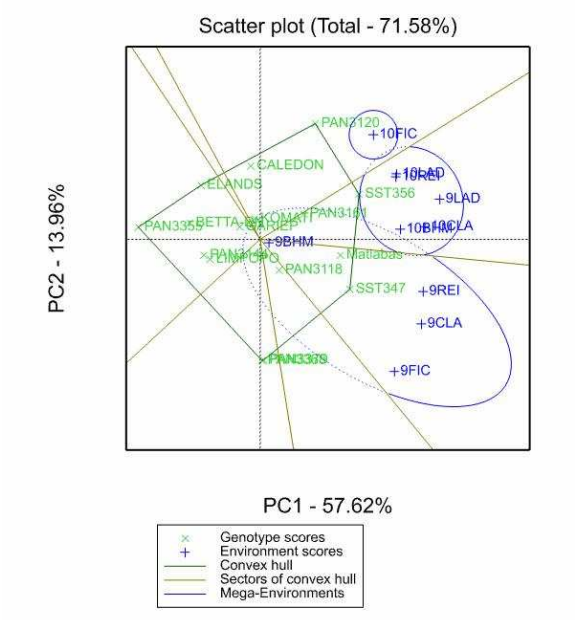
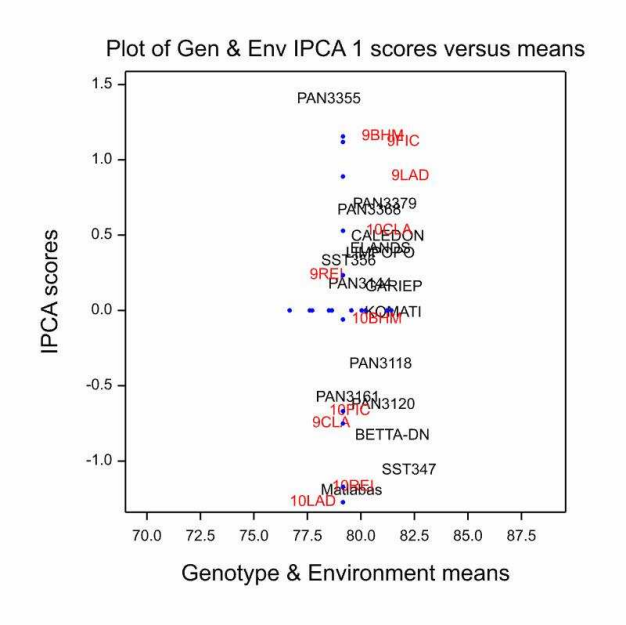


Figure 4.10 Genotype by environment biplot of planting date 1 in the eastern region of the Free State for yield (A) Principal Component I versus mean yield (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

The GGE biplot clearly shows three mega environments and the “which one wins where” genotypes are the ones on the vertices of the polygon. Genotype SST347 was the winning genotype in the environments 9CLA, 9FIC and 9REI. In the third mega environment — formed by 9BHM,10BHM, 0CLA, 9LAD, 10LAD and 10REI — genotype SST356 was the best adapted genotype.

The AMMI and GGE biplots presented for HLM of the first planting date in **Figure 4.11** show similarities. The GGE biplot clearly shows two mega environments. Komati and SST347 are the winning genotypes in 9CLA, 10CLA, 10BHM, 10FIC, 10LAD, 9REI and 10REI. The AMMI and GGE biplots clearly show PAN3355 as the winning genotype in the second mega environment, formed by the three environments, 9BHM, 9FIC and 9LAD.

A



B

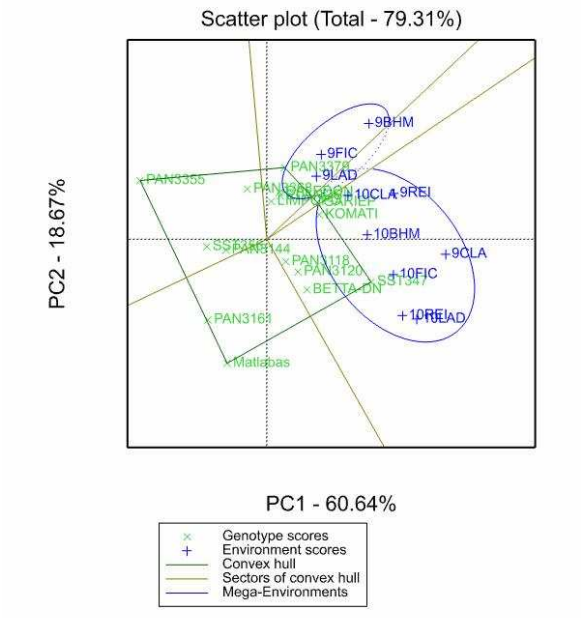
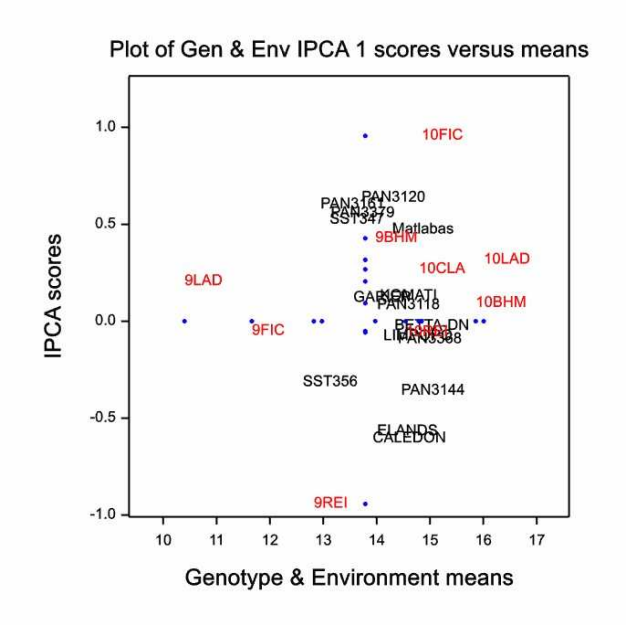


Figure 4.11 Genotype by environment biplot of planting date 1 in the eastern region of the Free State for hectolitre mass (HLM) (A) Principal Component I versus mean HLM (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot)

The AMMI and GGE biplots presented in **Figure 4.12** for protein content of the first planting date show similarities. The GGE biplot clearly shows three mega environments. PAN3144 is the winning genotype in 9CLA, 10CLA, 10BHM, 10FIC, 10LAD, 9REI and 10REI and in the second mega environment, formed by the three environments 9BHM, 9FIC and 9LAD. The AMMI biplot unfortunately did not show this result so clearly. In the GGE biplot environment, 10CLA formed the third mega environment with genotype PAN3355 as the winner. This differed from the AMMI analysis (results not shown) and AMMI biplot.

A



B

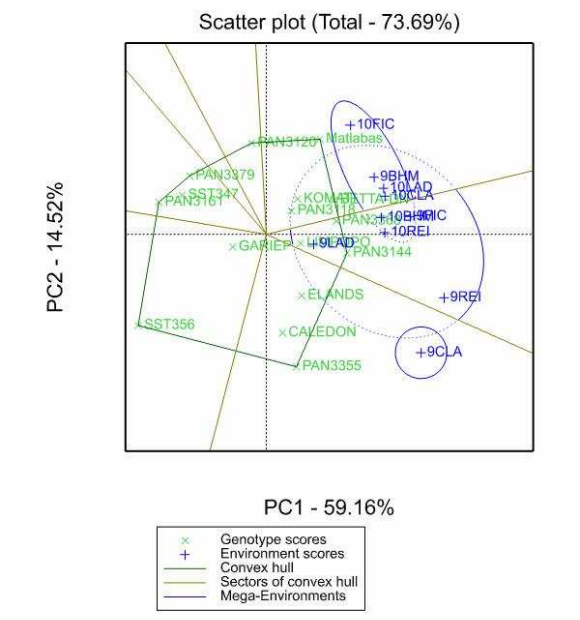
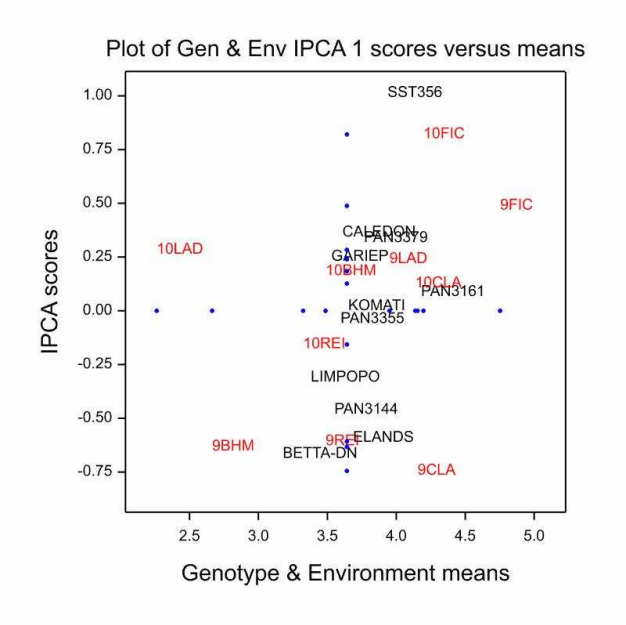


Figure 4.12 Genotype by environment biplot of planting date 1 in the eastern region of the Free State for protein content (A) Principal Component I versus mean protein content (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

Two mega environments were identified in the GGE biplot in **Figure 4.13** for yield of the second planting date. Mega environment 1 includes 9BHM, 9CLA, 10CLA, 9FIC and 9REI with best performing genotype, PAN3161. The second mega environment consists of environments 10BHM, 10FIC, 9LAD, 10LAD and 10REI and the best performing genotype was SST356. Although the AMMI biplot does not show the results so clearly, the same conclusions can be drawn.

A



B

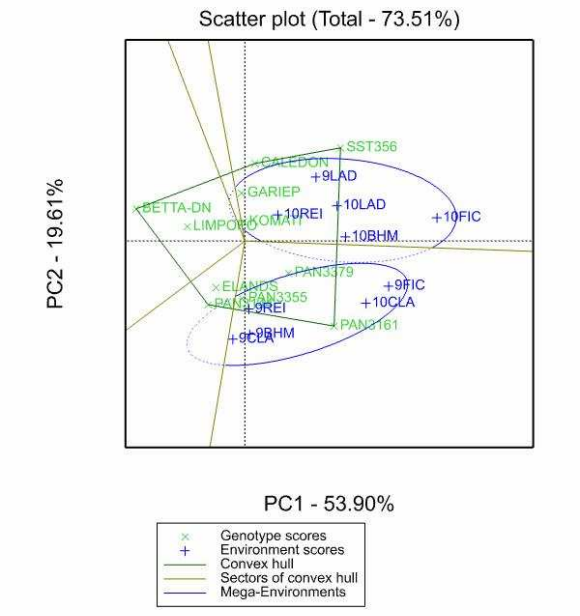
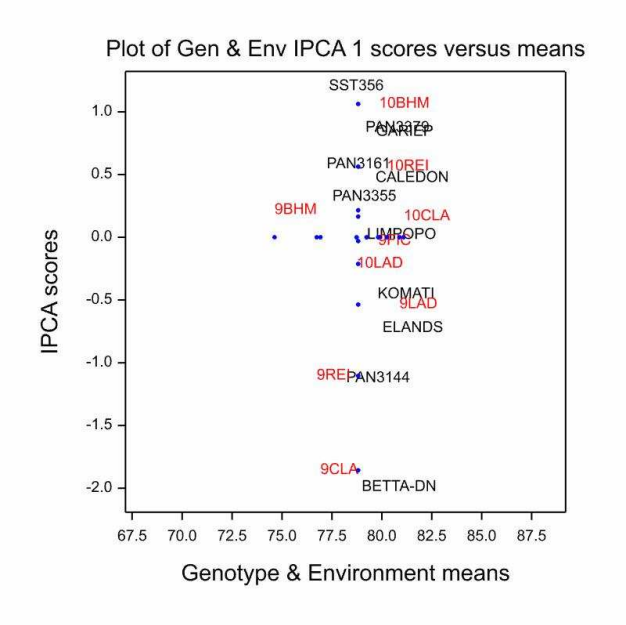


Figure 4.13 Genotype by environment biplot of planting date 2 in the eastern region of the Free State for yield (A) Principal Component I versus mean yield (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

The GGE biplot in **Figure 4.14 B** shows three mega environments for HLM of the second planting date. The first mega environment includes environments 10FIC, 10BHM and 10REI with winning genotype, Gariep. The winning genotype in the second mega environment (9BHM, 9FIC, 10LAD and 9CLA) is Elands. Betta-DN is the best performer in the third mega environment, which involves environments 9CLA, 9LAD and 9REI. The AMMI table of four best genotypes (not shown) were consulted to assist with the interpretations.

A



B

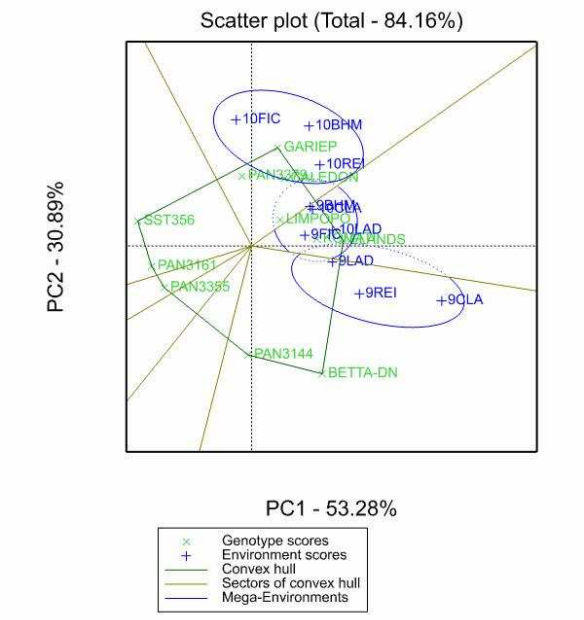
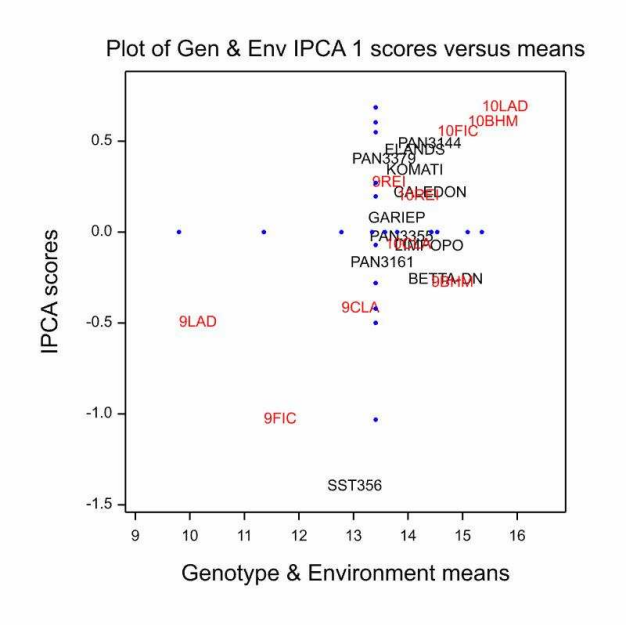


Figure 4.14 Genotype by environment biplot of planting date 2 in the eastern region of the Free State for hectolitre mass (HLM) (A) Principal Component I versus mean HLM (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

In **Figure 4.15** the GGE biplot analysis for protein content of the second planting date showed three mega environments. The first mega environment consisted of one environment, namely 10FIC and the best performing genotype for this environment was PAN3144. Betta-DN is the winning genotype of the second mega environment (9BHM, 9CLA, 10CLA, 9LAD and 9REI). The third mega environment involves environments 10BHM, 10REI, 10FIC and 10LAD where Elands was the best performing genotype. These conclusions were derived from the AMMI analysis (not shown), AMMI biplot and GGE biplot.

A



B

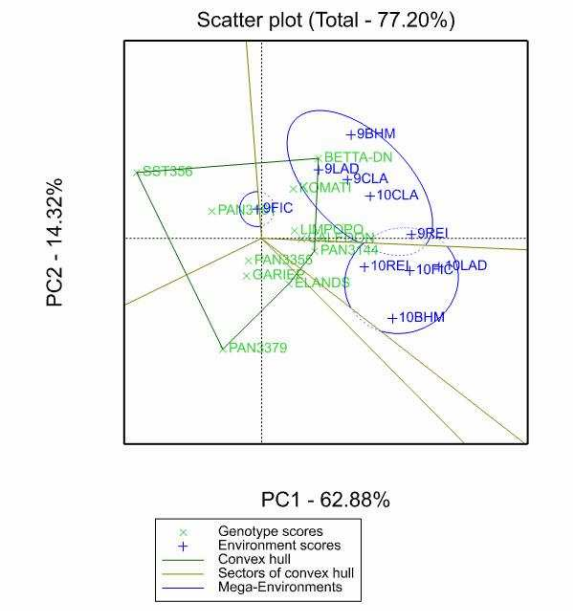


Figure 4.15 Genotype by environment biplot of planting date 2 in the eastern region of the Free State for protein content (A) Principal Component I versus mean protein (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

When the above results of both models are compared, the same genotypes performed the best. Also, in terms of stability and mega environment analysis, there are similarities between the two models. However, there are conflicting opinions about the reliability of both models and which one more credibly represents genotypes, environments (year by locality) and their interactions. According to Yan *et al.* (2007) the AMMI biplot is not accurately presented, especially when performances of individual genotypes in certain environments are observed and it is necessary to use AMMI2 charts for assessment of a so called “which won where“ model. In addition, the AMMI biplot always explains less G+GEI variation than the GGE biplot and its shape is completely subjective because the axes are in different units (original unit for the abscissa and square root of the original unit for the ordinate). The AMMI1 biplot also presents the environment main effects of the test environments or E, which is irrelevant to cultivar and test-environment evaluation (Yan and Kang 2003).

In contrast, Gauch *et al.* (2008) argued that a shortcoming of the GGE biplot method for evaluation of the genotypes is its inability to separate G from GEI, which is not the case in the AMMI. The abscissa of an AMMI1 biplot captures 100% of G. The ANOVA part of AMMI can separate G from GEI, which the PCA part of both the AMMI and GGE is not able to do.

This fact is very important because the large GEI relative to G could drive a large portion of G into the third and higher components that a GGE2 biplot misses. Additionally, more advantages of the AMMI biplot over the GGE biplot, according to Gauch *et al.* (2008) are that the AMMI biplot is simpler to construct and interpret because its axes are used directly and there is no need for the axes to be rotated.

Stojaković *et al.* (2012) used both models to analyse a set of 15 commercial maize hybrids in 30 environments across Serbia and also concluded that, besides methodological, there is no large difference in the results obtained by both models. Rad *et al.* (2013) found, in their study on winter bread wheat of 36 genotypes in six environments, that the results of the AMMI and GGE biplot were similar.

In this study both models proved to be very useful in assessing the performance of genotypes for yield, HLM and protein content for both planting dates. The models showed no differences in selection of best genotypes. The complementary nature of the models as shown in the biplots (**Fig 4.10** to **Fig 4.15**) is an advantage. The choice of the model (AMMI or GGE) depends on the aim of the study. The AMMI provides an analysis with an ANOVA table, mean and ranking values of the genotypes within the environments, a table of the best four genotypes in the environments and various biplots (not all the results are shown). The strength of the GGE biplot is its graphical presentation of the genotype by environment interaction. In this study the genotype by environment interaction biplots of the models were compared. It was found in most cases that the conclusions were similar.

4.4.6 Other techniques

Grain protein is an important factor affecting grain quality and thus a key determinant of both end use and market value in bread wheat. However, it is known that grain protein is negatively correlated with grain yield in wheat. Hence selection to increase one trait will tend to decrease the second trait (Rharrabti *et al.*, 2001). Despite this negative correlation, studies have reported simultaneous improvement of grain yield and grain protein content (Niu *et al.,* 2010).

From the Pearson's product moment correlation matrix (**Table 4.9**) it was observed that yield correlated significantly positive with HLM and was significantly negatively correlated with protein content. The relationship among these correlations and the genotypes is revealed through a PCA.

Table 4.9 Pairwise correlations between variables relating yield and quality attributes of the cultivars across the ten environments for the two planting dates

	Planting Date 1		**Planting Date 2**	
Variables	**Yield**	**HLM**	**Yield**	**HLM**
HLM	0.474 *		0.315 *	
Protein content	-0.678 *	-0.397 *	-0.567 *	-0.174

***P≤0.05 HLM=hectolitre mass**

It would have been ideal to investigate the simultaneous relationship of the three variables and genotypes with environment, but the biplots of the PCA were cluttered. Thus the results are given on the relationship among the variables and the genotypes. The PCA in **Figure 4.16** for planting date 1 showed protein content is located in quadrant 1 (nearly opposite of yield and HLM), thus confirming the negative correlations with yield and HLM (**Table 4.9**). Yield and HLM share the same quadrant (quadrant 2) thus indicating their positive relationship. Protein content is closely associated with genotypes Betta-DN, Elands and Limpopo, while HLM is associated with Komati, Gariep and SST347. A strong association among yield and genotypes PAN3118, PAN3368 and PAN3379 was observed.

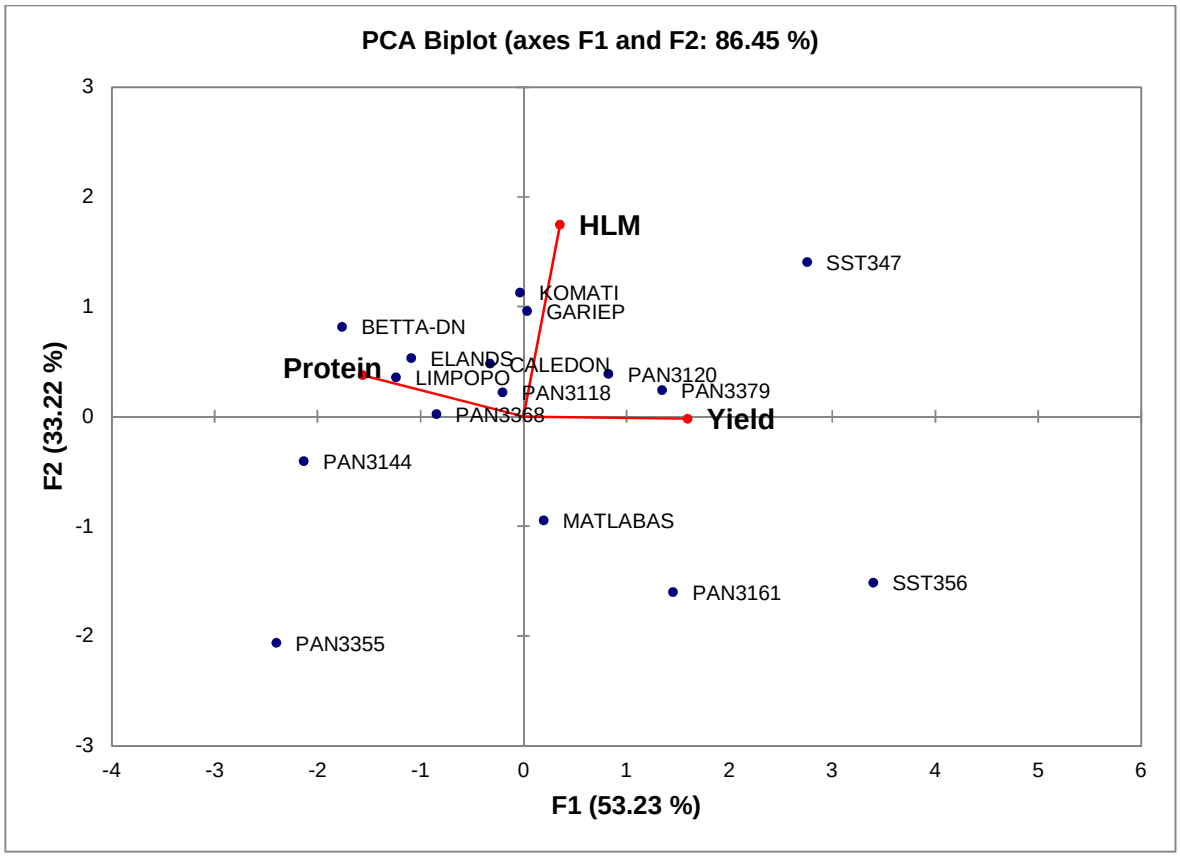


Figure 4.16 PCA biplot of relationship between all variables and the 16 genotypes of planting date 1.

