The nature and causes of sugarcane genotype x environment interactions: Integrated approaches to analysis and interpretation

Sanesh Ramburan

Submitted in fulfilment of the requirements for the degree

Philosophiae Doctor in Plant Breeding

in the Faculty of Natural and Agricultural Sciences, Department of Plant Sciences, University of the Free State, Bloemfontein

April 2012

Promoter:

Prof. Maryke Labuschagne

Co-promoter:

Dr. Marvellous Zhou

TABLE OF CONTENTS

DECLARATION

I hereby declare that the information contained in the following dissertation is the result of my own research efforts, unless otherwise stated. I further cede copyright of the thesis in favour of the University of the Free State.

Signed……………………………………..

Sanesh Ramburan

ACKNOWLEDGEMENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF OUTPUTS FROM STUDY

Ramburan, S., 2010. Analysis and interpretation of sugarcane cultivar adaptability to time of harvest in South Africa. *Proceedings of the American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America*. Long Beach, California, USA.

Ramburan, S., 2011. Sugarcane cultivar x time of harvest interactions in South Africa. *South African Journal of Plant and Soil* 28, 75-84.

Ramburan, S., Zhou, M., 2011. Investigating sugarcane genotype x environment interactions under rainfed conditions in South Africa using variance components