The PCA in **Figure 4.17** for planting date 2 confirmed the negative correlation between yield and protein content (yield and protein content were clearly in opposite sides of the biplot) and the positive correlations between protein content and HLM and among HLM and yield

respectively. HLM showed strong association with genotypes Elands, Gariep, Komati and Caledon. Protein content was strongly associated with Limpopo. Yield was strongly associated with PAN3379.

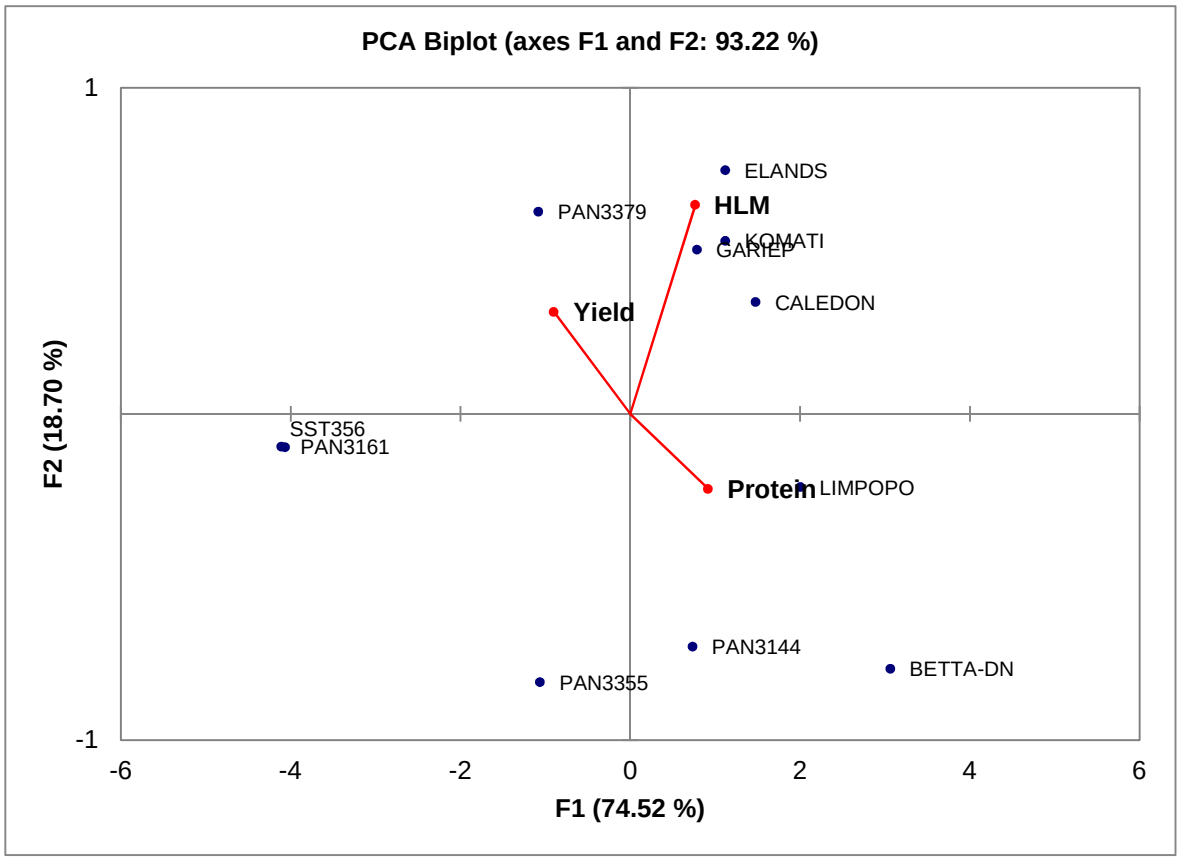


Figure 4.17 PCA biplot of relationship between all variables and the 16 genotypes of planting date 2.

These PCA results are similar to those encountered in AMMI and GGE biplots (**Fig 4.10** to **Fig 4.15**). The best performers (genotypes) for each variable in the AMMI and GGE biplots were mostly associated with the variables. These results were expected since the AMMI and GGE biplots rely heavily on principal component analysis.

In most studies cluster analysis is performed to identify the relationship among genotypes for a single variable, i.e. yield. Cluster analysis could be used in a multivariate scenario as well. It depends on the objectives of the study. In this study cluster analysis was used to

investigate the relationship among the environments for each variable and planting date, as well as to identify mega environments according to the definition of a mega environment. A mega environment is defined as a group of locations that consistently share the same best cultivar(s). Another essential requirement for mega environment differentiation is repeatability of the “which-won-where” pattern. Therefore, multi-site trials conducted over years are essential for addressing the mega environment issue (Yan *et al.*, 2007).

Three different groups could be identified from the dendrogrammes in **Figure 4.18** for the three variables.

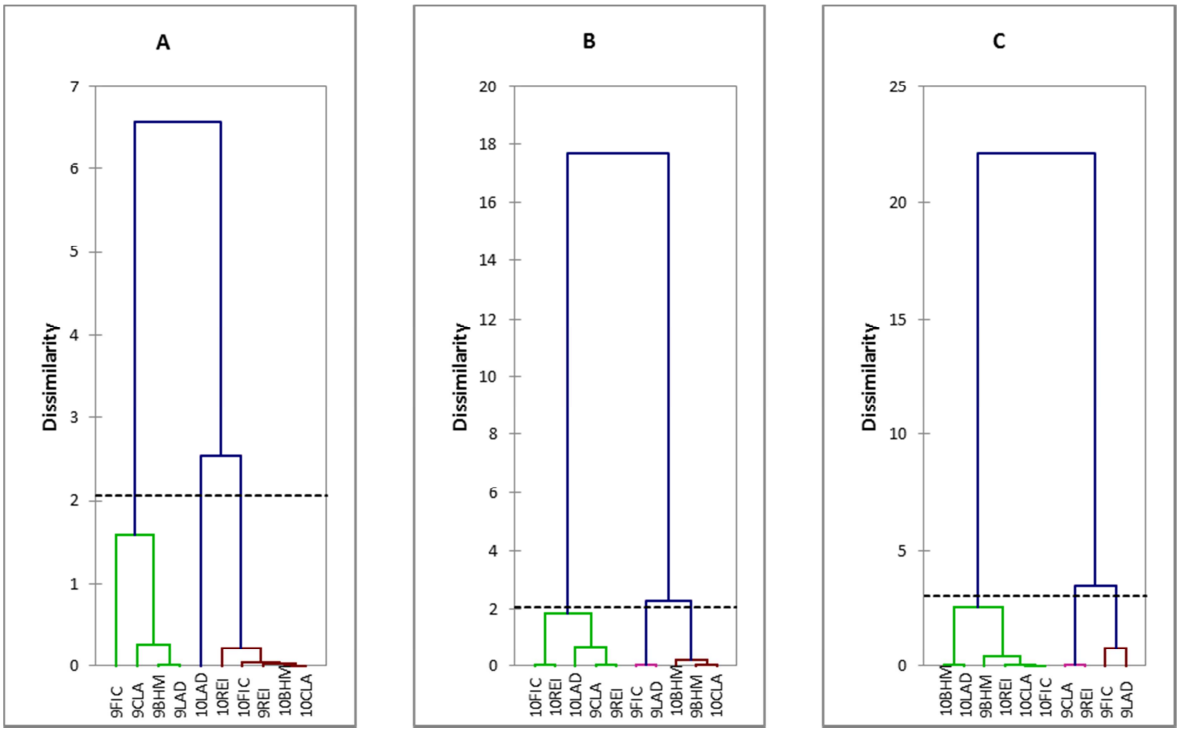


Figure 4.18 Dendrograms presenting hierarchical clustering of the 10 environments for planting date 1 for (A) yield, (B) hectolitre mass (HLM) and (C) protein content.

The groups could be identified as low, medium and high values for the three variables. For yield the high value group is green-coloured and on the left side. The intermediate (blue) is the value group (10LAD). The group on the right side (brown coloured) is the intermediate value group. For HLM the environments, 9BHM and 10BHM clustered in one group (intermediate value group). Unfortunately these environments per definition do not form a mega environment because they share the group with another environment.

In **Figure 4.19** a mega environment could be identified for protein content. Environments 9CLA and 10CLA with 9REI and 10REI formed a mega environment. These environments were repeatable over years. The dendrograms of the planting dates showed similarities. In both planting dates (**Figure 4.18** and **Figure 4.19**) environments, 9FIC and 9LAD produced low protein content.

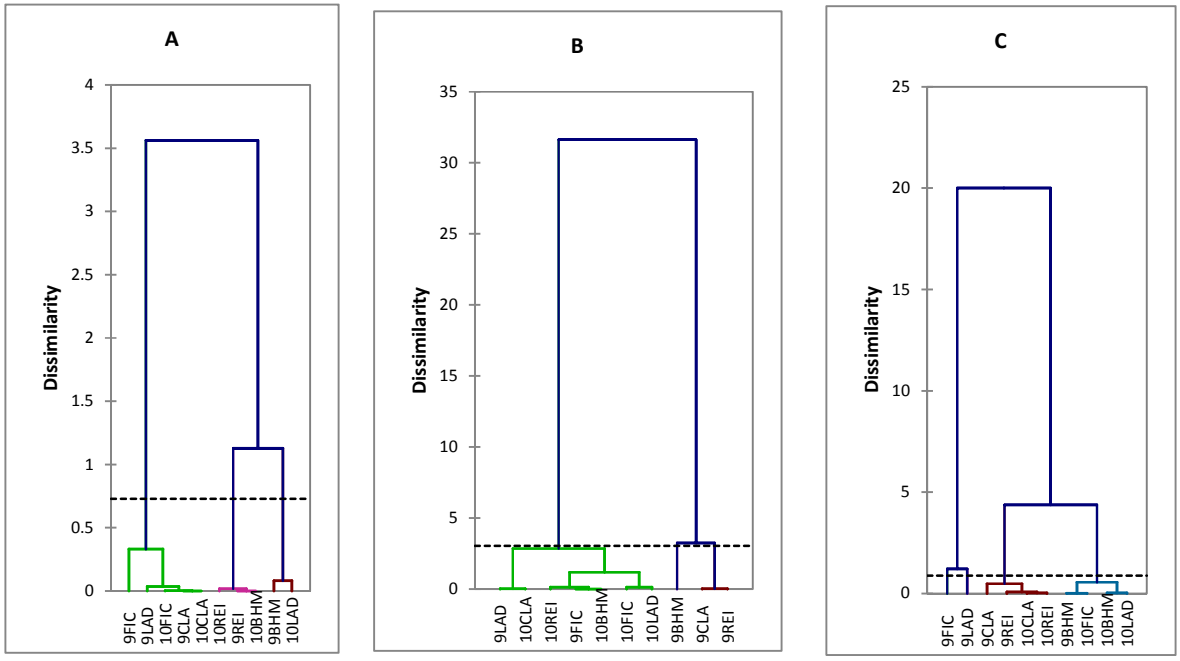


Figure 4.19 Dendrograms presenting hierarchical clustering of the 10 environments for planting date 2 for (A) yield, (B) hectolitre mass (HLM) and (C) protein content.

Linear discriminant analysis (DA) was used to determine whether the variables in this study could provide additional information on genetic advance. The first two discriminant variates (called factors) in **Figure 4.20** accounted for 95.69% variation among the years for the first planting date. In the plot the genotypes within a year are enclosed by 95% confidence ellipses. Overlapping of different environments' ellipses indicates no differences among environments. This plot indicates no separation among the years. Thus these variables are unable to discriminate among years.

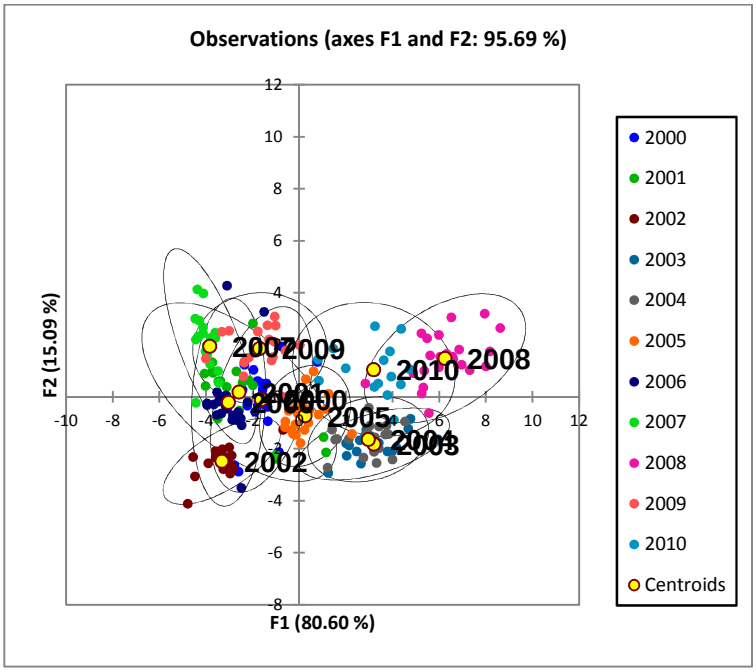


Figure 4.20 Linear discriminant biplot of the years of eastern region of the cultivar trial for planting date 1.

Van Lill and Purchase (1995) used discriminant analysis (also called canonical variate analysis) to investigate genetic advance in yield and protein content in winter wheat cultivar data for the period 1930-1994. They found a yield improvement of 1.3% per year and a stable performance in protein content. In this study no genetic improvement could be found with this technique. It was shown that these variables had no discriminating powers for the first planting date. No discriminant analysis could be performed for the second planting date of the cultivar trials of the eastern region.

4.5 Conclusions and recommendations

The proposed method for determining genetic advance over time was Trethowan (TRET), models 1 and 6. No significant trend was observed in the elite trials with TRET. Yield showed 0.5% and 0.6% improvement for the two planting dates in the cultivar trials of the eastern region. The effects of GEI on yield and quality traits were studied by comparing the additive main effects and multiplicative interaction (AMMI) and genotype plus genotype-by-environment interaction (GGE) biplot analysis. Application of AMMI and GGE biplots

facilitated visual comparison and identification of superior genotypes for each target set of environments. The complementary nature of the models, as shown in the biplots, is an advantage. The PCA showed the same genotypes as best performing genotypes with all three variables involved as the AMMI and GGE biplot analyses. Cluster analysis was unable to identify mega-environments except for protein content in the second planting date. The DA with the available variables could not indicate a direction in genetic advance.

References

Agricultural Research Council – Small Grain Institute (2010) ARC-SGI report, Bethlehem

Allard RW (1960) Principles of plant breeding. Wiley, New York

Austin RB, Ford MA, Morgan CL (1989) Genetic improvement in yield of winter wheat: A further evaluation. Journal of Agricultural Science 112:295-302

Comstock RE, Moll RH (1963) Genotype-environment interactions. In Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. National Academy of Science, Washington, pp 164-196

Cox TS, Shroyer JP Ben-Hui L, Sears RG, Martin TJ (1988) Genetic improvement in agronomic traits of hard red winter wheat cultivars from 1919 to 1987. Crop Science 28:756-760

Crossa J, Gauch HG, Zobel RW (1990) Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. Crop Science 30:493-500

Department of Agriculture (2012) Trends in the agriculture sector in 2012. In Agricultural statistics. Pretoria, pp1-11

Donmez E, Sears RG, Shroyer JP, Paulsen GM (2001) Genetic gain in yield attributes of winter wheat in the Great Plains. Crop Science 41:1412-1419

Fufa H, Baenziger PS, Beecher BS, Graybosch RA, Eskridge KM, Nelson LA (2005) Genetic improvement trends in agronomix performances and end-use quality characteristics among hard winter wheats in Nebraska. Euphytica 144:187-198

Gauch HG, Piepho HP, Annicchiarico P (2008) Statistical analysis of yield trials by AMMI and GGE: further considerations. Crop Science 48:866-889

Gauch HG, Zobel RW (1996) AMMI analysis of yield trials. In Kang MS, Gauch HG (eds) Genotype by environment interaction. Boca Raton CRC, New York, pp 85-122

Graybosch RA, Peterson CJ (2010) Genetic improvement in winter wheat yields in the Great Plains of North America, 1959-2008. Crop Science 50:1882-1890

Levene H (1960) Robust test in the equality of variance. In Olkin I (ed) Contributions to probability and statistics: Essays in honor of Harold Hotelling. Stanford UP, Stanford, pp 278-292

Niu N, Arief VN, DeLacy IH, Lush D, Sheppard J, Zhang G, Dieters MJ (2010) Genetic gain in yield and protein over two cycles of a wheat recurrent selection programme. Breeding Science 603:181-186

Payne R, Murray DA, Harding SA, Baird DB, Soutar D (2012) GenStat for Windows: Introduction, 15th edition. VSN International, Hemel Hempstead

Purchase JL, Hatting H, Van Deventer CS (2000) Genotype x environment interaction of winter wheat (*Triticum aestivum* L.) in South Africa II: Stability analysis of performance. South African Journal of Plant and Soil 17:101-107

Rad NM, Kadir MA, Rafii MY, Jaafar HZ, Naghavi MR, Ahmadi F (2013) Genotype x environment interaction by AMMI and GGE biplot analysis in three consecutive generations of wheat (*Triticum aestivum*) under normal and drought stress conditions. Australian Journal of Crop Science 7:956-961

Rharrabti Y, Villegas D, Garcia del Moral LF, Elhani S, Royo C (2001) Environmental and genetic determination of protein content and grain yield in durum wheat under Mediterranean conditions. Plant Breeding 120:381-388

SAS Institute (2012). The Statistical Procedure manual. North Carolina 27513: SAS Campus Drive, Cary

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52:591-611

Stojaković M, Ivanović M, Jocković Đ, Bekavac G, Purar B, Nastasić A, Stanisavljević D, Mitrović B, Treskić S, Laišić R (2010) NS maize hybrids in production regions of Serbia. Field and Vegetable Crops Research 47:93-102

Trethowan RM, Van Ginkel M, Rajaram S (2002) Progress in breeding wheat for yield and adaptation in global drought affected environments. Crop Science 42:1441-1446

Van Lill D, Purchase JL (1995) Directions in breeding for winter wheat yield in South Africa from 1930 to 1990. Euphytica 82:79-85

Van Niekerk HA (2001) The South African wheat pool. In Bonjean A, Angus W (eds) The world wheat book: A history of wheat breeding. Tec & Doc/Intercept, Paris, pp 923-936

Yan W, Cornelius PL, Hunt LA (2001) Two types of GGE biplots for analysing multi-environment trial data. Crop Science 41:643-355

Yan W, Kang MS (2003) GGE biplot analysis: A graphical tool for breeders, geneticists and agronomists. CRC Press, Boca Raton

Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007) GGE biplot analysis of genotype-by-environment data. Crop Science 47:643-655

Yasin BA, Ram M, Singh S, Wani B (2011) Genetic improvement in yield, yield attributes and leaf rust resistance in semi-dwarf wheat varieties developed in India from the last 40 years. International Journal of Agricultural Research 6:747-753

XLStat (2012) XLStat. Addinsoft ,Paris

Chapter 5
THE IRRIGATION TRIALS

Abstract

Spring wheat was evaluated for genetic improvement under irrigated conditions over a 16 year period (1995-2010) in 16 localities in the irrigation production areas of South Africa. The following statistical analyses and procedures were conducted to determine genetic improvement: (i) linear regression; (ii) a variance component method according to Comstock and Moll (1963); and (iii) a variance component method according to Allard (1960). A genetic advance of nearly 0.7% per year was found in the warm region of elite trials and 9% in the eastern region for the first planting date of the cultivar trials. The proposed method to determine genetic advance is TRET. The effects of GEI on yield and quality traits were studied by comparing the additive main effects and multiplicative interaction (AMMI) and genotype plus genotype-by-environment interaction (GGE) biplot analysis. Application of AMMI and GGE biplots facilitated visual comparison and superior genotypes were identified for each target set of environments. The results of the models, as shown in the biplots, are similar. Principal component analysis (PCA) and discriminant analysis (DA) provided additional information.

5.1 Introduction

South Africa (SA) is a relatively dry country, with very erratic rainfall. This leads to substantial variations in grain production, and irrigation plays an important role in stabilising grain production patterns. With regard to wheat, 12% is planted under irrigation and provides 30% of the wheat produced in the country (Department of Agriculture, 2012). This means that wheat production under irrigation plays an important role in warding off food security problems. It is unthinkable that food security could be sustained against the background of the current cost dispensation. A substantial part of production is done on irrigated lands with the accompanying cost of electricity (Fourie and Botha, 2011).

Since world food demand is growing continuously, new wheat varieties must have higher yield potential, tolerate warmer temperatures, and have improved water-use efficiency or

drought tolerance due to climate change and the dwindling supply of irrigation water. Recent genetic improvements of the yield of irrigated spring wheat lines reported by CIMMYT have been associated with significantly higher biomass, suggesting that either crops are becoming more water use efficient or are extracting more water from their environment (Pask and Reynolds, 2013).

Wheat yield stagnation has been reported in the past decade from regions such as South Asia (Mehla *et al*., 2000) and Europe (Brisson *et al*., 2010). Cargnin *et al*. (2008) found an annual yield progress between 1976-2005 of 1.84% in Brazil. Giunta *et al*. (2007) specified two periods. The period between 1950-1973 produced an increase of 1.7% and the second period (1974-2000) produced a 0.69% annual increase. Afridi and Khalil (2007) reported an annual increase of 1.08% per year over 45 years (1960-2005) in Iran.

There is currently no information available on genetic improvement (past or present) for wheat under irrigation in South Africa. Thus the objectives of the present investigation were to:

i. provide novel information about wheat yield improvement from the SA wheat irrigation breeding and cultivar programmes, more precisely from the Agricultural Research Council – Small Grain Institute (ARC-SGI) of the past 16 years (1995-2010);
ii. develop a method to easily quantify the trend in the grain yield of wheat cultivars;
iii. estimate genetic gains or losses in the additional agronomic traits of hectolitre mass (HLM) and protein content.
iv. compare the AMMI and GGE biplots in assessing GEI;
v. to investigate the relationship amongst the variables measured and/or factors (years, localities and genotypes).

In order to achieve these objectives for the irrigation trials, breeding and cultivar data from the ARC-SGI cultivar development programmes from the four distinct production regions were used. These trial dates ranged from 1995-2010. The SA production regions were the cooler central region, the warmer northern region, KwaZulu Natal; and the eastern Free State irrigation region.

5.2 Materials

Spring wheat under irrigation in the summer rainfall region is planted in late autumn, winter and early spring. Planting date varies over a period of 4 months leaving room for varietal development, including a range of maturities and adaptation characteristics.

Varieties differ in adaptation due to different factors, including planting dates and production potential. Varietal recommendations take this into account. Representative trials are planted in four regions:

i. cooler central irrigation region
ii. warmer northern irrigation region
iii. eastern Free State irrigation region
iv. KwaZulu Natal irrigation region.

In 2010 the cooler central areas (including the Orange, Vaal, Riet, Sand and Vet river irrigation areas) produced around 50% of the wheat under irrigation. The warmer northern areas account for 40% of the wheat production under irrigation, with the other areas supplying the balance of the production under irrigation (ARC-SGI Report, 2010).

5.2.1 Elite field trials

For this study, data for genotypes from the ARC-SGI irrigation trials in SA were used. These trials focused on two of the four geographical regions, namely, the cooler central and warmer northern regions (**Table 5.1**). The cooler central region was divided into two planting dates and the data was evaluated for continuity over 16 years (1995-2010). The trials in the warmer northern region consisted of only one planting date. The geographical position of the localities is depicted in the map (**Figure 5.1**).

In the cool and warm regions there were 345 and 335 genotypes investigated respectively and each trial included the same ARC-SGI check. Standard cultivation practices were followed. The experimental layout was a randomised complete block design (RCBD) with three replications. Yield (ton per hectare) and hectolitre mass (HLM) were measured in these trials.

Table 5.1 Listing of the ARC-SGI elite irrigation trials used in the study

Cooler central region	Warmer northern region
Riet river	Koedoeskop
Vaalharts	Loskop

5.2.2 Cultivar field trials

Data for genotypes from the ARC-SGI cultivar wheat yield trials in the irrigation areas were used in this study. Trials of three of the four geographical regions namely cooler central, eastern and warmer northern regions were studied (**Table 5.2**). The geographical position of the localities is depicted on the map (**Figure 5.1**). Each region was furthermore divided into different planting dates and the data was screened for continuity over 13 years (1998-2010) and consequently only two planting dates could be incorporated in this study. Data of the two planting dates were analysed separately. Although the same cultivars were used in all the localities, not all cultivars were equally represented and each trial included the same ARC-SGI check. Standard cultivation practices were followed in all of the regions. The experimental layout was a randomised complete block design with four replications. Measurements made on the trials and investigated in this study were yield in ton per hectare, HLM and protein content.

Table 5.2 Listings of the ARC-SGI cultivar irrigation trials used in the study

Cooler central region	Eastern region	Warmer northern region
Barkly West	Bethlehem	Atlanta
Bull Hill	Ladybrand	Brits
Christina		Koedoeskop
Douglas		Loskop
Hopetown		
Prieska		
Rama		
Remhoogte		
Riet river		
Upington		

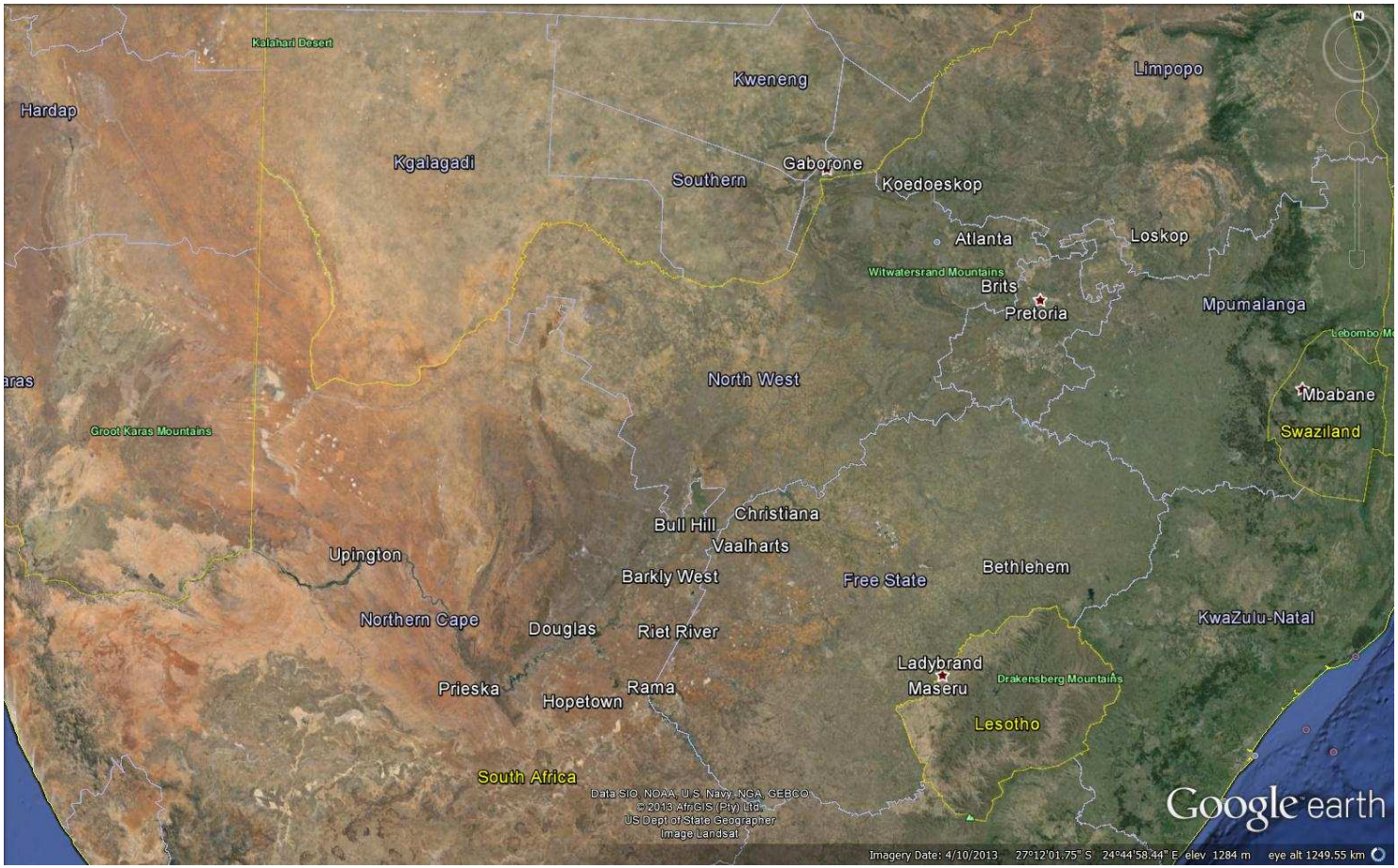


Figure 5.1 Google Map of the localities in the Irrigation regions.

5.3 Statistical techniques

Three statistical methods were compared to estimate genetic advance namely:

iv. Linear regression over years described by Trethowan *et al*. (2002);
v. Variance component method described by Comstock and Moll (1963);
vi. Variance component method described by Allard (1960).

Several other statistical techniques were used to investigate the relationship among the different factors and/or variables:

ix. Genotypes and environments ― AMMI and GGE biplot analyses were performed to evaluate the GEI and to compare the two methods for a single locality. Years within the locality were used as environments;
x. Pearson's product moment correlation was performed to determine the relationship among the three variables;
xi. PCA was performed to investigate the relationship between the factors (i.e. genotypes) and the variables;
xii. DA was used to determine whether the three variables could discriminate among the years within a single locality.

5.3.1 The linear regression method over years (TRET)

The first method to determine genetic improvement was a linear regression method described by Trethowan *et al*. (2002). The statistical procedure of this method was described in Section 3.3.1. The graphic presentations were done with Excel to provide a trend with R^2 values. The p values on the graphic presentations were provided by the regression generated by the PROC REG procedure of the software SAS (SAS Institute, 2012).

5.3.2 Sources of variation, heritability and genetic advance

Other methods to determine genetic advance are using sources of variation. This can either be fixed effects models or random effects models. In **Figure 3.2** the layout of the different methods to calculate genetic advance is shown. The statistical procedure of this method was described in Section 3.3.3.

5.3.2.1 Linear fixed model with expectations of the mean squares (Model 1)

The data was subjected to an ANOVA and analysed by the PROC GLM procedure of the SAS software (SAS Institute, 2012). Variance components were estimated using TYPE III expected mean squares of the ANOVA as pointed out in **Table 3.2** and discussed in Section 3.3.4.

5.3.2.2 Linear mixed model framework to calculate variance components and genetic advance (Model 2 to Model 5)

The statistical procedure of this method was described in Section 3.3.5.

The data was analysed by the PROC MIXED procedure of the SAS software (SAS Institute, 2012). The restricted (or residual or reduced) maximum likelihood (REML) was chosen as the estimation method.

5.3.3 The second variance components model (Model 6)

The statistical procedure of this method was described in Section 3.3.8.

This model (model 6) was tested in a linear fixed and linear mixed model environment.

5.3.4 AMMI versus GGE

These models are discussed in Section 3.3.9 and Section 4.3.4.

AMMI and GGE biplots were constructed using GenStat 15th edition (Payne *et al*., 2012). AMMI analysis was based on the model by Gauch and Zobel (1996) and GGE was based on the model for two principal components according to Yan and Kang (2003).

5.3.5 Other statistical techniques

The Pearson's product moment correlation matrix of the pairwise correlations among the three dependent variables was constructed to show their linear relationships. The two major components of PCA were then graphically represented to show the relationship

among the factors (genotypes) and the variables. Hierarchical cluster analysis of the environments was performed to identify the similarities among the environments. DA was performed to distinguish environments with the three variables. These methods were performed using XLSTAT software (2012).

5.4 Results and discussion

Preliminary ANOVA with PROC GLM of the software SAS (2012) using Error Type III was carried out for individual trials (each locality and year) to assess the residuals and predicted values. The residuals were tested by the Shapiro and Wilk test for normality (Shapiro and Wilk, 1965). Outliers (any residual value above the absolute value of three (ABS(rijkl) > 3) were replaced by the predicted values. Homogeneity of variances was tested by Levene's test (Levene, 1960). Heterogeneity in year variances and the interactions with year, genotype x year, genotype x location, and genotype x year x location were found. The occurrence of heterogeneity in multi-environments and possible solutions was discussed in Section 3.4.2

5.4.1 Linear regression over years using TRET

In this study mean yield for genotypes in experimental years were used to express genetic progress. A linear regression over time for each locality was performed (results not shown except where a significant trend was established). To facilitate comparison among the different methods to determine genetic advance, it was decided to perform the linear regression over localities in a region. Homogeneity of variances for localities (sites) within a region was found. The data of the elite trials and the cultivar trials were subjected to linear regression analysis for each region over time, according to Trethowan *et al*. (2002). In both trials non-recurrent genotypes occurred, leading to unbalanced data. The method proposed by Trethowan *et al.* (2002) renders a solution to the unbalanced nature of the data. The mean of the best five genotypes was expressed as a ratio (from now on called the Ratio) of the trial mean (TM). The same check was used throughout the period and for all localities in both the elite and cultivar trials. The modifications described in Section 3.3.1 – namely M_T (mean of five highest yielding genotypes minus trial mean), M_C (mean of five highest yielding genotypes minus check mean) – were relevant and therefore tested.

5.4.1.1 Elite trials

The TM, the check means (Check), Ratio and Mc across all environments within the regions are presented graphically. It was that found M_T added no extra value to the study. No significant trend in either Ratio or Mc was found for planting date 1 and planting date 2 of the cooler region of the elite trials (**Figure 5.2 and Figure 5.3**).

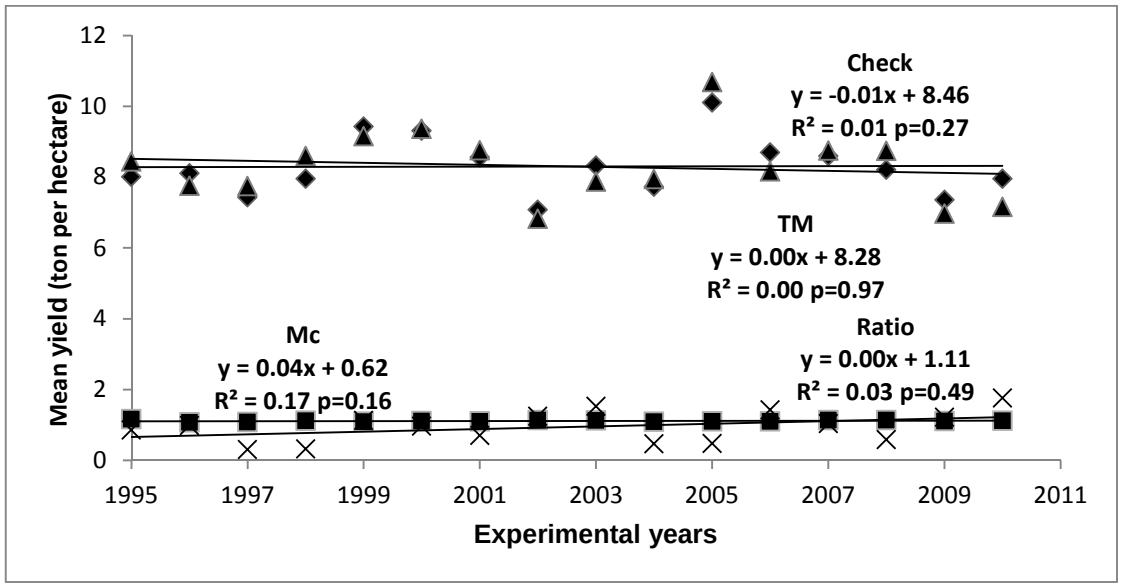


Figure 5.2 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the elite trial of the cool region for planting date 1. ♦ =trial means (TM); ▲ = Check; **X** = Mc; ■ = Ratio.

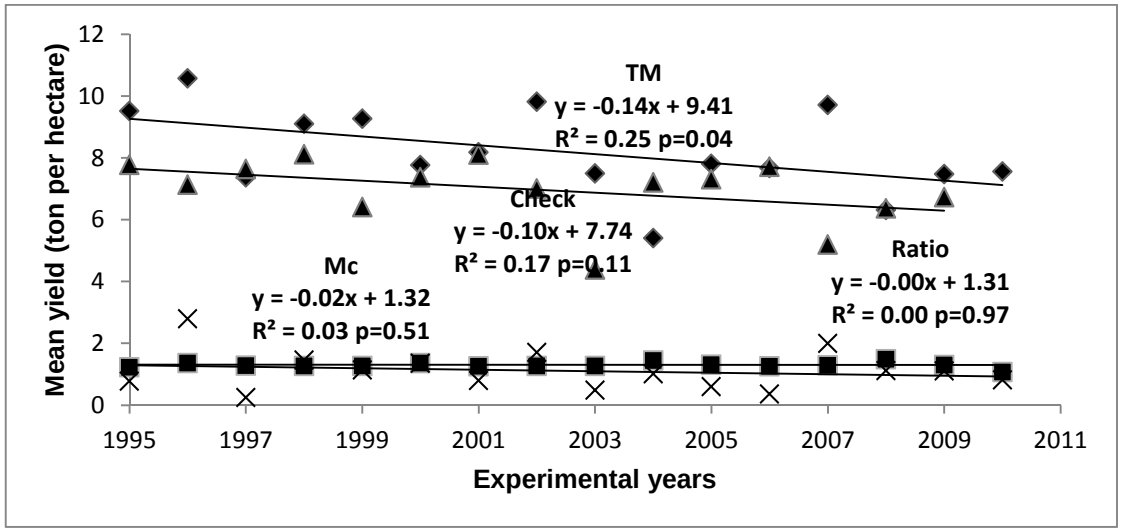


Figure 5.3 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the elite trial of the cool region for planting date 2. ♦ = trial means (TM); ▲ = Check; **X** = Mc; ■ = Ratio.

A significant trend at 10% was found for mean yield of Mc from the trials of the elite warm region with R^2 = 0.40 (**Figure 5.4**) and a regression coefficient, b = 0.10, which indicated genetic progress of 0.70% (b=0.10 15 years were involved, thus GA=0.10/15*100=0.70% per year).

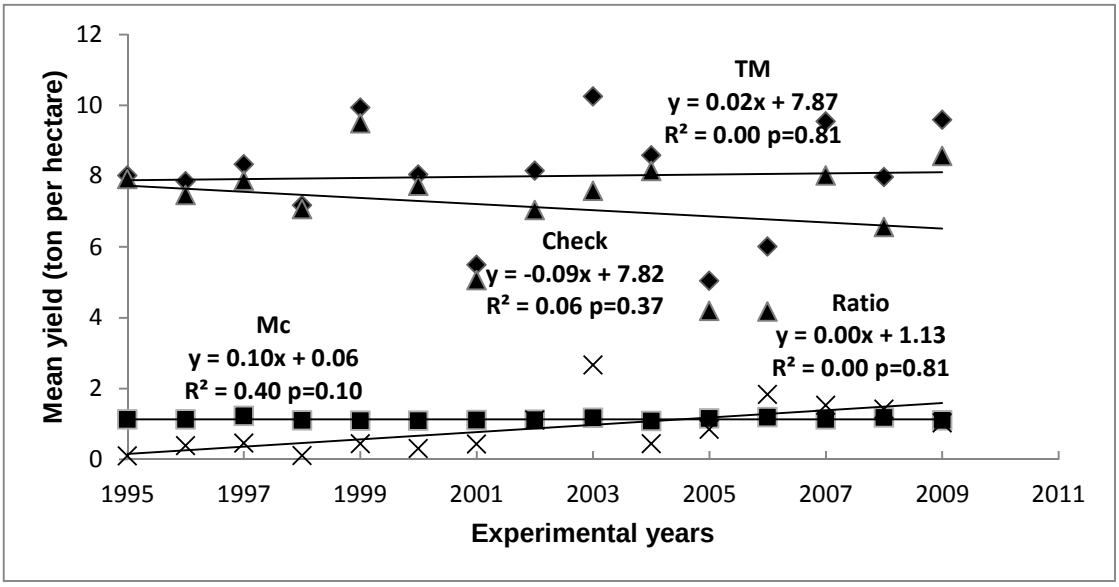


Figure 5.4 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the elite trials of the warm region. ♦ = trial means (TM); ▲ = Check; X = Mc; ■ = Ratio.

5.4.1.2 Cultivar trials

In this section of the study only TM, Check and Ratio were portrayed in the graphic presentations. Mc and M_T added no value to the research. In **Figure 5.5** the mean yield for Ratio displayed no trend in the cool regions for both planting dates in the cultivar trials. The R^2 values were respectively R^2 = 0.02 with p = 0.50 and R^2 = 0.12 with p = 0.25.

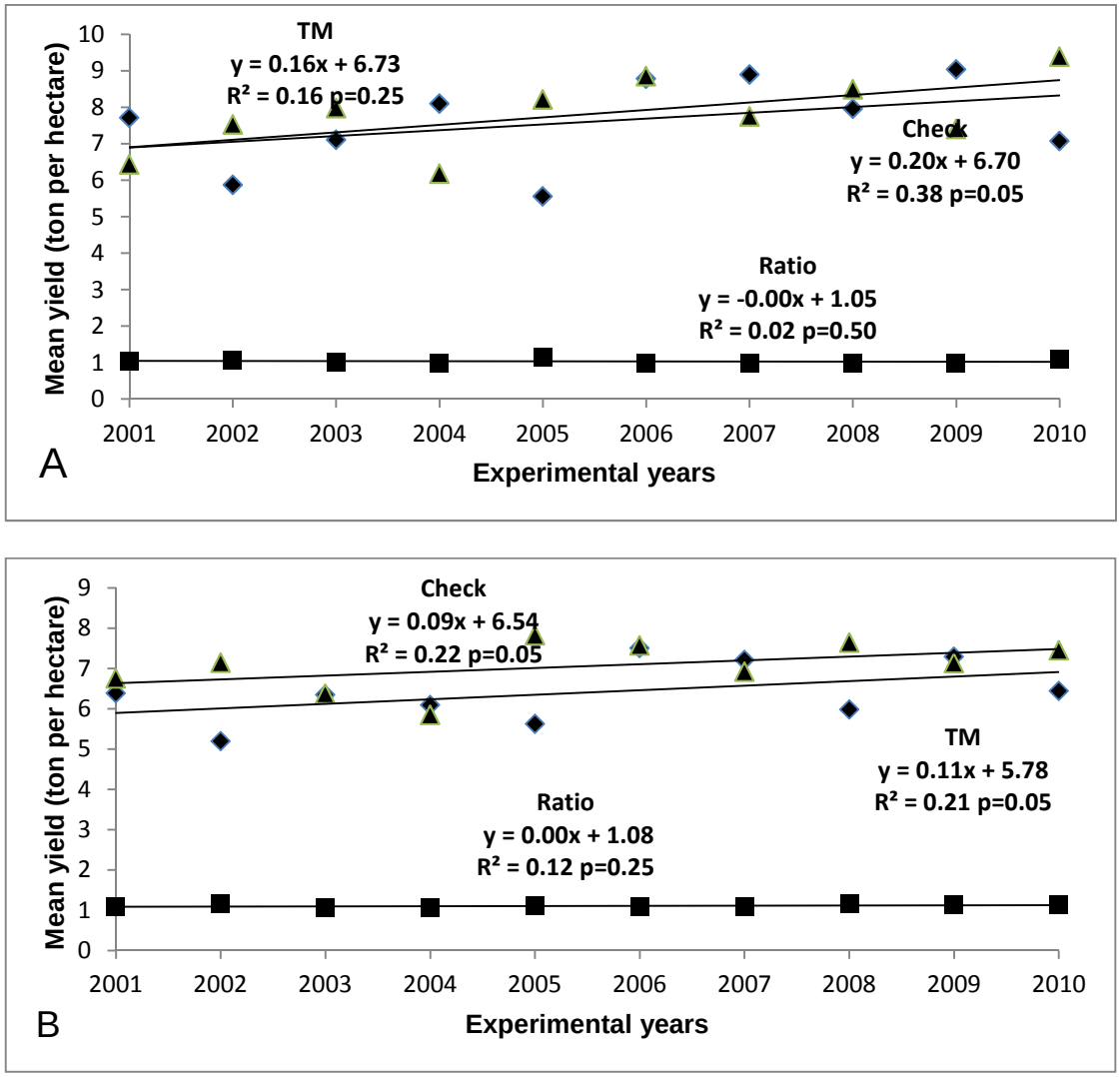


Figure 5.5 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trial of the cool region for A, the first planting date, and B, the second planting date. ♦ = trial means (TM); ▲ = Check; ■ = Ratio.

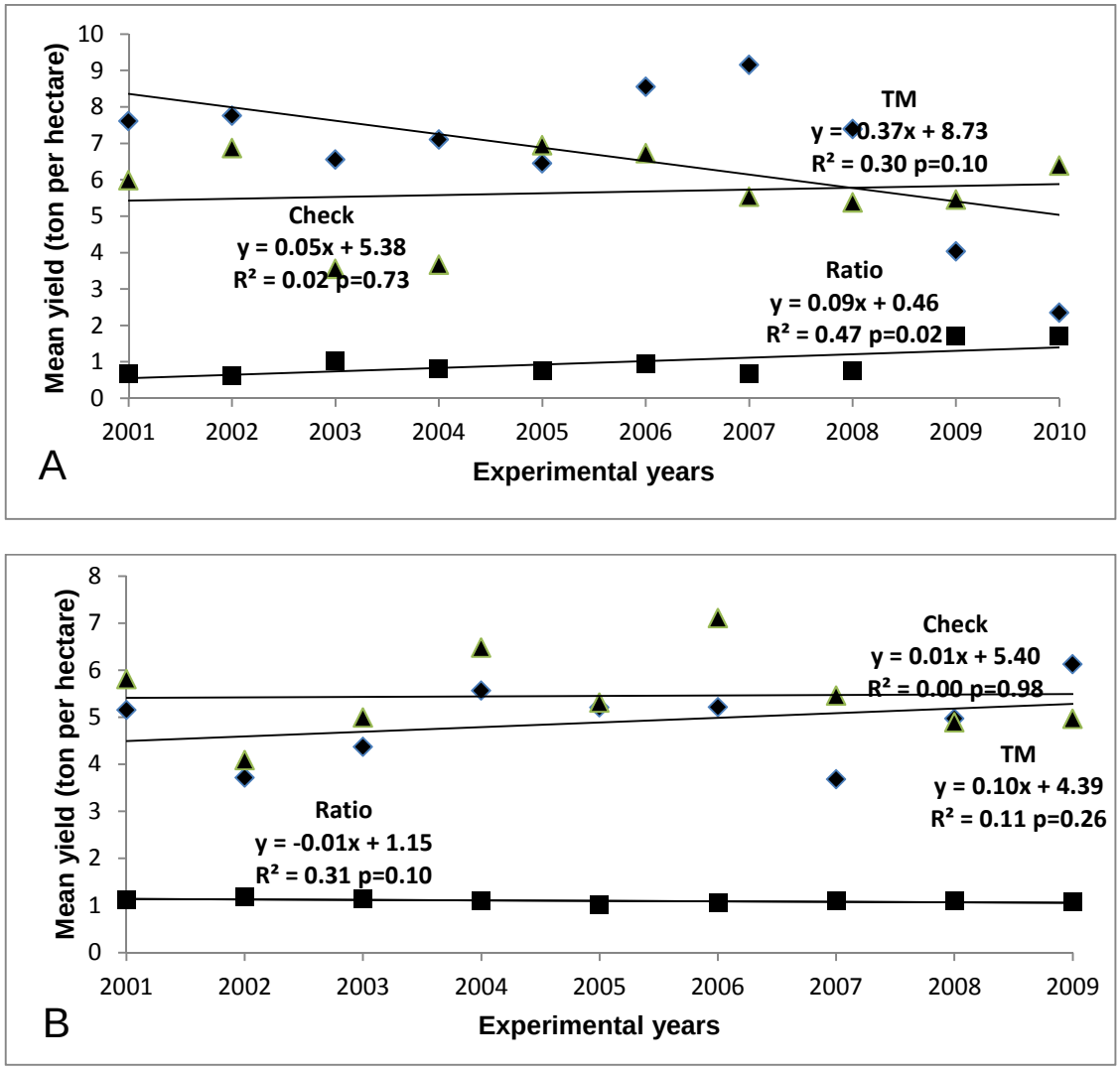


Figure 5.6 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trial of the eastern Free State region for A, the first planting date, and B, the second planting date. ♦ = trial means (TM); ▲ = Check; ■ = Ratio

The progress in mean yield for Ratio in the eastern region for planting date 1 (**Figure 5.6 A**) was b = 0.09 with a significant R^2 = 0.47 and p = 0.02. The mean yield for Ratio in the second planting date (**Figure 5.6B**) was significant at 10% with R^2 = 0.31 and showed a negative trend (b =- 0.01). No progress in mean yield for Ratio in the warm region for planting date 1 (**Figure 5.7A**) was observed. The slope value b = - 0.06 with a non-significant R^2 = 0.18 and p = 0.26. However, a significant positive trend was established for mean yield for Ratio in the second planting date (**Figure 5.7B**) with R^2 = 0.67 p = 0.02.

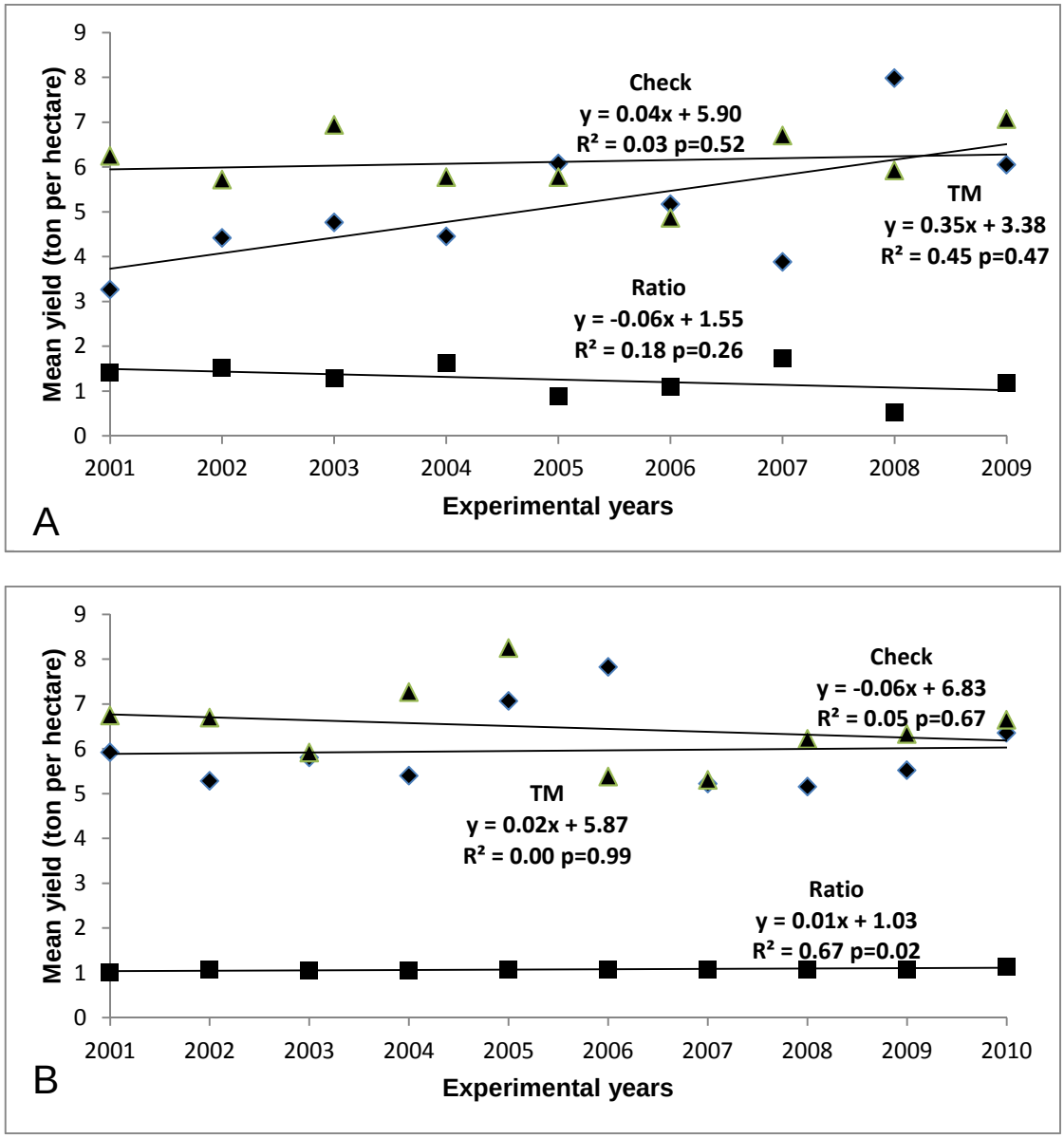


Figure 5.7 Mean yield of the five highest yielding genotypes expressed as a ratio of the trial mean of the cultivar trial of the warm region for A, the first planting date, and B, the second planting date. ♦ = trial means (TM); ▲ = Check; ■= Ratio.

The evaluation of historical series of genotypes in irrigation trials is a common approach used in several countries to quantify the achievements of past breeding efforts (Cargnin *et al.,* 2008; Giunta *et al.,* 2007).

The present study examined a set of approximately 340 bread wheat genotypes in the elite trials and 38 bread wheat genotypes in the cultivar trials. The methods used to

combat environmental fluctuations were Ratio (TRET) and Mc. Elite trials Mc showed the progress or lack thereof successfully. In the cultivar trials, Ratio was a far better predictor of the progress than Mc.

Genetic improvement determined by linear regression could only be shown in the warm region of the elite trials, the eastern region for planting date 1 and the warm region, planting date 2 of the cultivar trials. A negative trend was observed for planting date 2 of the eastern region's cultivar trials. Similar results, namely a decline in genetic progress, were found by Mehla *et al.* (2000) in South Asia and by Brisson *et al*. (2010) in Europe, more precisely in France. In contrast, genetic improvement was found in Brazil (Cargnin *et al.*, 2008), Italy (Giunta *et al*., 2007), and in Iran (Afridi and Khalil, 2007). Oury *et al.* (2012) stated that genetic progress has been partly or totally counterbalanced by adverse effects of climate change.

5.4.2 Sources of variation, heritability and genetic advance

The mean squares obtained from the combined analysis of variance showed all effects were significant ($p<0.0001$) for yield (Yld), HLM, protein content (Protein), logarithm transformed variables of yield (LYld), HLM (LHLM), protein content (LProtein), and the yield, HLM and protein content from the weighted analyses (WYld, WHLM and WProtein) respectively from the elite trials and cultivar trials of the different regions for both planting dates (Appendix C Tables C1 to Table C12).

5.4.3 Comparison of linear fixed models and linear mixed models

The genetic advance values of model 6 were determined in the linear fixed model and linear mixed model scenarios for the three variables and the logarithmic transformed variables. The variables were subjected to weighted analyses in the linear fixed milieu not in the linear mixed environment, which were depicted in **Table 5.3** to **Table 5.7**. The linear regression coefficient b (TRET), genetic advance (ΔG) of the fixed model (model 1), mixed model (model 4) and the model using heritability (model 6) for yield are compared.

Table 5.3 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) calculated from the various proposed models for yield (Yld) of the elite irrigation trials (LYld = logarithmic transformed yield, WYld = weighted analysis of yield)

			Fixed Effects Model			Mixed Effects Model			
			% ΔG			%ΔG			
Planting date	**Region**	**Variable**	**H^2**	**Model 1**	**Model 6**	**H^2**	**Model 4**	**Model 6**	**TRET**
1	Cool	Yld	0.15	1.29	0.16	0.05	2.78	2.34	Ns
1	Cool	LYld	0.12	1.36	0.02	0.09	4.10	1.01	
1	Cool	WYld	0.15	1.28	0.23	0.00			
2	Cool	Yld	0.00	0.36	-0.01	0.04	2.05	0.37	Ns
2	Cool	LYld	0.10	0.07	0.01	0.00	0.00	0.00	
2	Cool	WYld	0.00	-1.30	-0.19	0.00			
1	Warm	Yld	0.09	0.75	0.10	0.07	3.83	0.91	0.7
1	Warm	LYld	0.02	0.14	0.00	0.05	3.25	3.26	
1	Warm	WYld	0.02	1.64	0.31	0.00			

Ns = no significant trend

Table 5.4 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) calculated from the various proposed models for hectolitre mass (HLM) of the elite irrigation trials (LHLM = logarithmic transformed HLM and WHLM = HLM from weighted analysis)

			Fixed Effects Model			Mixed Effects Model		
			% ΔG			% ΔG		
Planting date	**Region**	**Variable**	**H^2**	**Model 1**	**Model 6**	**H^2**	**Model 4**	**Model 6**
1	Cool	HLM	0.02	2.38	0.44	0.12	7.06	2.50
1	Cool	LHLM	0.02	2.44	0.00	0.00	0.00	0.00
1	Cool	WHLM	0.00	0.00	0.00			
2	Cool	HLM	0.03	2.00	0.44	0.34	10.06	6.00
2	Cool	LHLM	0.03	1.94	0.00	0.32	9.69	2.38
2	Cool	WHLM	0.04	2.63	0.63			
1	Warm	HLM	0.02	1.47	0.33	0.28	8.40	5.20
1	Warm	LHLM	0.02	1.47	0.00	0.28	8.33	2.13
1	Warm	WHLM	0.06	2.87	0.87			

Table 5.5 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) calculated from the various proposed models for yield of the cultivar irrigation trials (LYld = logarithmic transformed yield, WYld = weighted analysis of yield)

			Fixed Effects Model			Mixed Effects Model			
				% ΔG			% ΔG		
Planting date	**Region**	**Variable**	**H^2**	**Model 1**	**Model 6**	**H^2**	**Model 4**	**Model 6**	**TRET**
1	Cool	Yld	0.00	0.30	0.10	016	9.50	0.32	Ns*
1	Cool	LYld	0.00	0.70	0.00	0.13	8.70	1.32	
1	Cool	WYld	0.03	4.10	0.80				
1	East	Yld	0.01	1.20	0.30	0.17	10.00	2.50	9.00
1	East	LYld	0.06	3.90	0.20	0.75	45.90	20.10	
1	East	WYld	0.07	5.50	1.90				
1	Warm	Yld	0.01	1.40	0.20	0.20	10.60	2.70	
1	Warm	LYld	0.01	1.00	0.00	0.18	10.20	14.00	
1	Warm	WYld	0.03	3.50	0.70				
2	Cool	Yld	0.03	3.70	0.60	0.24	12.70	5.00	Ns*
2	Cool	LYld	0.03	4.10	0.10	0.25	14.50	24.60	
2	Cool	WYld	0.03	4.30	0.90				
2	East	Yld	0.03	3.10	0.30	0.14	8.90	1.80	-1.00
2	East	LYld	0.02	2.60	0.10	0.15	9.40	12.80	
2	East	WYld	0.03	3.20	0.70				
2	Warm	Yld	0.01	1.80	0.20	0.14	8.50	1.80	Ns*
2	Warm	LYld	0.01	2.10	0.00	0.14	8.50	1.80	
2	Warm	WYld	0.01	1.70	0.30				

*TRET Ns = no significant trend

Table 5.6 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) calculated from the various models proposed for hectolitre mass (HLM) of the cultivar trials (LHLM = logarithmic transformed HLM and WHLM = HLM from weighted analysis)

			Fixed Effects Model			Mixed Effects Model		
				% ΔG			% ΔG	
Planting date	Region	Variable	H^2	Model 1	Model 6	H^2	Model 4	Model 6
1	Cool	HLM	0.03	2.40	6.30	0.17	10.50	4.40
1	Cool	LHLM	0.04	2.80	0.50	0.17	10.50	1.80
1	Cool	WHLM	0.03	4.10	0.80			
1	East	HLM	0.06	3.80	11.80	0.79	50.60	35.00
1	East	LHLM	0.06	3.80	0.70	0.82	56.30	16.40
1	East	WHLM	0.28	14.80	8.10			
1	Warm	HLM	0.01	1.30	2.70	0.11	7.50	1.10
1	Warm	LHLM	0.02	1.60	0.20	0.13	8.00	2.90
1	Warm	WHLM	0.03	3.70	0.70			
2	Cool	HLM	0.03	4.90	1.00	0.19	14.60	6.90
2	Cool	LHLM	0.03	4.60	0.00	0.16	13.80	2.50
2	Cool	WHLM	0.03	5.30	1.00			
2	East	HLM	0.04	3.90	1.00	0.22	12.60	6.10
2	East	LHLM	0.04	3.70	0.00	0.21	12.40	2.50
2	East	WHLM	0.03	3.00	0.80			
2	Warm	HLM	0.01	1.80	0.30	0.11	6.90	2.90
2	Warm	LHLM	0.01	1.60	0.00	0.11	6.90	2.90
2	Warm	WHLM	0.02	1.70	0.30			

Table 5.7 Estimates of broad sense heritability (H^2) and genetic advance (ΔG) calculated from the various models proposed for protein content (Prot) of the cultivar trials (LProt=logarithmic transformed protein content, WProt=weighted analysis of protein content)

			Fixed Effects Model			Mixed Effects Model		
				%ΔG			%ΔG	
Planting date	**Region**	**Variable**	**H^2**	**Model 1**	**Model 6**	**H^2**	**Model 4**	**Model 6**
1	Cool	Prot	0.04	2.73	0.27	0.02	7.73	1.60
1	Cool	LProt	0.04	2.73	0.00	0.02	7.73	4.60
1	Cool	WProt	0.03.	2.47	0.47	0.00	0.00	0.00
1	East	Prot	0.07	3.20	0.67	0.02	7.60	2.60
1	East	LProt	0.07	3.20	0.07	0.02	7.47	5.93
1	East	WProt	0.07	3.33	1.07	0.00	0.00	0.00
1	Warm	Prot	0.04	2.40	0.33	0.02	8.80	2.67
1	Warm	LProt	0.04	2.40	0.00	0.02	8.53	6.87
1	Warm	WProt	0.04	2.33	0.53	0.00	0.00	0.00
2	Cool	Prot	0.02	1.53	0.20	0.01	3.00	0.53
2	Cool	LProt	0.04	2.13	0.00	0.04	4.87	2.33
2	Cool	WProt	0.06	3.47	1.00	0.00	0.00	0.00
2	East	Prot	0.02	1.07	0.13	0.60	3.53	0.80
2	East	LProt	0.01	0.60	0.00	0.67	4.13	2.33
2	East	WProt	0.02	1.33	0.27	0.00	0.00	0.00
2	Warm	Prot	0.02	1.53	0.13	0.02	7.67	2.20
2	Warm	LProt	0.02	1.47	0.00	0.02	7.67	2.20
2	Warm	WProt	0.04	2.33	0.67	0.00	0.00	0.00

Interpreting the genetic advance (more precisely change in genetic advance) values in **Table 5.3** to **Table 5.7** for yield, HLM and protein content was difficult. Visually it shows a specific picture. Statistically speaking it tells us nothing because there is no standard error to describe the variation around a specific value. To recommend one of these models for predicting genetic advance, a comparator should be found. Pearson's product moment correlation and/or Cronbach's alpha (Gregory, 2000) are used in psychometrics to evaluate agreement (correlation) among different tests (in layman's terms: do the tests test the same construct?). In this study the construct was genetic advance from model 1 (the linear fixed effects model), model 4 (linear mixed model), model 6 (method according to Allard) and TRET (the slope from the linear regression). The genetic

advance values of model 6 were calculated in a linear fixed environment (called model 6F) and from the linear mixed scenario (called model 6R).

The agreement (correlation) between yield for model 1 and model 6F was strong and significant ($r = 0.89$, $p = 0.00$). The correlation coefficient between model 1 and model 6R was significant at 10% ($r = 0.59$, $p = 0.08$).The agreement between model 6F and model 4 was moderate ($r = 0.69$, $p = 0.03$). The Pearson's correlation coefficient between model 4 and model 6R was $r = 0.75$ with a $p = 0.02$. Cronbach's alpha = 0.93 when omitting TRET in the comparison of the models. This is an indication that TRET was not consistent with the other models.

For the genetic advance values of logarithmic transformed yield a significant agreement was found between model 1 and model 6F of $r = 0.82$ and $p = 0.00$, model 1 and model 4 with $r = 0.68$ and $p = 0.04$, model 1 and model 6R with $r = 0.60$ and $p = 0.08$. The agreement of Model 6F with model 4 was $r = 0.86$ with $p = 0.00$, model 6F with $r = 0.86$ and $p = 0.00$, model 4 and model 6R with $r = 0.96$, $p<0.00$. Cronbach's alpha = 0.97. Agreement of the genetic advance values from the weighted analysis of yield was found for model 1 and model 6F was $r = 0.90$ and $p<0.00$.

TRET was not performed for HLM and protein content; therefore only the variance components methods (models 1, 4 and 6) were compared for the quality traits. For HLM agreement was found between model 6F and model 4 ($r = 0.83$ $p = 0.00$), model 6F and model 6R ($r = 0.80$ $p = 0.00$). Cronbach's alpha = 0.93. For the genetic advance values of logarithmic transformed HLM (LHLM) no significant agreement occurred between model 1 and any of the other models. Significant agreement of model 6F with model 4 and 6R of $r = 0.73$, $p = 0.00$ and $r = 0.72$, $p = 0.02$, respectively, were found. A strong agreement between model 4 and model 6R of $r = 0.98$ $p<0.00$ was found. Cronbach's alpha = 0.91. Agreement of the genetic advance values from the weighted analysis of HLM was found for model 1 and model 6F and was $r = 0.97$ and $p<0.00$.

For the genetic advance of protein content a significant agreement was found between model 1 and model 6F ($r = 0.86$, $p<0.00$) and between model 4 and model 6R with r =

0.93 and p<0.00. Cronbach's alpha = 0.94. No significant agreement was found among the models of the genetic advance values for the logarithmic transformed protein content. For weighted protein content an agreement of r = 0.94 p<0.00 between model 1 and model 6F was found.

Selection efficiency is related to magnitude of heritability and genetic advance (Johnson *et al*., 1955). In this study low broad sense heritabilty (not given in %) and low genetic advance (-0.00 to 2.50, given in % in the tables), depending on the model and variable were found. The low heritabilty estimates suggest that environmental effects constituted a major portion of the total phenotyic variation in these traits. Simlar results were found by Kashif and Khaliq (2004), Bilgin *et al*. (2011), as well as Khan and Naqvi (2011).

5.4.4 AMMI and GGE

Investigating GEI by AMMI and GGE analyses, is important to select stable and adapted genotypes. To illustrate the simarilities and dissimilarities of above mentioned methods, 10 genotypes over 6 years (environments) from a single locality (Barkly West) in the cool region were submitted to these methods. The prerequisite of combined ANOVA with balanced data indicated highly signficant environment, genotype and genotype x environment interaction between genotypes and envionments effects (results not shown).

The AMMI ANOVA of yield, HLM and protein content for planting date 1 is shown in **Table 5.8**. The distribution of total sum of squares due to treatments (genotypes + environment + GEI) was 65.26%, 78.00% and 89.31% and accounted for by environments for yield, HLM and protein content respectively. Percentage variation explained by the genotypes for yield, HLM and protein content was 11.79%, 4.34% and 4.34% respectively. The proportion of variation for the interaction was 22.98%, 17.66% and 6.37% respectively.

In **Table 5.9** the percentage of the total treatment variation for environments of 77.00%, 75.84% and 76.60% for yield, HLM and protein respectively is shown. The percentage genotype variation was 8.24%, 6.53% and 11.12% respectively for yield, HLM and

protein content. The percentage of variation for the interaction was 14.76%, 17.63% and 12.28% respectively. The large proportion of variation counted for by environments showed the environments were diverse. It is to be expected because the environments in the example were years. The genotype × environment interaction sum of squares was approximately double that of the genotypes, which determined differences in genotypic response across environments.

Table 5.8 Analysis of variance from AMMI analysis of yield, hectolitre mass (HLM) and protein content of the Barkly West cultivar trials from planting date 1 during 2004-2010

		Yield		HLM		Protein content	
Source	**DF**	**MS**	**% SS**	**MS**	**% SS**	**MS**	**% SS**
Block(Env)	18	0.57		2.94		0.79	
Treatments	**65**	**7.21**		**13.92**		**6.74**	
Environments	**5**	**61.11**	**65.26**	**141.11**	**78.00**	**78.19**	**89.31**
Genotypes	**10**	**5.52**	**11.79**	**3.93**	**4.34**	**1.90**	**4.34**
Interactions	**50**	**2.15**	**22.98**	**3.19**	**17.66**	**0.56**	**6.37**
IPCA1	14	2.68		4.26		0.96	
IPCA2	12	2.55		4.49		0.78	
Residuals	24	1.64		1.92		0.21	
Error	180	0.28		0.72		0.23	
Total	263	2.01		4.13		1.87	

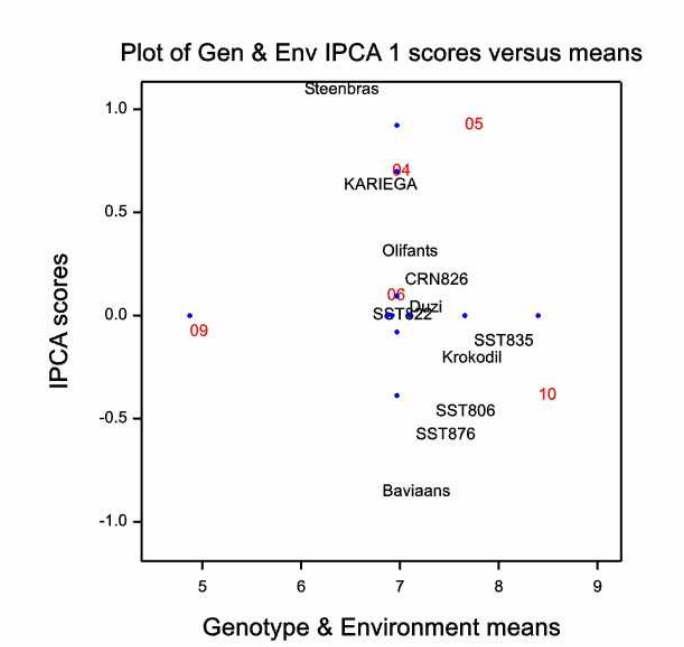
Table 5.9 Analysis of variance from AMMI analysis of yield, HLM and protein content of the Barkly West cultivar trials from planting date 2 during 2004-2010

		Yield		HLM		Protein content	
Source	**DF**	**MS**	**% SS**	**MS**	**% SS**	**MS**	**% SS**
Block(Env)	21	0.53		4.56		0.44	
Treatments	**76**	**10.72**		**24.87**		**2.90**	
Environments	**6**	**104.57**	**77.00**	**238.87**	**75.84**	**28.16**	**76.60**
Genotypes	**10**	**6.71**	**8.24**	**12.34**	**6.53**	**2.45**	**11.12**
Interactions	**60**	**2.00**	**14.76**	**5.55**	**17.63**	**0.45**	**12.28**
IPCA1	15	3.46		13.06		1.08	
IPCA2	13	2.47		4.04		0.48	
Residuals	32	1.13		2.65		0.15	
Error	210	0.38		1.99		0.18	
Total	307	2.95		7.83		0.87	

The AMMI biplot and the GGE biplot in **Figure 5.8** (genotypes and environments for yield for the first planting), at first glance, do not show similarities. The AMMI table of four best genotypes in the six environments (Appendix C Table C13) assisted in interpreting the similarities in the biplots. The GGE biplot shows more clearly three mega-environments

and the “which one wins where” genotypes that are the ones on the vertices of the polygon. Genotype Kariega was the winning genotype in the year 05. In the second mega-environment – formed by years 06, 07 and 10 – the best performing genotype was SST835. In the third mega-environment, containing years 04 and 09, genotype SST806 was the best adapted genotype.

A



B

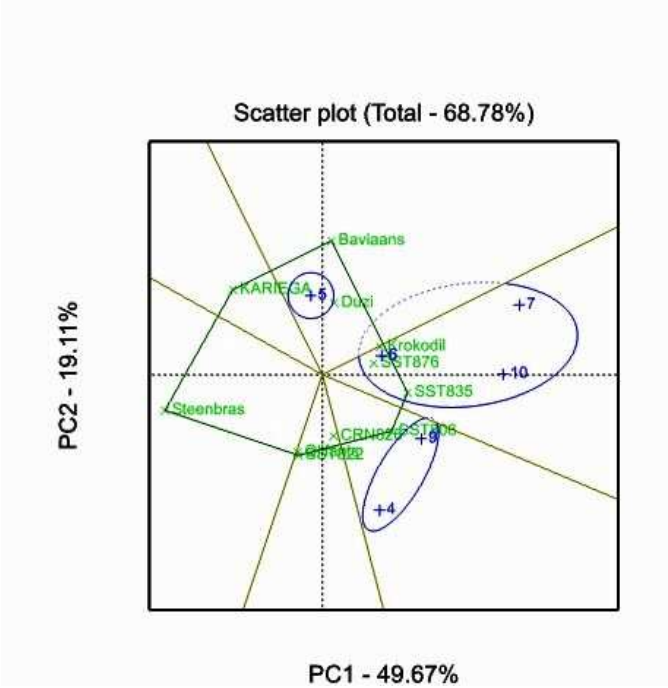
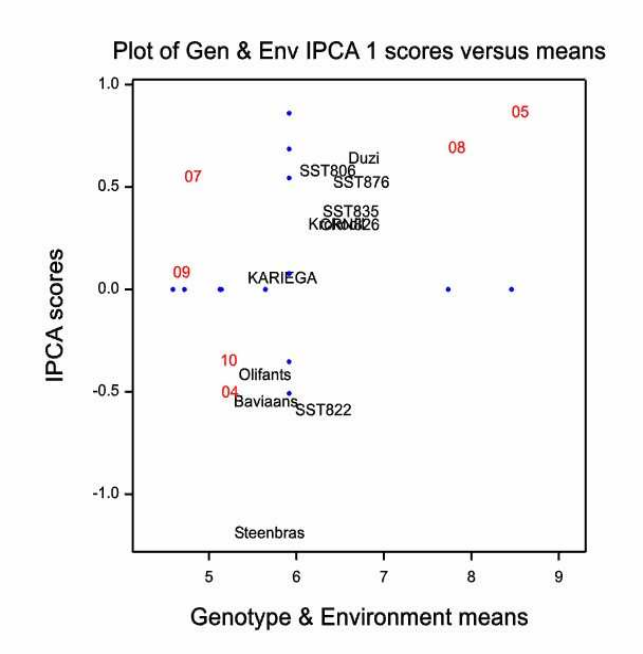


Figure 5.8 Genotype by environment biplot of planting date 1 of Barkly West for yield (A) Principal Component I versus mean yield (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

The GGE biplot **in Figure 5.9 B** shows clearly four mega-environments and the best performing genotypes in year 05 which were SST835 and SST876. The winning genotype in the environment year 06 was SST822 and in year 07 genotype SST806 was the best adapted. In the fourth mega-envronment – formed by years 04, 08, 09 and 10 – the best performing genotype was Duzi. The similarities shown by the AMMI are confirmed by the AMMI selections in Appendix C Table C14.

A



B

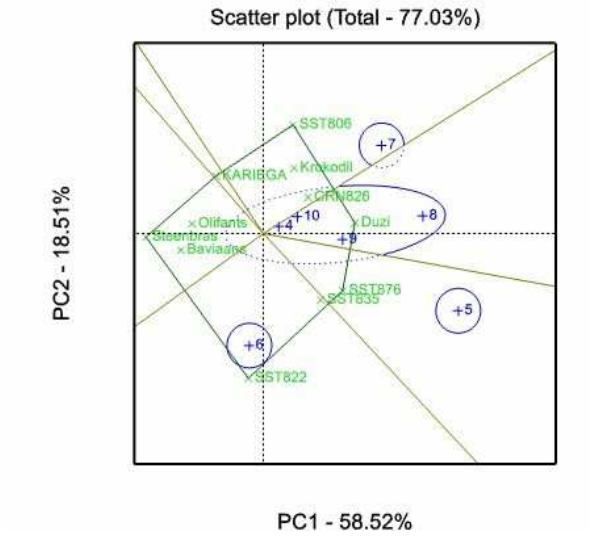
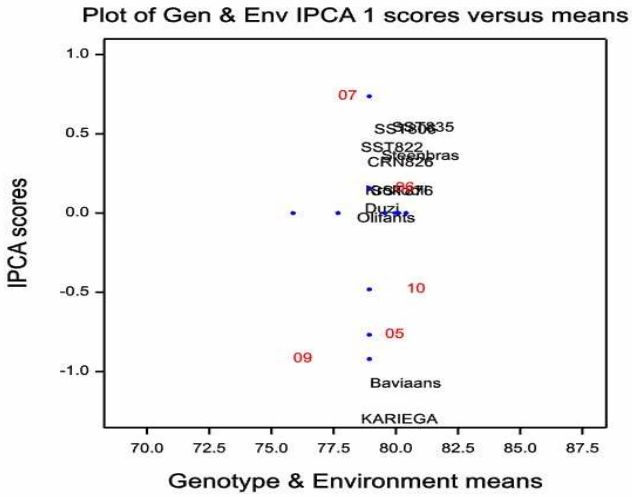


Figure 5.9 Genotype by environment biplot of planting date 2 of Barkly West for yield (A) Principal Component I versus mean yield (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

The AMMI biplot of HLM for the first planting date in **Figure 5.10 A** is not very clear. Consulting the table of the four best genotypes in Appendix C Table 13 confirms the results of the GGE biplot in **Figure 5.10 B**. The first mega-environment is formed by year 05 and year 09 with winning genotype Olifants. The second mega-environment includes years 04 and 07 with best performing genotype SST835. Genotype CRN826 is the best adapted genotype in the third mega-environment (years 06 and 10).

A



B

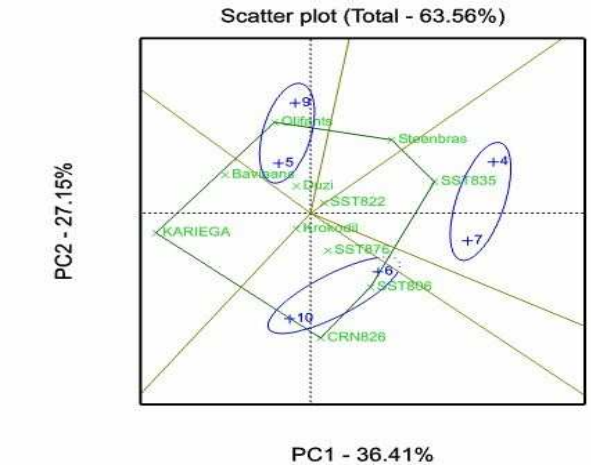
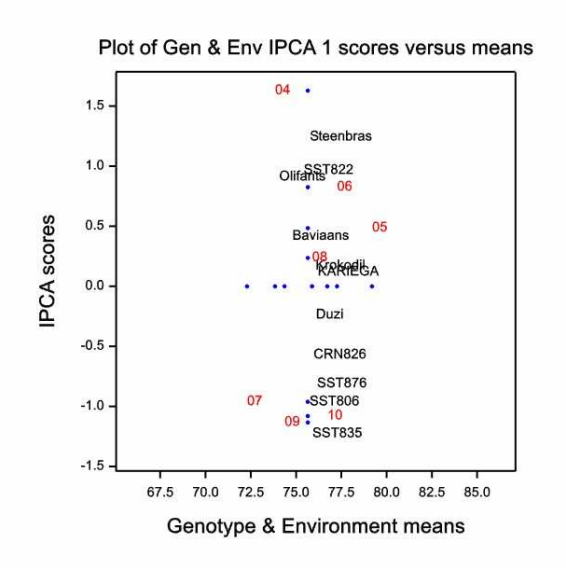


Figure 5.10 Genotype by environment biplot of planting date 1 of Barkly West for hectolitre mass (HLM) (A) Principal Component I versus mean HLM (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

A



B

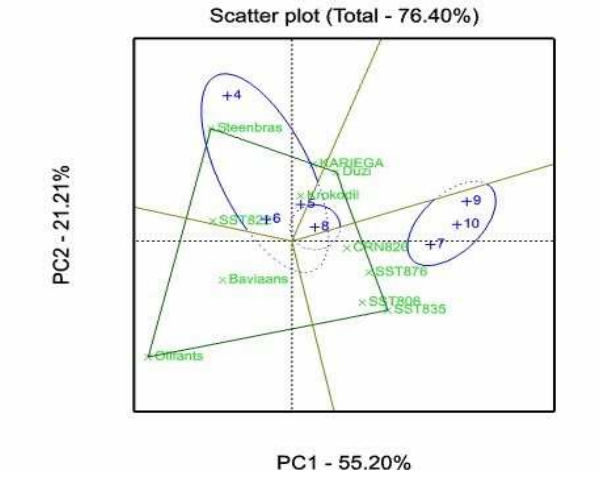
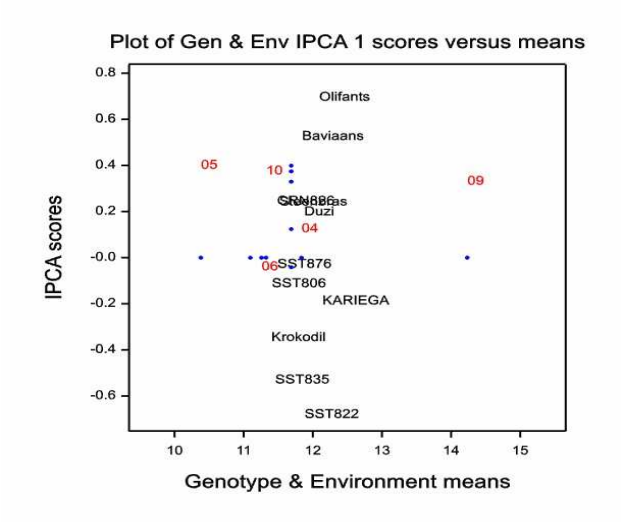


Figure 5.11 Genotype by environment biplot of planting date 2 of Barkly West for hectolitre mass (HLM) (A) Principal Component I versus mean HLM (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

The GGE biplot of HLM second planting for HLM (**Figure 5.11 B**) shows three mega-environments with two mega-environments overlapping. The first mega-environment includes years 04, 05 and 06 overlapping with year 08. In these mega-environments, the best performing genotype was Steenbras. The third mega-environment is formed by years 07, 09 and 10 with best performing genotypes SST806 and SST835. The AMMI

biplot shows similar results. In the GGE biplot in **Figure 5.12 B** three mega-environments are visible. The genotype SST822 proved to be the best performer in mega-environment 1 (year 07). The AMMI biplot (**Figure 5.12 A**) shows the best performing genotype is Olifants for years 04, 05, 09 and 10. These years form a mega environment in the GGE biplot (**Figure 5.12 B**).

A



B

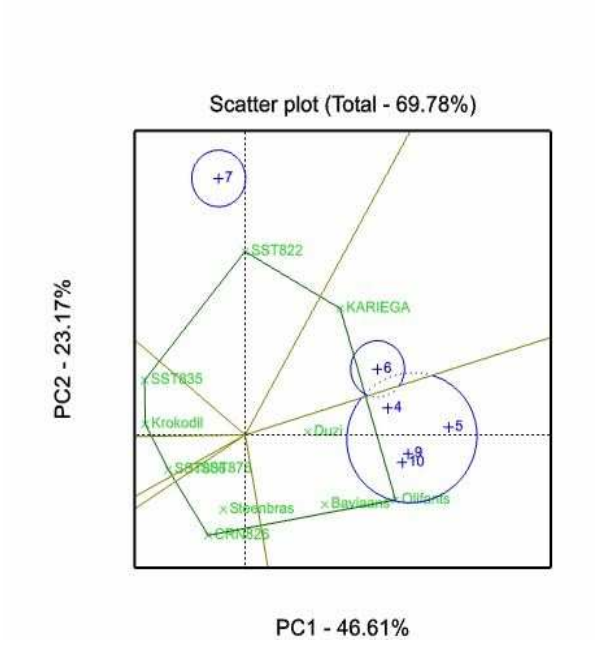
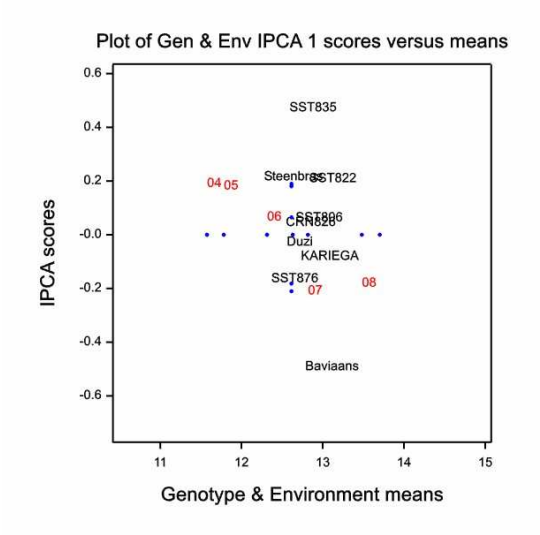


Figure 5.12 Genotype by environment biplot of planting date 1 of Barkly West for protein content (A) Principal Component I versus mean protein content (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

The AMMI and GGE biplot results for protein content (planting date 2), as shown in **Figure 5.13**, show the best performing genotype in mega-environment 1 (year 05 and year 10) is SST822. In the second mega-environment (year 06 and year 07, year 08 and year 09) Olifants performed the best. The table of AMMI selections (Appendix C Table C18) confirmed these results

A



B

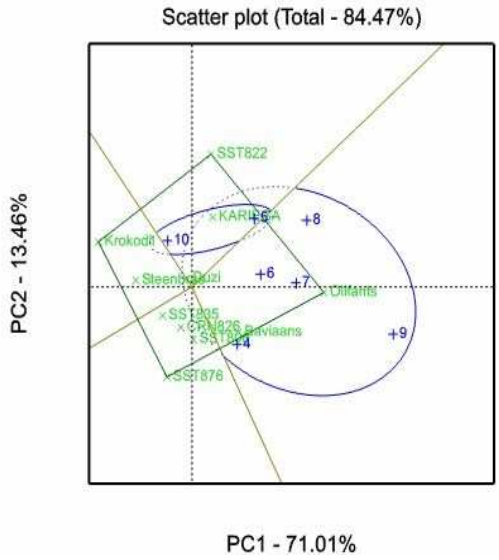


Figure 5.13 Genotype by environment biplot of planting date 2 of Barkly West for protein content (A) Principal Component I versus mean protein content (AMMI biplot) (B) Principal Component I versus Principal Component II (GGE biplot).

The overlapping of the two mega-environments shows that genotype Olifants performed well in all environments.

In conclusion, the GGE biplot graphic analysis complements the environmental stratification of the AMMI biplot by defining mega-environments and showing the genotypes that optimise performance in such mega-environments. Both methods are suitable to explain the GEI, although the AMMI analysis provides statistical support for its biplot results. Utilising both methods simultaneously provides an innovative approach to the interpretation of genotype by environment interactions.

5.4.5 Other statistical techniques

Grain protein is an important factor affecting grain quality and thus a key determinant of both end use and market value in bread wheat. However, it is known that grain protein is negatively correlated with grain yield in wheat. To illustrate the relationship,among the three variables (yield, HLM and protein content) 10 genotypes over 6 years (environments) from a single locality (Barkly West) in the cool region were submitted to these methods.

From the Pearson's product moment correlation matrix (Table 5.10) it was observed that yield correlated significantly positive to HLM and was significantly negative correlated to protein content. The relationship among these correlations and the genotypes is revealed through a principal component analysis (PCA).

Table 5.10 Phenotypic correlation of yield and quality traits for the two planting dates

	Planting date 1		**Planting date 2**	
	Yield	**HLM**	**Yield**	**HLM**
HLM	0.69 **		-0.01	
Protein content	-0.62 **	-0.63 **	-0.23	-0.38

****P≤0.01 HLM = hectolitre mass**

It would have been perfect to investigate the simultaneous relationship of the three variables and genotypes with environments, but the biplots of the PCA were cluttered; thus the results are given on the relationship among the variables and the genotypes.

Table 5.11 Loadings of the variables onto the first two principle components (PC) for the genotypes

	Planting date 1		Planting date 2	
	PC1	**PC2**	**PC1**	**PC2**
Yield	**0.78**	0.08	0.18	**0.75**
HLM	**0.79**	0.05	**0.53**	0.26
Protein content	**0.73**	0.27	**0.73**	0.00

Values in bold correspond for each variable to the principal component (PC) for which the loading is the largest. HLM = hectolitre mass

Table 5.11 presents the correlation (loading) of each variable with the first two principal components (PC) for the two planting dates respectively. For the first planting date all three variables were significantly correlated to PC1 which accounts for 76.63% of the total variation of 89.84%. The relationship among the variables and genotypes for planting date 1 is depicted in **Figure 5.14**. The genotypes SST822, Baviaans, Kariega, Steenbras and SST806 show strong correlation with PC1. Thus these genotypes are strongly associated with the three variables.

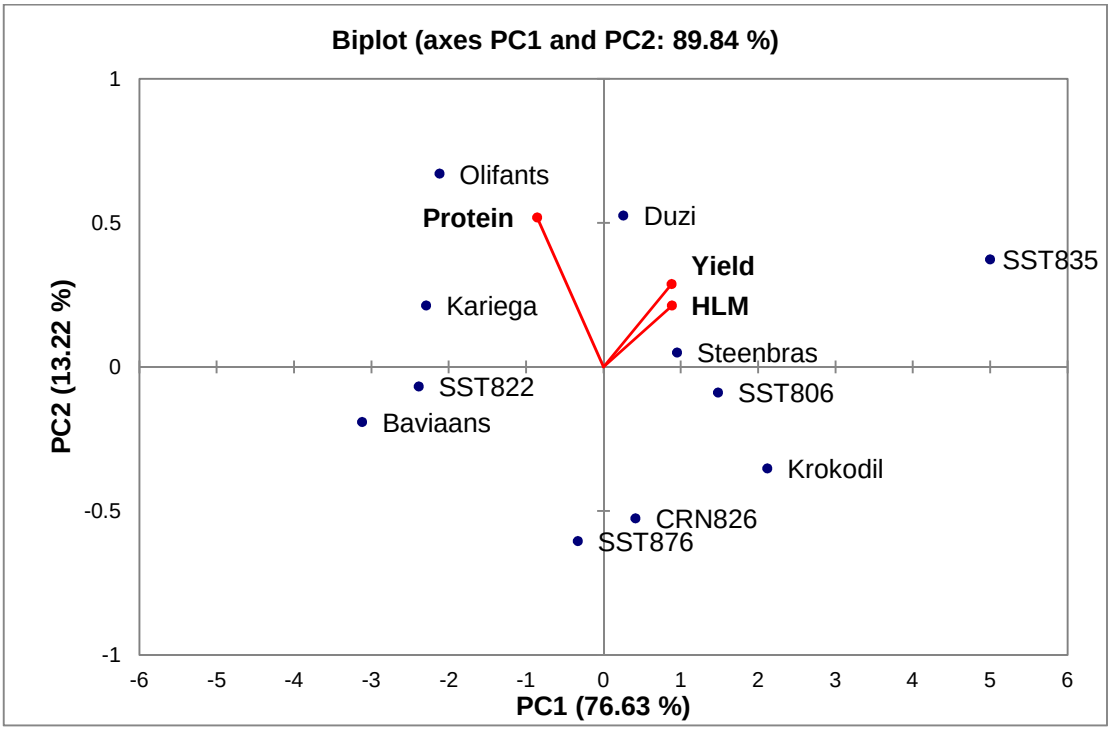


Figure 5.14 PCA biplot of relationship between all variables and the 11 genotypes for planting date 1.

For the second planting date, HLM and protein content were strongly correlated to PC1 which accounts for 47.95% of the total variation of 81.62%. **Table 5.11** shows the strong correlation between PC1 and HLM, and between PC1 and protein content. The correlation between PC2 and yield was the strongest. **Figure 5.15** shows the relationship among the genotypes and variables. The genotypes Baviaans, Krokodil, Olifants and SST806 are strongly correlated to PC1. These genotypes are strongly associated with HLM and protein content. Genotypes Duzi, Kariega, CRN826, SST822, SST876 and SST835 are strongly correlated to PC2. This indicates that yield is strongly associated with these genotypes.

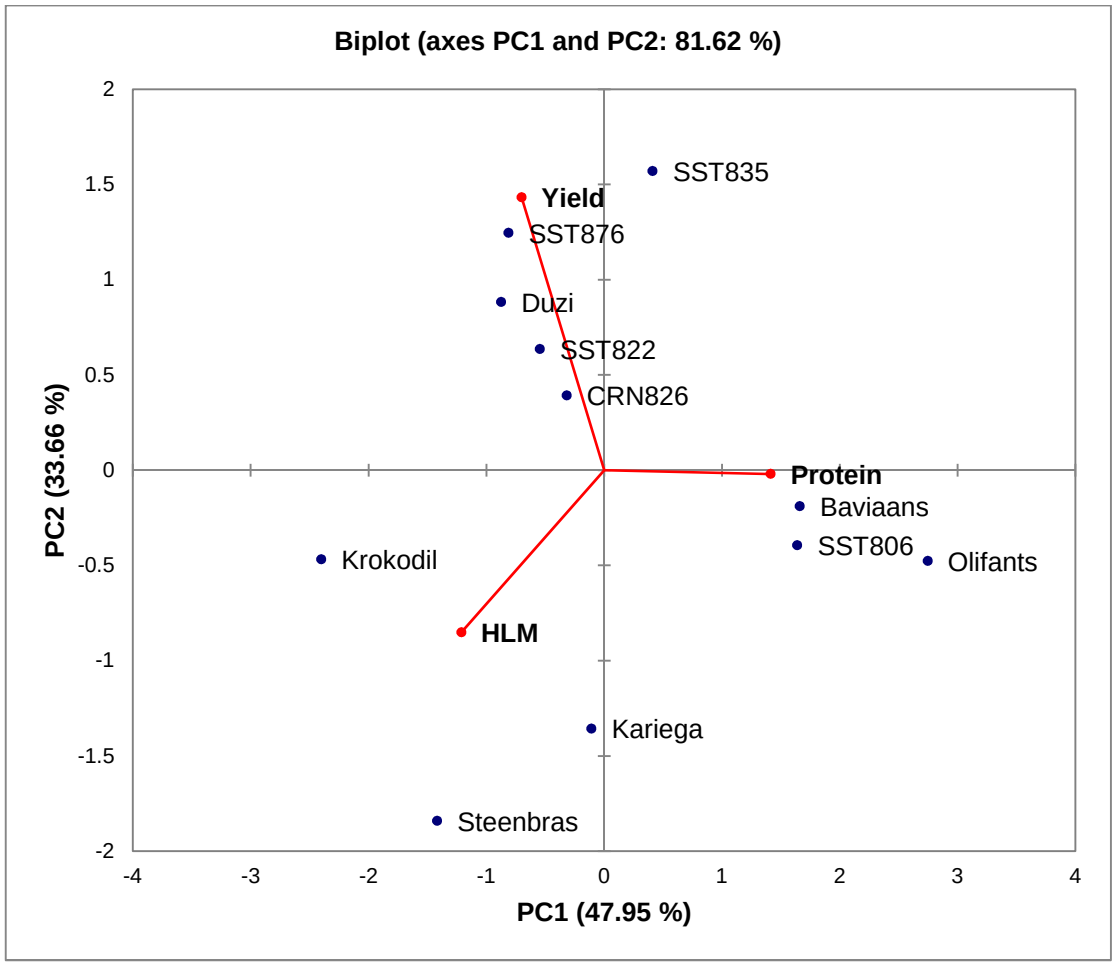


Figure 5.15 PCA biplot of relationship between all variables and the 11 genotypes of planting date 2.

Discriminant analysis (DA) was used to determine whether the variables in this study could provide additional information on genetic advance. The first two discriminant variates (called factors) in **Figure 5.16** and **Figure 5.17** accounted for 96.89 and 83.28% of the variation among the environments respectively. For the first planting date the plot

of the second discriminant variate (factor) versus the first factor is shown in **Figure 5.16**. In this plot the genotypes of an environment are enclosed by a 95% confidence ellipse. Overlapping of different environments' ellipses indicates no differences among environments. This plot indicates two distinct groupings of environments. Year 2009 is grouped on the left side; and years 2004, 2005, 2006, 2007 and 2010 are in the middle of the plot. The scores found for each of the factors were then correlated with the original variables to indicate those that were most important in discriminating between the two groups (results not shown). Protein content had the highest loading (-0.98), yield (0.77) and HLM (0.76) on Factor 1. Only HLM had a moderate loading of 0.62 on Factor 2.

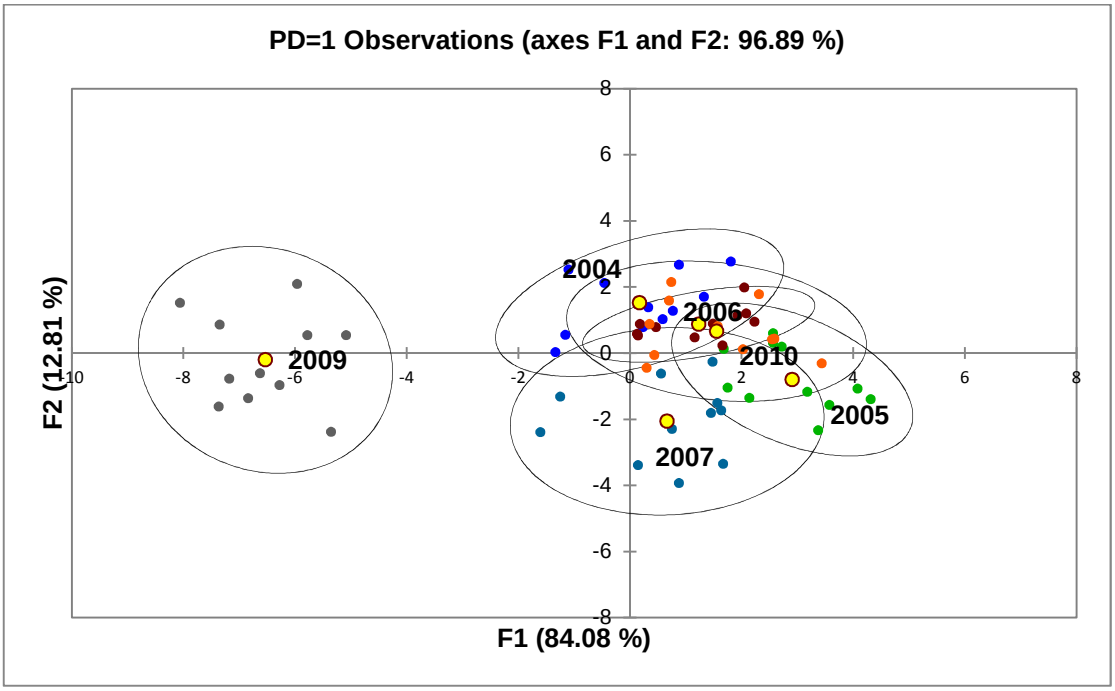


Figure 5.16 Linear discriminant biplot of the environments of Barkly West for planting date 1.

For the second planting date the plot of the second discriminant variate (factor) versus the first factor is shown in **Figure 5.17**. In this plot the confidence ellipses of the environments are overlapping. No distinct groupings are visible. On the most lenient terms years 2004 and 2007 differ from year 2008. Factor 2 is responsible for this separation. Protein content has the highest loading of 0.97 on this factor; thus protein content is most likely responsible for the separation.

Van Lill and Purchase (1995) used discriminant analysis (also called canonical variate analysis) to investigate genetic advance in yield and protein content in winter wheat cultivar data for the period 1930-1994. They found a yield improvement of 1.3% per year and a stable performance in protein content. In this study no genetic improvement could be found with this technique. But it was shown that these variables had discriminating powers for the first planting date.

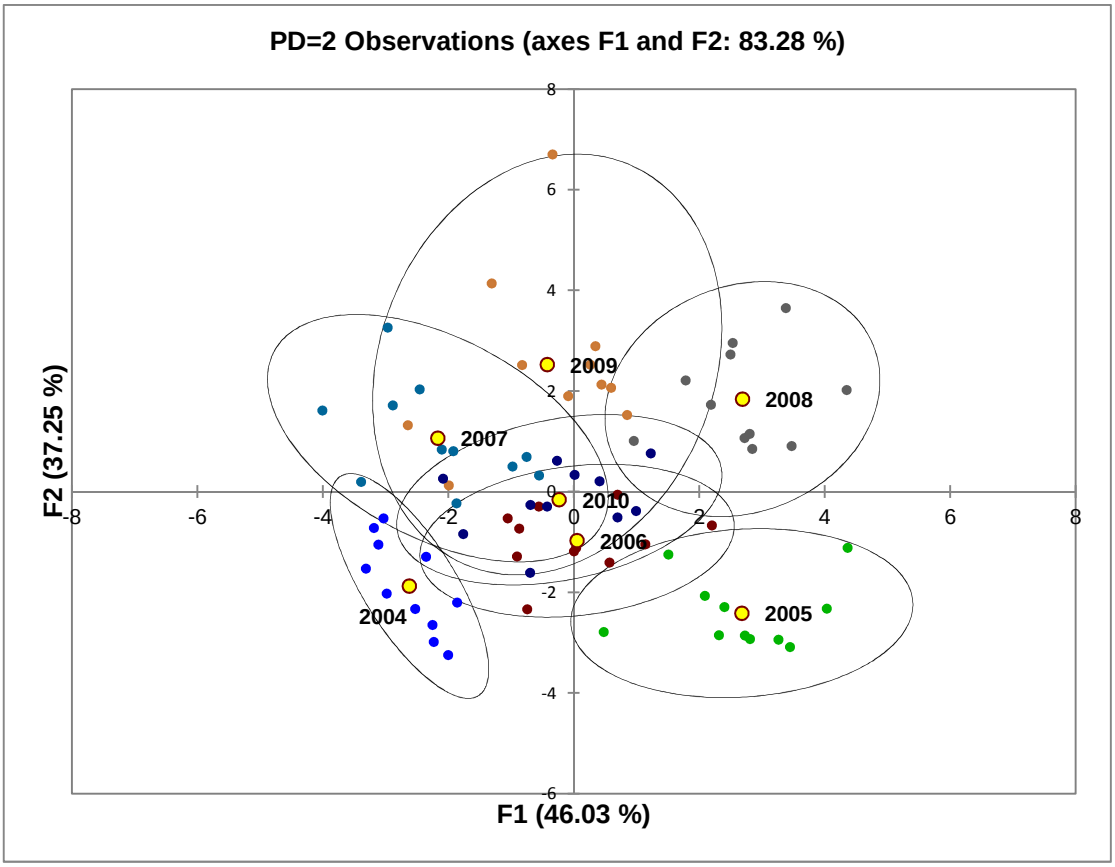


Figure 5.17 Linear discriminant biplot of the environments of Barkly West for planting date 2.

5.5 Conclusions and recommendations

In summary, model 1 to model 6R could be used to calculate genetic advance for yield, HLM and protein content. Although TRET showed no agreement with the other models, it does not mean the linear regression could not be used as a method of determining genetic improvement. The effect of environments (most likely the year effects) is too large to provide significant results. For the regions where TRET provided a significant regression coefficient the progress was about 1% per year. The complementary nature

of the AMMI and GGE models, as shown in the biplots, is an advantage. The PCA and provide additional information. It depends on the objectives of the study which methods should be used for determining genetic advance and modelling GEI. Using multivariate techniques like PCA or DA provides additional information in the holistic picture of genotype selection.

References

Afridi N, Khalil IH (2007) Genetic improvements in yield related traits of wheat. Sarhad Journal of Agriculture 23:965-972

Agricultural Research Council – Small Grain Institute (2010) ARC-SGI report, Bethlehem

Allard RW (1960) Principles of plant breeding. Wiley, New York

Bilgin O, Korkut K, Baser I, Daglioglu O, Ozturk I, Kahraman T, Balkan A (2011) Genetic variation and interrelationship of morpho-physiological traits in durum wheat (*Triticum durum* (L.) Desf.). Pakistan Journal of Botany 43:253-260

Brisson N, Gouache D, Harmet G, Oury FX, Huard F (2010) Why are wheat yield stagnating in Europe? Field Crops Research 119:201-212

Cargnin A, De Souza M, Fronza V (2008) Genetic and environmental contributions to increased wheat yield in Minas Gerais. Crop Breeding and Applied Biotechnology 8:39-46

Comstock RE, Moll RH (1963) Genotype-environment interactions. In Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. National Academy of Science, Washington, pp 164-196

Department of Agriculture (2012) Abstract in agricultural statistics. Pretoria

Falconer DS, MacKay TFC (1996) Introduction to quantitative genetics. Longman, London pp 184-207

Fourie P, Botha P (2011) The role of irrigation in grain production. Senwes. http://www.senwes.co.za/en-ZA/Articles/View/The_role_of_irrigation_in_grain_production.aspx. Accessed 10 November 2013

Gauch HG, Zobel RW (1996) AMMI analysis of yield trials. In Kang MS, Gauch HG (eds) Genotype by environment interaction. Boca Raton CRC, New York, pp 85-122

Giunta F, Motzo R, Pruneddo G (2007) Trends since 1900 in yield potential of Italian-bred durum wheat cultivars. European Journal of Agronomy 27:12-24

Gordon IL, Byth DE, Balam LN (1972) Variance of heritability ratios estimated by phenotypic components. Biometrics 28:401-415

Gregory RJ (2000) Psychological testing: History, principles and applications, 3rd edn. Allyn & Bacon, Boston

Johnson HW, Robinson HF, Comstock RE (1955) Estimates of genetic and environmental variability in soya beans. Agronomy Journal 47:318-324

Kashif M, Khaliq I (2004) Variability and correlation between metric traits in wheat. Journal of Agriculture and Biology 6:138-142

Khan N, Naqvi FN (2011) Heritability of morphological traits in bread wheat advanced lines under irrigated and non-irrigated conditions. Asian Journal of Agricultural Science 3:215-222

Levene H (1960) Robust test in the equality of variance. In Olkin I (ed) Contributions to probability and statistics: Essays in honor of Harold Hotelling. Stanford UP, Stanford, pp 278-292

Mehla RS, Verma JK, Gupta RK, Hobbs PR (2000) Stagnation in the productivity of wheat in the Indo-Gangetic plains: Zero-till-seed-cum-fertilizer drill as an integrated solution. Rice-wheat consortium paper: Series 8

Oury FX, Godin C, Mailliard A, Chassin A, Gardet O, Giraud A, Heumez E, Morlais J, Rolland B, Rousset M, Trottet M, Charmet G (2012) A study of genetic progress due to the selection reveals a negative effect of climate change on bread wheat yield in France. European Journal of Agronomy 40:23-38

Payne R, Murray DA, Harding SA, Baird DB, Soutar D (2012) GenStat for Windows: Introduction, 15th edition. VSN International, Hemel Hempstead

Pask AJD, Reynolds MP (2013) Breeding for yield potential has increased deep soil water extraction capacity in irrigated wheat. Crop Science 53:2091-2104

SAS Institute (2012). The Statistical Procedure manual. North Carolina 27513: SAS Campus Drive, Cary.

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52:591-611

Trethowan RM, Van Ginkel M, Rajaram S (2002) Progress in breeding wheat for yield and adaptation iun global drought affected environments. Crop Science 42:1441-1446

Van Lill D, Purchase JL (1995) Directions in breeding for winter wheat yield in South Africa from 1930 to 1990. Euphytica 82:79-85

XLStat (2012) XLStat., Addinsoft Paris.

Yan W, Kang MS (2003) GGE biplot analysis: A graphical tool for breeders, geneticists and agronomists. Boca Raton CRC, New York

Chapter 6
Conclusions and recommendations

This study was undertaken to compare various statistical methods of analysis to determine the most suitable procedure to evaluate genetic improvement in the three wheat production areas of South Africa, namely the Western Cape province, the Free State province and the irrigation areas. The principal objective of the study was thus to recommend the most appropriate statistical procedure(s) to estimate wheat yield improvements. The second objective was to demonstrate the trend of yield and two quality traits (HLM and protein content) over 16 years (1995-2010) by various statistical techniques. The third objective was to compare the AMMI and the GGE analyses in assessing GEI for yield and the two quality traits. The fourth objective was to study the relationship among wheat grain yield and the two quality traits by various statistical techniques (*inter alia* PCA, cluster analysis and DA).

The three production areas were investigated separately in Chapters 3 to 5 due to their unique climatic and water requirements (rainfed or dryland and irrigated). The similarities among the chapters were the division of the data into elite and cultivar trials within geographical regions and the various statistical techniques performed on the data.

6.1 Outcomes of this study

The primary outcome of the study was the determination of genetic advance (trend) from the elite and cultivar trials of the three production areas by an appropriate statistical method. TRET provided, in most cases, a reliable approximation of genetic advance. Hence the estimated genetic advance as calculated by TRET was as follows:

(i) In the Rûens and Swartland regions of the Western Cape yield increase since 1995 was 1% per year for both regions of the elite trials. No yield increase was observed in the cultivar trials. Genetic advance of 1% per year was observed for protein content for both regions of the cultivar trials.

(ii) No genetic improvement was measured in the elite trials of the three regions of the Free State. Yield improvement of respectively 0.5% and 0.6% per year were found for the two planting dates in the cultivar trials of the eastern region. Yield progress of 0.3% was measured for both planting dates in the central region.

(iii) A genetic advance of nearly 0.7% per year was found in the warm region of the elite trials and 9% (calculated by TRET, by using the formula in Section 2.9.3) in the eastern region for the first planting date of the cultivar trials under irrigation.

The genetic advance as portrayed by the different trials provides a favourable picture for both plant breeders and industry. Most of genetic advance estimates from this study are in line with the world trend of approximately 1% per year. Irrigation trials, as expected, produced much higher yield than dryland trials.

While there is no denial that genetic improvement was made by the breeding process and selection of superior genotypes, caution should be taken how to proceed in future.

The two other objectives, namely the comparison between the AMMI and GGE analyses and the relationship between the factors and the variables are discussed in Section 6.5 to Section 6.8.

The advantages and disadvantages of the statistical techniques investigated in this study are discussed below.

6.2 Linear regression according to Trethowan *et al.* (2002)

The linear regression analysis in this study was called TRET or Ratio [mean of the best five genotypes in a trial as a proportion of the trial mean (TM)] to determine genetic advance. A modification of their technique called Mc (the difference between the mean of the five best genotypes in a trial and the check genotype) was applied. A second modification M_T (difference between the mean of the five best genotypes in a trial and TM) added no value to the study.

Comments: TRET Ratio and/or Mc provided adequate evidence of genetic progress or lack thereof in the trials studied. Ratio was proposed by Trethowan *et al*. (2002) as a method to combat year fluctuations in yield. TRET was performed in Chapter 3 for yield and the quality traits. TRET was significant for only one of the two quality traits. TRET for protein content showed a significant trend in the two regions. No significant trends in Ratio or Mc were found in the quality traits in Chapter 4 and Chapter 5. Therefore it was

not included in these chapters. TRET is recommended for yield trials. It is a straightforward technique which shows reliable results and it combats the yearly deviations very well. Manès *et al*. (2012) proposed a fixed percentage of superior lines should be chosen rather than the mean of the top five genotypes. In this research however a maximum of five genotypes were entered in the breeding trials of a specify year and locality. Thus the top five genotypes were used in the calculation of TRET.

6.3 Variance component methods

The variance component method according to Comstock and Moll (1963) and the method proposed by Allard (1960) were investigated. Model 1 to model 5 depicted the method by Comstock and Moll (1963) and model 6 represented the method by Allard (1960). These analyses were performed as proposed in a linear fixed scenario to determine genetic advance. These methods were also evaluated in this study in the linear mixed model scenario (**Figure 3.2**).

Model 1 (the combined analysis of variance with fixed effects) - based on the results of this study this is not an appropriate model for analysing regional trials. The specification of locality as a fixed effect is true. Year is a spatial effect and should be modelled as such (Smith *et al*., 2001 and Oakley *et al.*, 2007). Genotype in the breeding trials and even in cultivar trials may be non-recurrent and should be seen as random.

Comments: The fixed effects model should render reliable results when the assumptions of the model are fulfilled (namely normality of residuals and homogeneous variances) and when all effects specified in the model are fixed. This is not a straightforward technique and becomes very tedious in calculating genetic advance.

Model 2 - linear mixed model analysis for each year and each locality with genotype effects as fixed and blocks as random effects. Homogeneity of variances among localities within a year resulted in combining localities within a year, hence producing model 3.

Model 3 - linear mixed model analysis for each year with genotype and the genotype by locality interaction specified as random effects. Locality and blocks within locality were defined as fixed effects. The trials were planted on the same localities for the period evaluated in this study. The trials were analysed for each year. Results from a number of

test data sets showed the pooled error (pooled error term over the years) were similar to the error term of model 4.

Model 4 - linear mixed model analysis for the full model without taking heterogeneity of variances into account (**Figure 3.2**). Genotype and genotype interactions with year and locality were specified as random effects. Year, locality, the interaction of year and locality and block within year and locality were defined as fixed.

This model was chosen for analysing the data of this study. This model produced the required variance components to calculate GA estimates. Unfortunately this model produced inflated results. The results could most likely be blamed on ignoring heterogeneity of year variances.

Model 5 - linear mixed model analysis for the full model (**Table 3.2**) with taking heterogeneity of variances into account. Genotype and genotype interactions with year and locality were specified as random effects. Year, locality, the interaction of year and locality and block within year and locality were specified as fixed effects. This model (model 4 plus accounting for heterogeneity of year variances) produced frequent run-time errors. This could probably be due to the limitation of computer capacity - data sets are either too large or too unbalanced for the more parsimonious or elaborate models (both sides of the coin).

Several solutions to the year variance heterogeneity were tried. To compensate for the heterogeneous year variances (or more accurately tried to compensate) the GROUP option in the REPEATED statement of PROC MIXED was added. This was unsuccessful due to computational inadequacies. A second option was tried: fitting the year effect as a spatial effect, e.g. autoregressive (AR), without any success.

Factor analytic (FA) models provide a natural framework for modeling GEI type problems. Mixed model analyses fitting such models are likely to see an increased usage due to the parsimonious description of covariance structures available. The scope for direct interpretation of factors as well as computational advantages are addressed by the FA models (Meyer, 2009) The FA framework has been increasingly used in plant breeding *inter alia* as reported by Meyer (2009), Beeck *et al.* (2010), Cullis *et al*. (2010) and Raman *et al*. (2011). It has become accepted that such structures can be fitted

directly within mixed model approaches commonly applied for estimation of variance components and genetic parameters for selection purposes.

Problems with fitting FA models to the data and the estimation of variance components were encountered in this study. Oakey *et al*. (2007) reported computational difficulties and convergence problems as drawbacks of FA models. Smith *et al.* (2001) attributed these to poor starting values.

Comment: Calculating variance components and thus genetic parameters and GA from the linear mixed models are a cumbersome process. It depends on the objective of the study which model(s) to choose. Therefore these problems should be addressed in future before estimating genetic parameters for the wheat production areas of South Africa.

Model 6 – variance component model proposed by Allard (1960)
In Chapter 3 and Chapter 4 this model was administrated under a linear fixed model scenario. In Chapter 5 GA estimates were calculated from variance components of both linear fixed and mixed models. Both models produced unreliable GA estimates.

Comment: The formula for calculation GA from this model is straightforward but the disadvantages are similar to those discussed for model 1 to model 5.

6.4 Addressing the computational problems in this study

So (2009) suggested that, with the typical statistical complexities of MET data, there are two options available to the analyst:
(i) choose a statistical approach to estimate GA from data with unequal information;
(ii) sub setting the data into balanced subsets;

The first approach (choosing a statistical model) to calculate GA estimates: is discussed in Section 6.2 and Section 6.3.

The second approach - sub setting the data into balanced data sets brought problems of its own. The GA estimates of the sub sets were very different from the GA estimates of the original data set and therefore not shown in the study. From a statistical viewpoint it

is hard to imagine any situation in which information is omitted or ignored. This leads to poorer estimates than using all available information. This option is not recommended.

Comment: The preferred method in literature of estimating genetic parameters in MET is FA models therefore the problems encountered in this study should be addressed in future research.

6.5 AMMI versus GGE analyses

The evaluation of GEI for the full data set used to calculate GA estimates for the data in each chapter was not possible. Although GEI was present in the full data set the AMMI and GGE analyses required a balanced data set. An example (a sub set of the data) was used to perform these techniques.

The features of the AMMI and GGE models entail much more than the two biplots compared in this study as summarised by **Table 6.1**. The objective was to compare the models by evaluating the biplots containing both the genotypes and the environments. By comparing the biplots of the two methods it was found in most cases that the conclusions from the biplots were similar. The AMMI and GGE models showed no differences in the selection of best genotypes in environments.

The strength of the AMMI is the statistical analysis e.g.an analysis with an ANOVA table, mean and ranking values of the genotypes within the environments, a table of the best four genotypes in the environments, various biplots and AMMI Stability Value (ASV) developed by Purchase *et al.* (2000). The strength of the GGE biplot is its graphical presentation of GEI through a variety of biplots according to the objective of a specific study.

Comments: The complementary nature of the models is an advantage. The choice of the model (AMMI or GGE) depends on the goal of a specific study. In performing both models the advantage is to obtain statistical evidence from the AMMI model to validate the conclusions from the various biplots of the GGE analysis.

Figure 6.1 Summary of the features of the AMMI and GGE analyses

	AMMI	GGE Biplot
Prerequisite	Assumptions of ANOVA should be fulfilled. Identifying significant GEI from combined ANOVA	
Definition	Genotype by environment interaction (GEI)	Genotype + Genotype by environment interaction (G+GEI)
Math. definition	$\bar{y}_{ij.} = \mu + \tau_i + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_i$	$\bar{y}_{ij.} = \mu + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_i$
ANOVA	Yes	Yes, in some software packages
Ranking stats	Yes	No, provide a ranking plot
Strong point	Statistics e.g. ranking statistics & ASV	GGE biplot with confidence ellipses

6.6 Principal component analysis (PCA)

PCA is a useful tool in the investigation of the relationship of a numbers of factors (e.g. years, localities and/or genotypes) and a number of variables. PCA was successful in identifying association of certain genotypes with a specific variable in the different chapters. The association in each chapter was unique due to the diverse requirements of the geographical regions which were portrayed in a specific chapter.

Comment: PCA is a very useful multivariate technique in exploring patterns in MET data.

6.7 Cluster analysis

Cluster analysis has a number of uses, e.g. determination of relationship among genotypes or environments. Cluster analysis was used as a test to identify “mega environments” in Chapter 4, but could not identify mega environments except for protein content in the second planting date. Cluster analysis was omitted from Chapter 3 and Chapter 5.

Comment: The advantage of cluster analysis is that it can be used in a univariate (one variable) or multivariate (more than one variable) scenario to determine “relationship” among factor(s) e.g. genotypes, years, and/or environments.

6.8 Discriminant analysis (DA)

DA is sometimes called the t test of a multivariate set-up. Van Lill and Purchase (1995) used this technique to show genetic advance in their data consisting of a number of quality traits. DA did not show genetic advance in this study.

Comment: DA is a suitable multivariate technique to use when the objective of the study is group separation with a number of variables, where groups are: e.g. genotypes, years, and/or environments.

6.9 Other comments

6.9.1 Data transformations

The evaluated data transformation strategies (logarithmic transformation and weighted analysis) were not efficient in eliminating the effects of heterogeneity among years in the estimation of genetic advance in the fixed effects model (model 1).

6.9.2 Coefficient of variation (CV)

There are some requirements that must be met in order for the CV to be interpreted. The most obvious problem arises when the mean of a variable is zero. In this case, the CV cannot be calculated. Even if the mean of a variable is not zero, but the variable contains both positive and negative values and the mean is close to zero, then the CV can be misleading. The CV of a variable or the CV of a prediction model for a variable can be considered as a reasonable measure if the variable contains only positive values. This is a definite disadvantage of CVs.

Comment: In Chapter 3 Phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) were used in interpreting the results. In Chapter 4 and Chapter 5 these statistics were not calculated because CVs could be misleading. Other statistics to describe the variation will be investigated in a future study.

6.9.3 BLUP versus BLUE

There are two general approaches for a mixed model analysis. The first is using REML to estimate variance components and then using BLUP for estimation random effects or

BLUE for estimating fixed effects. In this study the main objective was to determine genetic advance from variance component models. Genotype means of each trial (for each year and each locality) were called in TRET with genotype effect was specified as fixed thus estimating genotype means from a BLUE prospective.

Comment: Means from BLUP should be investigated in future for the trials of the three production areas within the framework of this technique.

6.9.4 Partial Least – Squares model (PLS) and Factorial Regression (FR)

PLS and FR models use explicit information about the environment. This information can be used directly in the model by including it in the form of explanatory variables. GEI is then described as differential genotypic sensitivity to explicit environmental factors such as temperature, precipitation, water availability etc. Reynolds *et al.* (2002) stated that PLS may assist with both conventional and molecular approaches in breeding. Malosetti *et al.* (2013) and Vargas *et al.* (2013) recommended FR for auxiliary information in GEI investigations.

6.10 Recommendations

The recommended method to estimate genetic advance for the three production areas is TRET. An ordinary linear regression over time on TM, Check genotype means or each genotype means could be misleading due to climatic fluctuations. TRET Ratio provided the solution – dividing the means of the five highest yielding genotypes by TM thus compensating for yearly differences. TRET Mc had a similar effect by subtracting the check mean from the mean of the five highest yielding genotypes. Although TRET did not show significant trends in some cases (Chapter 4), it is still recommend as the more reliable method to predict genetic advance in this study.

6.11 Future objectives

The way forward following this study will be to:

(i) address the problems with the linear mixed model, heterogeneity and estimation of variance components;

(ii) utilise BLUE and BLUP in multi-environmental trials;

(iii) find a more reliable method to determine genetic advance in quality traits;

(iv) evaluate more reliable parameters instead of using the coefficient of variation as measure of variability in multi-environmental trials;

(v) investigate FR and PLS in multi-environmental trials.

This may be a challenge for future research to find a solution to these problems.

References

Allard RW (1960) Principles of plant breeding. Wiley, New York

Beeck CP, Cowling WA, Smith AB, Cullis BR (2010). Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. Genome 53:992-1001.

Comstock RE, Moll RH (1963) Genotype-environment interactions. In Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. National Academy of Science, Washington, pp 164-196

Cullis BR, Smith A, Beeck CP, Cowling W (2010) Analysis of yield and oil from a series of canola breeding trials, Part II: Exploiting VxE using factor analysis. Genome 53:1002-1016

Manès Y, Gomez HF, Puhl L, Reynolds M, Braun HJ, Tethrowan R (2012) Genetic yield gains of CIMMYT International Semi-Arid Wheat Yield Trials from 1994-2010. Crop Science 52:1543-1552

Malosetti M, Ribaut JM, Van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modelling genotype-by-environment interaction and its genetic basis. Frontiers in Physiology 4 (article 44):1-17

Meyer K (2009) Factor-analytic models for genotype × environment type problems and structured covariance matrices. Genetics Selection Evolution 41:21-32

Oakey H, Verbyla AP, Cullis BR, Wei X, Pitchford WS (2007) Joint modelling of additive and non-additive (genetic line) effects in multi-environment trials. Theoretical and Applied Genetics 114:1319-1332

Purchase JL, Hatting H, Van Deventer CS (2000) Genotype x environment interaction of winter wheat (*Triticum aestivum* L.) in South Africa II: Stability analysis of performance. South African Journal of Plant and Soil 17:101-107

Raman A, Ladha JK, Kumar V, Sharma S, Piepho HP (2011) Stability analysis of farmer participatory trials for conservation agriculture using mixed models. Field Crops Research 121:450-459.

Reynolds MP, Trethowan R, Crossa J, Vargas M, Sayre KD (2002) Physiological factors associated with genotype by environment interaction in wheat. Field Crops Research 75:139-160

Smith AB, Cullis BR, Thompson R (2001) Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. Biometrics 57:1138-1147

So Y (2009) Prediction of cultivar performance and heterogeneity of genotype variance, correlation and error variance in Iowa crop performance test in corn (*Zea mays* L.). Dissertation, Iowa State University.

Trethowan RM, Van Ginkel M, Rajaram S (2002) Progress in breeding wheat for yield and adaptation in global drought affected environments. Crop Science 42:1441-1446

Van Lill D, Purchase JL (1995) Directions in breeding for winter wheat yield in South Africa from 1930 to 1990. Euphytica 82:79-85

Vargas M, Combs E, Alvarado G, Atlin G, Matthews K, Crossa J (2013) META: A Suite of SAS Programs to Analyze Multi-environment Breeding Trials. Agronomy Journal 105:11-19

Summary

Wheat is the biggest winter cereal crop in South Africa and the second largest cereal to feed the population of South Africa. The population of South Africa grows with approximately one million people a year. Consistent wheat production is necessary for food security and is therefore of extreme agricultural and economic significance. Future production increases depend on the ability to improve, or at least maintain, the rate of increase to feed the population. The study was undertaken to investigate genetic improvement (genetic advance) in wheat by various statistical methods of analysis. This was done to determine the most suitable procedure to evaluate genetic improvement in the three wheat production areas of South Africa, namely the Western Cape province, the Free State province and the irrigation areas. The second objective of this study was to demonstrate the trend of yield and the two quality traits [HLM (hectolitre mass) and protein content] over 16 years (1995-2010) by various statistical techniques. The third objective was to compare the AMMI (additive main effects and multiplicative interaction) and the GGE (genotype plus genotype-by-environment interaction) analyses in assessing genotype-by-environment interaction (GEI) for yield and the two quality traits. The fourth objective was to study the relationship among wheat grain yield and the two quality traits by various statistical techniques. Linear regression (TRET) and various variance component methods were investigated to determine genetic advance. The recommended method of determining genetic advance in this study is TRET. In the Western Cape elite trials TRET predicted a genetic advance of 1% per year and genetic advance estimated at 1% genetic improvement for protein content in the cultivar trials. No significant trend was observed in the elite trials of the Free State with TRET. Yield showed 0.5% and 0.6% per year improvement for the two planting dates of the eastern cultivar trials of Free State. A yield improvement of 0.3% per year improvement for the two planting dates of the central cultivar trials of Free State was determined. A genetic advance for yield of 0.7% per year was found in the warm region of the elite irrigation trials and 9% yield improvement per year for the first planting date of the eastern region of the cultivar irrigation trials. A negative trend was observed for the second planting date of eastern region of both elite and cultivar irrigation trials.

The effects of GEI on yield and quality traits were studied by comparing the AMMI and GGE analyses. These methods portrayed similar results. An advantage of these techniques is their complementary nature. Although both models portray GEI in various biplots, the AMMI provides statistical evidence to the visual presentation of the GGE biplots. Pearson product moment correlation matrix provided a linear relationship among the variables studied. Principal component analysis (PCA), cluster analysis (CA) and discriminant analysis (DA) offered auxiliary information on the relationship among the factors (e.g. genotypes, years, localities and/or environments) and the variables. DA was not able to indicate direction of genetic improvement in either of the three production areas in this study.

Keywords: wheat, genetic advance, genetic improvement, trend, linear regression, variance component methods, genotype-by-environment interaction, AMMI, GGE, cluster analysis, PCA, DA.

Opsomming

Koring is die grootste winter graangewas in Suid-Afrika en die tweede grootste graangewas om die mense van Suid-Afrika te voed. Die bevolking van Suid-Afrika vermeerder met naastenby ‘n miljoen mense per jaar. Volhoubare koringproduksie is noodsaaklik om voedselsekerheid te verskaf en is daarom van uiters hoë landbou en ekonomiese waarde. Toekomstige opbrengsvermeerdering is afhanklik van die verbetering in potensiaal en ten minste instandhouding van die opbrengs om in die kosbehoeftes van die bevolking te voorsien. Hierdie studie is onderneem om die genetiese vordering in koring deur verskeie statistiese metodes te ondersoek. Die drie koringproduksie streke van Suid-Afrika, naamlik die Wes-Kaap provinsie, die Vrystaat provinsie en die besproeiingsgebiede is ondersoek om aan hierdie doelwit te voldoen. Die tweede doelwit van die studie was om die tendens in opbrengs en die twee kwaliteitsveranderlikes, hektolitermassa (HLM) en proteïeninhoud oor 16 jaar (1995-2010) met behulp van verskeie statistiese tegnieke te ondersoek. Die derde doelwit was om vir elkeen van die drie veranderlikes die AMMI (additiewe hoofeffek en multiplikatiewe interaksie) en GGE (genotipe plus genotipe-omgewingswisselwerking) modelle wat genotipe-by-omgewingsinteraksie (GEI) beskryf, te vergelyk. Die vierde doelwit was om die verwantskap tussen die koringopbrengs en die twee kwaliteitseienskappe met behulp van verskeie statistiese tegnieke te beskryf. Metodes soos liniêre regressie (TRET) en verskeie variansiekomponentmetodes is ondersoek in die bepaling van genetiese vordering. Die voorgestelde metode om genetiese vordering te bepaal is TRET. In beide streke in die Wes-Kaap toon TRET genetiese vordering van 1% per jaar vir opbrengs in die elite proewe. In die cultivar proewe van hierdie streke kon TRET geen vordering in opbrengs toon nie, maar wel 1% vordering per jaar vir proteïeninhoud. Geen betekenisvolle tendens kon in die elite proewe van enige van die Vrystaat streke bepaal word nie. In die oostelike streek van die cultivarproewe is daar onderskeidelik 0.5% en 0.6% per jaar genetiese vordering in opbrengs van die twee plantdatums getoon. ‘n Vordering van 0.3% in opbrengs is in die sentrale streek van die Vrystaat bepaal. Met behulp van TRET is genetiese vordering in opbrengs van 0.7% per jaar in die warm gedeeltes van die elite besproeiïngsproewe en 9% vordering per jaar vir die eerste plantdatum van die oostelike cultivar besproeiïngsproewe bepaal. ‘n Negatiewe tendens vir die tweede plantdatum van beide die elite en cultivar proewe van die oostelike besproeiïngsgebied is gevind.

Die effek van genotipe-omgewingswisselwerking vir opbrengs en die kwaliteitseienskappe is bestudeer en vergelyk met behulp van die AMMI en GGE. Hierdie metodes verskaf soortgelyke antwoorde. Die voordeel van hierdie tegnieke is hulle aanvullende aard. Alhoewel beide metodes genotipe-omgewingswisselwerking in verskeie ”biplots” weergee, verskaf die AMMI statistiese bewyse vir die visuele voorstelling van die GGE “biplots”. Pearson produk-moment korrelasiekoeffisiënt verskaf ‘n liniêre verwantskap tussen die veranderlikes. Hoofkomponentanalise (PCA), trosanalise en diskriminantanalise (DA) verskaf bykomende inligting tot die verwantskap tussen die faktore (o.a. genotipes, jare, lokaliteite en/of omgewings) en die veranderlikes. Diskriminantanalise kon nie die rigting van genetiese vordering aandui nie.

Sleutelwoorde: koring, genetiese vordering, tendens, liniêre regressie, variansiekomponent metodes, genotipe-omgewingswisselwerking, AMMI, GGE, PCA, DA, trosanalise.

“I have walked that long road to freedom. I have tried not to falter; I have made missteps along the way. But I have discovered the secret that after climbing a great hill, one only finds that there are many more hills to climb. I have taken a moment here to rest, to steal a view of the glorious vista that surrounds me, to look back on the distance I have come. But I can only rest for a moment, for with freedom come responsibilities, and I dare not linger, for my long walk is not ended.” – Nelson Mandela

Appendices

Appendix A

This appendix corresponds to Chapter 3 (Dryland Western Cape).

Table A1 Sources of variation (mean squares) for yield of the elite trials of the two regions

	Rûens				Swartland				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	14	57.55	7.09	436.64	14	65.56	5.25	340.29	<0.00
Lok	3	156.90	21.71	1144.39	3	17.71	2.49	108.02	<0.00
Lok*Year	21	14.71	2.40	97.97	21	26.20	3.05	120.48	<0.00
Blok(Lok*Year)	111	3.05	0.17	8.71	110	1.49	0.16	7.62	<0.00
GEN	284	2.69	0.40	16.99	282	1.30	0.14	6.77	<0.00
Lok*GEN	475	0.73	0.11	4.20	499	0.32	0.05	1.87	<0.00
Year*GEN	155	0.69	0.08	6.59	154	0.72	0.08	4.02	<0.00
Lok*Year*GEN	188	0.46	0.05	3.17	198	0.40	0.04	2.19	<0.00
Error	3140	0.20	0.02	1.00	3181	0.19	0.02	1.00	
Corrected Total	4391				4465				

Table A2 Sources of variation (mean squares) for yield of the cultivar trials

	Rûens				Swartland				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	8	114.45	17.43	1393.76	9	31.75	1.88	100.37	<0.00
Lok	3	236.51	25.12	2109.55	2	83.71	6.05	331.33	<0.00
Lok*Year	10	24.23	5.53	229.69	12	17.25	1.18	86.02	<0.00
Block(Lok*Year)	66	0.91	0.10	6.52	72	0.77	0.06	4.12	<0.00
Gen	29	1.84	0.17	14.94	30	1.54	0.12	10.21	<0.00
Lok*Gen	58	0.76	0.05	6.15	53	0.63	0.05	3.00	<0.00
Year*Gen	62	0.80	0.07	5.46	69	0.644	0.057	3.471	<0.00
Lok*Year*Gen	67	0.50	0.07	4.55	88	0.53	0.048	2.789	<0.00
Error	651	0.15	0.01	1.00	888	0.26	0.021	1.00	
Corrected Total	955				1223				

Table A3 Sources of variation (mean squares) for HLM of the cultivar trials

	Rûens				Swartland				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	7	54.64	0.01	328.16	8	158.85	0.03	155.20	<0.00
Lok	3	100.51	0.02	407.88	2	621.82	0.11	340.97	<0.00
Lok*Year	6	154.96	0.03	574.02	11	307.73	0.05	240.31	<0.00
Block(Lok*Year)	51	1.19	0.00	2.61	66	4.29	0.00	2.50	<0.00
Gen	29	8.00	0.00	28.67	30	13.89	0.00	10.43	<0.00
Lok*Gen	49	2.77	0.00	14.18	53	3.36	0.00	2.24	<0.00
Year*Gen	47	2.46	0.00	8.25	63	3.29	0.00	2.48	<0.00
Lok*Year*Gen	31	2.28	0.00	9.25	82	2.87	0.00	2.41	<0.00
Error	477	0.32	0.00	1.00	852	2.22	0.00	1.00	
Corrected Total	703	3180.04			1167				

Table A4 Sources of variation (mean squares) for protein content of the cultivar trials

	Rûens				Swartland				
Source	**DF**	**Prot**	**LProt**	**WProt**	**DF**	**Prot**	**LProt**	**WProt**	**P**
Year	6	40.43	0.30	264.58	8	114.82	0.81	0.48	<0.00
Lok	3	142.33	1.06	932.54	2	4.45	0.04	0.04	<0.00
Lok*Year	5	15.73	0.12	125.10	11	69.59	0.48	0.01	<0.00
Block(Lok*Year)	45	0.79	0.01	5.64	66	1.33	0.01	0.04	<0.00
Gen	29	2.27	0.02	12.38	30	5.90	0.04	0.00	<0.00
Lok*Gen	49	0.52	0.00	3.76	53	0.47	0.00	0.00	<0.00
Year*Gen	41	0.37	0.00	2.67	63	0.61	0.00	0.00	<0.00
Lok*Year*Gen	25	0.28	0.00	1.87	82	0.42	0.00	0.00	<0.00
Error	441	0.15	0.00	1.00	852	0.33	0.00	1.00	
Corrected Total	647				1167				

Appendix B

This appendix corresponds to Chapter 4 (Dryland Free State).

Table B1 Sources of variation (mean squares) for yield of the elite trials in the intermediate trial of the eastern region and winter trial of eastern Free State

	Intermediate trial: Eastern Free State				Winter trial: Eastern Free State				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	14	282.53	42.79	3729.36	14	357.15	55.87	4096.55	<0.00
Locality	8	154.26	33.08	1860.17	8	231.56	47.74	2849.77	<0.00
Locality*Year	84	20.95	3.82	251.76	87	29.54	5.85	369.85	<0.00
Block(Locality*Year)	319	0.63	0.22	9.85	330	0.72	0.20	9.70	<0.00
Genotype	264	3.31	0.90	46.44	279	4.15	1.11	49.12	<0.00
Locality*Genotype	1686	0.44	0.13	6.34	1783	0.58	0.17	8.47	<0.00
Year*Genotype	201	1.38	0.33	18.15	233	1.59	0.35	19.42	<0.00
Locality*Year*Genotype	1153	0.33	0.07	4.82	1396	0.42	0.12	5.73	<0.00
Error	9827	0.08	0.02	1.00	11073	0.08	0.02	1.00	
Corrected Total	13556				15203				

Table B2 Sources of variation (mean squares) for yield of the elite winter trials of the central and western Free State

	Winter trial: Central Free State				Winter trial: Western Free State				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	13	25.27	11.35	714.60	13	80.46	12.50	880.37	<0.00
Locality	2	5.34	0.68	121.28	1	44.57	6.69	542.62	<0.00
Locality*Year	5	12.27	2.97	276.69	10	69.89	13.19	1078.06	<0.00
Block(Locality*Year)	63	0.68	0.78	10.77	75	1.00	0.15	11.26	<0.00
Genotype	262	0.37	0.19	8.77	262	1.40	0.25	18.36	<0.00
Locality*Genotype	182	0.24	0.06	5.53	213	0.60	0.11	8.20	<0.00
Year*Genotype	195	0.25	0.16	6.76	196	0.77	0.14	10.72	<0.00
Locality*Year*Genotype	63	0.20	0.05	4.87	161	0.36	0.07	5.24	<0.00
Error	2139	0.05	0.05	1.00	2526	0.09	0.02	1.00	
Corrected Total	2935				3467				

Table B3 Sources of variation (mean squares) for yield of the cultivar trials of the eastern Free State for both planting dates

	Planting Date=1: Eastern Free State				Planting Date=2: Eastern Free State				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	12	214.79	26.94	1914.15	12	170.94	19.18	1579.21	<0.00
Locality	6	163.11	19.71	1344.77	6	95.87	6.55	930.60	<0.00
Locality*Year	52	42.26	5.90	431.30	47	24.99	2.90	245.73	<0.00
Block(Locality*Year)	210	0.60	0.11	6.60	196	0.57	0.06	5.51	<0.00
Genotype	52	5.86	0.88	46.47	51	6.72	0.84	55.35	<0.00
Locality*Genotype	271	0.79	0.18	7.09	269	1.42	0.33	15.73	<0.00
Year*Genotype	199	1.22	0.18	9.85	192	1.68	0.19	14.99	<0.00
Locality*Year*Genotype	730	0.48	0.08	4.12	595	0.64	0.07	5.72	<0.00
Error	3719	0.14	0.02	1.00	3317	0.12	0.01	1.00	
Corrected Total	5252				4686				

Table B4 Sources of variation (mean squares) for yield of the cultivar trials of the central Free State for both planting dates

Planting Date=1: Central Free State					Planting Date=2: Central Free State				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	11	45.41	15.42	1076.51	12	52.16	15.57	1118.75	<0.00
Locality	2	10.92	8.19	182.03	2	12.86	9.36	229.00	<0.00
Locality*Year	7	22.32	7.17	346.30	7	23.59	5.87	366.54	<0.00
Block(Locality*Year)	63	0.40	0.20	4.26	66	0.26	0.06	4.71	<0.00
Genotype	47	1.38	0.44	22.83	48	0.83	0.25	18.04	<0.00
Locality*Genotype	78	0.21	0.08	3.43	72	0.26	0.19	4.22	<0.00
Year*Genotype	167	0.28	0.13	6.91	179	0.39	0.19	9.88	<0.00
Locality*Year*Genotype	85	0.17	0.08	3.22	78	0.21	0.22	4.08	<0.00
Error	1131	0.06	0.02	1.00	1134	0.05	0.01	1.00	
Corrected Total	1591				1599				

Table B5 Sources of variation (mean squares) for yield of the cultivar trials of the north western Free State for both planting dates

Planting Date=1: North-Western Free State					Planting Date=2: North Western Free State				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	10	80.39	8.93	763.19	12	109.96	14.57	944.37	<0.00
Locality	2	261.29	25.63	2016.45	2	142.46	12.59	1293.25	<0.00
Locality*Year	10	7.48	0.69	61.68	10	25.26	1.99	184.17	<0.00
Block(Locality*Year)	69	1.09	0.10	6.19	75	0.57	0.07	4.19	<0.00
Genotype	45	2.64	0.20	16.54	47	2.54	0.32	19.42	<0.00
Locality*Genotype	72	0.80	0.07	5.60	61	0.83	0.07	6.67	<0.00
Year*Genotype	134	0.66	0.06	4.68	171	0.61	0.09	4.85	<0.00
Locality*Year*Genotype	125	0.49	0.05	3.62	126	0.51	0.05	4.24	<0.00
Error	1164	0.16	0.02	1.00	1302	0.14	0.02	1.00	
Corrected Total	1643				1819				

Table B6 Sources of variation (mean squares) for yield of the cultivar trials of the southern Free State for both planting dates

Planting Date=1: Southern Free State					Planting Date=2: Southern Free State				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	11	40.55	12.76	768.10	11	36.22	10.58	647.02	<0.00
Locality	2	43.60	9.21	654.61	2	59.35	11.73	1033.22	<0.00
Locality*Year	11	17.79	4.91	191.45	12	19.81	6.44	420.32	<0.00
Block(Locality*Year)	75	0.25	0.13	6.79	78	0.17	0.07	3.84	<0.00
Genotype	47	1.54	0.76	27.66	45	0.94	0.29	18.03	<0.00
Locality*Genotype	77	0.31	0.28	7.63	78	0.30	0.15	5.62	<0.00
Year*Genotype	171	0.32	0.21	8.31	172	0.31	0.15	6.17	<0.00
Locality*Year*Genotype	145	0.23	0.10	4.78	167	0.23	0.10	3.95	<0.00
Error	1388	0.05	0.02	1.00	1485	0.05	0.02	1.00	
Corrected Total	1927				2051				

Table B7 Sources of variation (mean squares) for HLM of the cultivar trials in the central Free State for both planting dates

Planting Date=1: Central Free State					Planting Date=2: Central Free State				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	11	348.63	0.06	684.18	12	535.39	0.13	435.88	<0.00
Locality	2	149.26	0.02	305.93	2	1080.65	0.03	1929.43	<0.00
Locality*Year	7	91.36	0.01	176.71	7	345.36	0.01	587.79	<0.00
Block(Locality*Year)	63	1.40	0.00	3.64	66	2.09	0.00	3.07	<0.00
Genotype	47	23.34	0.00	58.34	48	62.23	0.01	63.12	<0.00
Locality*Genotype	78	2.11	0.00	4.96	72	263.55	0.00	309.69	<0.00
Year*Genotype	167	5.13	0.00	14.51	179	141.22	0.00	213.99	<0.00
Locality*Year*Genotype	85	2.96	0.00	7.50	78	254.18	0.00	384.90	<0.00
Error	1131	0.52	0.00	1.00	1134	1.70	0.00	1.00	
Corrected Total	1591				1599				

Table B8 Sources of variation (mean squares) for HLM of the cultivar trials of the eastern Free State for both planting dates

Planting Date=1: Eastern Free State					Planting Date=2: Eastern Free State				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	12	666.38	0.11	1582.76	12	1162.86	0.20	1657.00	<0.00
Locality	6	218.09	0.04	409.07	6	915.23	0.06	1005.01	<0.00
Locality*Year	52	136.36	0.02	333.72	48	185.63	0.03	362.17	<0.00
Block(Locality*Year)	210	2.54	0.00	4.06	199	2.41	0.00	4.76	<0.00
Genotype	52	89.39	0.02	202.49	51	216.30	0.02	276.98	<0.00
Locality*Genotype	272	5.43	0.00	12.80	270	169.26	0.00	825.03	<0.00
Year*Genotype	199	12.20	0.00	28.45	192	14.93	0.00	24.76	<0.00
Locality*Year*Genotype	730	3.67	0.00	8.69	609	5.90	0.00	11.08	<0.00
Error	3721	0.49	0.00	1.00	3362	0.71	0.00	1.00	
Corrected Total	5255				4750				

Table B9 Sources of variation (mean squares) for HLM of the cultivar trials of the north western Free State for both planting dates

Planting Date=1: North Western Free State					Planting Date=2: North Western Free State				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	10	349.42	0.06	606.95	12	356.11	0.06	482.15	<0.00
Locality	2	90.62	0.02	139.10	2	312.72	0.05	666.98	<0.00
Locality*Year	10	54.63	0.01	101.24	10	118.75	0.02	377.19	<0.00
Block(Locality*Year)	69	3.20	0.00	5.07	75	2.07	0.00	4.67	<0.00
Genotype	45	28.74	0.00	52.08	47	39.79	0.00	77.33	<0.00
Locality*Genotype	72	4.36	0.00	6.80	61	4.24	0.00	9.52	<0.00
Year*Genotype	134	5.09	0.00	9.01	171	4.90	0.00	11.06	<0.00
Locality*Year*Genotype	125	3.43	0.00	5.88	126	4.07	0.00	9.34	<0.00
Error	1164	0.64	0.00	1.00	1302	0.55	0.00	1.00	
Corrected Total	1643				1819				

Table B10 Sources of variation (mean squares) for HLM of the cultivar trials of the southern Free State for both planting dates

	Planting Date=1: Southern Free State				Planting Date=2: Southern Free State				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	11	330.61	0.06	795.83	11	452.79	0.08	1147.20	<0.00
Locality	2	463.91	0.03	1299.03	2	212.19	0.03	487.29	<0.00
Locality*Year	11	411.83	0.05	1053.53	12	283.69	0.05	479.21	<0.00
Block(Locality*Year)	75	2.15	0.00	5.03	78	2.32	0.00	4.76	<0.00
Genotype	47	119.84	0.00	231.29	45	30.58	0.00	61.21	<0.00
Locality*Genotype	77	106.10	0.00	173.66	78	6.26	0.00	9.15	<0.00
Year*Genotype	169	49.97	0.00	73.56	172	6.29	0.00	12.48	<0.00
Locality*Year*Genotype	145	27.29	0.00	51.48	167	4.36	0.00	7.23	<0.00
Error	1382	0.41	0.00	1.00	1485	0.66	0.00	1.00	
Corrected Total	1919.0				2051				

Table B11 Sources of variation (mean squares) for protein content of the cultivar trials of the central Free State for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**Prot**	**LProt**	**WProt**	**DF**	**Prot**	**LProt**	**WProt**	**P**
Year	10	319.74	2.05	1101.29	11	242.91	1.54	1103.46	<0.00
Locality	2	49.16	0.25	152.48	2	100.69	0.52	478.75	<0.00
Locality*Year	7	69.16	0.43	232.34	7	37.24	0.21	179.31	<0.00
Block(Locality*Year)	60	1.65	0.01	6.56	63	1.07	0.01	5.99	<0.00
Genotype	46	5.42	0.03	17.00	47	4.02	0.02	17.16	<0.00
Locality*Genotype	78	1.26	0.01	3.76	72	0.84	0.01	4.15	<0.00
Year*Genotype	152	1.18	0.01	3.63	164	0.91	0.00	4.30	<0.00
Locality*Year*Genotype	85	0.88	0.01	2.39	78	0.73	0.00	4.27	<0.00
Error	1083	0.33	0.00	1.00	1086	0.22	0.00	1.00	
Corrected Total	1523				1531				

Table B12 Sources of variation (mean squares) for protein content of the cultivar trials of the eastern Free State for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**Prot**	**LProt**	**WProt**	**DF**	**Prot**	**LProt**	**WProt**	**P**
Year	11	672.31	3.88	2481.26	11	541.08	0.55	2127.17	<0.00
Locality	6	159.83	1.10	549.69	6	76.21	3.17	233.94	<0.00
Locality*Year	46	81.21	0.53	376.09	42	73.97	0.51	301.32	<0.00
Block(Locality*Year)	189	2.12	0.01	7.79	178	1.51	0.01	7.11	<0.00
Genotype	51	12.05	0.07	38.64	50	15.10	0.09	49.56	<0.00
Locality*Genotype	253	1.45	0.01	4.71	251	0.97	0.01	4.43	<0.00
Year*Genotype	181	2.13	0.01	7.36	173	1.94	0.01	7.32	<0.00
Locality*Year*Genotype	654	0.73	0.00	2.69	532	0.69	0.00	2.82	<0.00
Error	3379	0.29	0.00	1.00	3017	0.30	0.00		
Corrected Total	4771				4262				

Table B13 Sources of variation (mean squares) for protein content of the cultivar trials of the north western Free State for both planting dates

Planting Date=1: North Western Free State					Planting Date=2: North Western Free State				
Source	**DF**	**Prot**	**LProt**	**WProt**	**DF**	**Prot**	**LProt**	**WProt**	**P**
Year	8	134.41	0.75	602.78	11	179.77	0.95	819.15	<0.00
Locality	2	16.30	0.01	51.89	2	42.00	0.26	135.67	<0.00
Locality*Year	6	27.20	0.15	99.31	7	12.06	0.07	49.44	<0.00
Block(Locality*Year)	51	2.25	0.01	7.34	63	1.86	0.01	6.49	<0.00
Genotype	39	3.04	0.02	8.80	47	4.15	0.02	13.99	<0.00
Locality*Genotype	60	0.85	0.01	2.93	57	0.58	0.00	2.43	<0.00
Year*Genotype	107	1.34	0.01	4.56	155	0.93	0.01	3.83	<0.00
Locality*Year*Genotype	71	0.77	0.00	2.54	81	0.91	0.01	3.54	<0.00
Error	867	0.32	0.00	1.00	1107	0.30	0.00	1.00	
Corrected Total	1223				1543				

Table B14 Sources of variation (mean squares) for protein content of the cultivar trials of the southern Free State for both planting dates

Planting Date=1: Southern Free State					Planting Date=2: Southern Free State				
Source	**DF**	**Prot**	**LProt**	**WProt**	**DF**	**Prot**	**LProt**	**WProt**	**P**
Year	10	173.37	1.328	532.19	9	143.88	1.05	260.66	<0.00
Locality	2	50.18	0.451	121.04	2	67.14	0.57	160.12	<0.00
Locality*Year	10	149.82	1.090	372.04	11	156.14	1.05	525.19	<0.00
Block(Locality*Year)	69	1.78	0.012	6.51	69	2.49	0.02	7.27	<0.00
Genotype	47	4.35	0.028	13.17	42	5.76	0.04	13.83	<0.00
Locality*Genotype	74	1.02	0.007	2.71	76	1.03	0.01	2.70	<0.00
Year*Genotype	155	1.73	0.011	4.96	142	1.36	0.01	3.28	<0.00
Locality*Year*Genotype	132	1.06	0.008	2.88	153	1.30	0.01	3.08	<0.00
Error	1224	0.41	0.003	1.00	1267	0.47	0.00	1.00	
Corrected Total	1723				1772				

Appendix C

This appendix corresponds to Chapter 5 (Irrigation trials).

Table C1 Sources of variation (mean squares) for yield of the elite trials for the cool irrigation region

	Planting Date=1:				Planting Date=2:				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	15	49.10	0.75	92.27	15	84.61	2.04	263.06	<.0.00
Locality	1	241.36	5.12	493.71	1	0.04	0.29	0.10	<.0.00
Locality*Year	13	166.69	2.79	432.20	9	52.58	1.27	125.56	<.0.00
Block(Locality*Year)	60	3.41	0.06	8.04	67	1.49	0.04	4.28	<.0.00
Genotype	344	2.63	0.05	6.13	344	1.83	0.04	4.84	<.0.00
Locality*Genotype	318	1.46	0.03	3.31	244	1.25	0.03	3.50	<.0.00
Year*Genotype	214	1.68	0.04	3.93	214	1.62	0.05	6.74	<.0.00
Locality*Year*Genotype	168	1.81	0.05	3.94	100	1.09	0.03	2.83	<.0.00
Error	2105	0.48	0.01	1.00	1809	0.47	0.01	1.00	
Corrected Total	3238				2804				

Table C2 Sources of variation (mean squares) for HLM of the elite trials of the cool irrigation region

	Planting Date=1:				Planting Date=2:				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	15	205.35	0.03	347.73	15	440.16	0.08	558.11	<0.00
Locality	1	1882.3	0.32	1612.17	1	24.83	0.01	23.52	<0.00
Locality*Year	13	175.96	0.03	189.47	9	405.24	0.07	293.72	<0.00
Block(Locality*Year)	44	8.81	0.00	8.68	57	6.42	0.00	3.90	<0.00
Genotype	344	8.61	0.00	22.61	344	8.00	0.00	9.84	<0.00
Locality*Genotype	318	3.35	0.00	2.21	242	3.12	0.00	2.42	<0.00
Year*Genotype	214	5.76	0.00	23.43	214	2.80	0.00	3.07	<0.00
Locality*Year*Genotype	166	6.38	0.00	2.78	100	3.03	0.00	2.96	<0.00
Error	1546	0.9	0.00	1.00	1459	1.02	0.00	1.00	
Corrected Total	2661				2442				

Table C3 Sources of variation (mean squares) for yield and HLM of the elite trials of the warm irrigation region

Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	14	186.36	20.84	337.57	12	223.23	0.04	308.39	<0..00
Locality	1	165.25	68.35	0.11	1	1445.41	0.26	2189.20	<.0.00
Locality*Year	9	174.03	27.55	378.58	9	154.49	0.03	341.84	<0..00
Block(Locality*Year)	50	2.74	0.06	4.94	34	3.67	0.00	6.39	<0..00
Genotype	334	2.86	0.08	5.25	302	5.90	0.00	14.30	<0..00
Locality*Genotype	237	1.97	0.07	3.20	237	2.44	0.00	5.57	<0..00
Year*Genotype	189	1.93	0.06	2.42	151	2.25	0.00	2.70	<.0.00
Locality*Year*Genotype	108	1.84	0.05	3.46	108	2.17	0.00	2.92	<0..00
Error	1755	0.62	0.01	1.00	1196	1.31	0.00	1.00	
Corrected Total	2698				2051				

Table C4 Sources of variation (mean squares) for yield of the cultivar trials of the cool irrigation region for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	12	354.76	5.57	810.15	12	215.68	4.56	585.58	<.0.00
Locality	9	188.80	3.08	423.79	9	294.20	7.83	703.02	<.0.00
Locality*Year	78	72.87	1.50	286.86	75	63.30	1.69	235.97	<.0.00
Block(Locality*Year)	300	1.18	0.02	2.83	291	1.33	0.03	3.55	<.0.00
Genotype	36	21.00	0.44	43.95	37	22.30	0.55	52.09	<.0.00
Locality*Genotype	241	2.76	0.05	6.04	251	3.34	0.10	8.43	<.0.00
Year*Genotype	115	3.36	0.09	7.66	116	3.13	0.07	8.31	<.0.00
Locality*Year*Genotype	640	1.99	0.04	6.24	3123	2.00	0.01	5.77	<.0.00
Error	3152	0.66	0.01	1.00	618	0.60	0.05	1.00	
Corrected Total	4583				4532				

Table C5 Sources of variation (mean squares) for HLM of the cultivar trials of the cool irrigation region for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	12	5517.43	0.12	652.34	12	4829.64	0.13	656.00	<.0.00
Locality	9	2036.35	0.07	414.58	9	1936.13	0.09	489.10	<0.00
Locality*Year	71	2581.46	0.03	317.27	67	2818.43	0.03	215.62	<0.00
Block(Locality*Year)	276	3.44	0.00	3.83	264	4.20	0.00	4.03	<0.00
Genotype	36	36.50	0.01	41.05	37	68.45	0.01	65.51	<0.00
Locality*Genotype	232	4.45	0.00	6.30	237	12.57	0.00	11.28	<0.00
Year*Genotype	114	5.18	0.00	5.89	115	9.05	0.00	8.56	<0.00
Locality*Year*Genotype	621	4.02	0.00	6.10	584	9.99	0.00	11.29	<0.00
Error	3041	0.98	0.00	1.00	2952	1.11	0.00	1.00	
Corrected Total	4412				4277				

Table C6 Sources of variation (mean squares) for protein content of the cultivar trials of the cool irrigation region for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**Prot**	**LProt**	**WProt**	**DF**	**Prot**	**LProt**	**WProt**	**P**
Year	11	233.11	0.85	612.98	11	204.95	0.55	591.09	<0.00
Locality	9	127.56	0.19	111.94	9	103.45	0.41	391.91	<0.00
Locality*Year	66	137.74	0.21	214.94	64	99.60	0.25	351.30	<0.00
Block(Locality*Year)	261	0.95	0.01	6.40	255	0.68	0.01	4.53	<0.00
Genotype	32	7.32	0.06	42.68	34	23.53	0.09	84.51	<0.00
Locality*Genotype	218	0.67	0.00	10.09	226	15.71	0.03	19.49	<0.00
Year*Genotype	108	0.83	0.01	5.44	109	0.82	0.01	3.87	<0.00
Locality*Year*Genotype	597	0.58	0.00	4.81	567	0.36	0.00	2.73	<0.00
Error	2893	0.19	0.00	1.00	2865	0.25	0.00	1.00	
Corrected Total	4195				4140				

Table C7 Sources of variation (mean squares) for yield of the cultivar trials of the eastern irrigation region for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yild**	**LYld**	**WYld**	**P**
Year	12	60.51	2.33	271.25	12	49.18	1.75	231.10	<0.00
Locality	1	827.07	28.26	3597.07	1	100.83	3.54	787.91	<0.00
Locality*Year	4	22.19	0.51	114.66	4	18.01	0.49	160.58	<0.00
Block(Locality*Year)	54	1.18	0.05	6.03	51	0.97	0.04	4.62	<0.00
Genotype	35	5.07	0.44	27.63	36	2.62	0.11	11.36	<0.00
Locality*Genotype	23	1.95	0.07	8.13	16	1.00	0.05	4.36	<0.00
Year*Genotype	111	1.39	0.06	6.25	110	1.25	0.05	5.32	<0.00
Locality*Year*Genotype	51	0.92	0.04	4.07	21	1.07	0.04	4.69	<0.00
Error	660	0.25	0.01	1.00	557	0.24	0.01	1.00	
Corrected Total	951				808				

Table C8 Sources of variation (mean squares) for HLM of the cultivar trials of the eastern irrigation region for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	12	309.47	454.55	491.41	12	130.97	0.02	179.40	<0.00
Locality	1	800.35	524.28	1354.12	1	43.57	0.01	109.99	<0.00
Locality*Year	4	41.41	26.73	73.09	4	13.87	0.00	38.25	<0.00
Block(Locality*Year)	54	2.83	3.71	3.98	51	3.64	0.00	4.40	<0.00
Genotype	35	42.37	17.40	90.98	36	16.75	0.00	17.32	<0.00
Locality*Genotype	23	3.32	2.37	5.65	16	6.41	0.00	7.25	<0.00
Year*Genotype	111	3.63	4.43	5.82	110	5.17	0.00	7.03	<0.00
Locality*Year*Genotype	51	2.94	1.83	4.81	21	3.55	0.00	4.40	<0.00
Error	660	0.59	1.00	1.00	557	0.86	0.00	1.00	
Corrected Total	951				808				

Table C9 Sources of variation (mean squares) for protein content of the cultivar trials of the eastern irrigation region for both planting dates.

	Planting Date=1				Planting Date=2				
Source	**DF**	**Prot**	**LProt**	**WProt**	**DF**	**Prot**	**LProt**	**WProt**	**P**
Year	11	83.99	0.47	454.55	12	157.48	1.25	638.79	<0.00
Locality	1	190.62	1.12	524.28	1	1.69	0.02	30.34	<0.00
Locality*Year	4	16.15	0.10	26.73	4	33.64	0.18	128.36	<0.00
Block(Locality*Year)	51	0.99	0.01	3.71	51	1.14	0.01	3.87	<0.00
Genotype	33	5.79	0.03	17.40	36	3.02	0.02	14.49	<0.00
Locality*Genotype	23	0.84	0.00	2.37	16	1.48	0.01	3.99	<0.00
Year*Genotype	106	1.17	0.01	4.43	110	1.27	0.01	4.75	<0.00
Locality*Year*Genotype	51	0.58	0.00	1.83	21	0.67	0.00	2.19	<0.00
Error	639	0.40	0.00	1.00	557	0.32	0.00	1.00	
Corrected Total	919				808				

Table C10 Sources of variation (mean squares) for yield of the cultivar trials of the warm irrigation region for both planting dates.

	Planting Date=1				Planting Date=2				
Source	**DF**	**Yld**	**LYld**	**WYld**	**DF**	**Yld**	**LYld**	**WYld**	**P**
Year	12	46.25	1.32	159.16	9	48.88	1.59	236.47	<0.00
Locality	3	219.19	6.4	587.44	3	89.01	3.77	409.73	<0.00
Locality*Year	16	38.76	1.27	128.22	6	53.81	1.97	234.7	<0.00
Block(Locality*Year)	96	1.17	0.04	3.64	57	1.87	0.06	5.37	<0.00
Genotype	38	4.66	0.15	14.7	38	2.22	0.08	7.83	<0.00
Locality*Genotype	58	1	0.03	3.08	48	0.88	0.03	3.77	<0.00
Year*Genotype	114	1.18	0.04	3.92	74	0.99	0.03	3.74	<0.00
Locality*Year*Genotype	142	1	0.04	3.48	46	0.85	0.03	3.59	<0.00
Error	1055	0.32	0.01	1.00	617	0.29	0.01	1.00	
Corrected Total	1534				898				

Table C11 Sources of variation (mean squares) for HLM of the cultivar trials of the warm irrigation region for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**HLM**	**LHLM**	**WHLM**	**DF**	**HLM**	**LHLM**	**WHLM**	**P**
Year	12	292.13	0.05	547.05	9	597.17	0.10	520.13	<0.00
Locality	3	393.34	0.07	238.72	3	1053.17	0.19	732.15	<0.00
Locality*Year	16	289.45	0.05	333.14	6	254.47	0.04	237.64	<0.00
Block(Locality*Year)	96	4.85	0.00	4.29	57	6.97	0.00	5.04	<0.00
Genotype	38	10.35	0.00	15.17	38	11.15	0.00	11.86	<0.00
Locality*Genotype	58	3.82	0.00	4.15	48	5.34	0.00	5.69	<0.00
Year*Genotype	114	2.54	0.00	3.62	74	3.46	0.00	3.74	<0.00
Locality*Year*Genotype	142	3.79	0.00	4.55	46	3.08	0.00	2.30	<0.00
Error	1055	0.93	0.00	1.00	617	1.17	0.00	1.00	
Corrected Total	1534				898				

Table C12 Sources of variation (mean squares) for protein content of the cultivar trials of the warm irrigation region for both planting dates

	Planting Date=1				Planting Date=2				
Source	**DF**	**Prot**	**LProt**	**WProt**	**DF**	**Prot**	**LProt**	**WProt**	**P**
Year	11	45.65	0.33	194.39	8	55.98	0.39	291.33	<0.00
Locality	3	33.23	0.25	128.84	3	4.57	0.04	22.32	<0.00
Locality*Year	14	27.8	0.21	102.65	5	27.99	0.22	133.08	<0.00
Block(Locality*Year)	87	1.77	0.01	7.3	51	2.31	0.02	6.45	<0.00
Genotype	35	4.94	0.03	16.62	35	2.52	0.02	8.93	<0.00
Locality*Genotype	50	1.23	0.01	4.73	44	0.72	0	3.67	<0.00
Year*Genotype	108	0.75	0.01	2.37	68	0.59	0	2.23	<0.00
Locality*Year*Genotype	132	0.57	0	2.49	41	0.41	0	1.89	<0.00
Error	975	0.28	0	1.00	563	0.29	0	1.00	
Corrected Total	1415				818				

The AMMI selections (genotypes) per environment are given in Table C13 to Table C18

Table C13 Table of first four AMMI selections per environment for yield for planting date 1

Number	Environment	Mean	Score	1	2	3	4
2	05	7.657	0.9226	KARIEGA	Duzi	SST835	Krokodil
1	04	6.922	0.699	SST835	SST806	Olifants	CRN826
3	06	6.868	0.0944	SST835	Krokodil	Duzi	SST806
5	09	4.872	-0.0795	SST835	SST806	Krokodil	SST876
6	10	8.399	-0.3877	SST835	Krokodil	SST806	SST876
4	07	7.097	-1.2488	Baviaans	SST835	SST876	SST806

Table C14 Table of first four AMMI selections per environment for yield for planting date 2

Number	Environment	Mean	Score	1	2	3	4
2	05	8.46	0.8601	SST876	SST835	Duzi	SST822
5	08	7.735	0.6861	Duzi	SST876	SST806	CRN826
4	07	4.721	0.5444	SST806	Duzi	Krokodil	CRN826
6	09	4.59	0.0769	Duzi	SST876	SST835	CRN826
7	10	5.127	-0.3532	Duzi	CRN826	Krokodil	SST806
1	04	5.144	-0.508	Duzi	CRN826	Krokodil	Steenbras
3	06	5.646	-1.3062	SST822	Steenbras	SST835	SST876

Table C15 Table of first four AMMI selections per environment for HLM for planting date 1

Number	Environment	Mean	Score	1	2	3	4
1	4	80.11	1.2754	SST835	Steenbras	SST806	SST822
4	7	77.68	0.7365	SST835	SST806	Steenbras	CRN826
3	6	79.97	0.1574	SST835	CRN826	SST806	SST876
6	10	80.41	-0.4813	CRN826	KARIEGA	SST806	SST876
2	5	79.56	-0.7676	Baviaans	SST835	KARIEGA	Steenbras
5	9	75.87	-0.9204	Baviaans	Steenbras	SST835	KARIEGA

Table C16 Table of first four AMMI selections per environment for HLM for planting date 2

Number	Environment	Mean	Score	1	2	3	4
1	4	73.84	1.6284	Steenbras	Duzi	KARIEGA	Krokodil
3	6	77.26	0.825	SST822	Olifants	Krokodil	KARIEGA
2	5	79.2	0.4848	Steenbras	KARIEGA	Krokodil	SST822
5	8	75.88	0.2365	KARIEGA	Krokodil	SST876	SST822
4	7	72.29	-0.9615	SST835	SST876	SST806	Duzi
7	10	76.72	-1.0795	SST835	SST876	SST806	CRN826
6	9	74.36	-1.1337	SST835	SST876	SST806	CRN826

Table C17 Table of first four AMMI selections per environment for protein content for planting date 1

Number	Environment	Mean	Score	1	2	3	4
2	5	10.38	0.3991	Baviaans	Duzi	KARIEGA	Olifants
6	10	11.32	0.3743	Duzi	Baviaans	KARIEGA	Olifants
5	9	14.23	0.3296	Olifants	KARIEGA	SST806	Baviaans
1	4	11.84	0.124	Olifants	KARIEGA	SST822	Baviaans
3	6	11.26	-0.0413	Olifants	KARIEGA	SST822	Duzi
4	7	11.1	-1.1857	SST822	KARIEGA	SST835	Krokodil

Table C18 Table of first four AMMI selections per environment for protein content for planting date 2

Number	Environment	Mean	Score	1	2	3	4
7	10	12.63	0.9419	SST835	SST822	Krokodil	SST806
1	4	11.58	0.1897	SST835	Olifants	SST806	CRN826
2	5	11.78	0.1806	SST822	Olifants	KARIEGA	Baviaans
3	6	12.31	0.0648	Olifants	SST822	Baviaans	KARIEGA
5	8	13.48	-0.1816	Olifants	SST822	KARIEGA	Baviaans
4	7	12.82	-0.2103	Olifants	Baviaans	SST822	KARIEGA
6	9	13.7	-0.9851	Olifants	Baviaans	SST806	KARIEGA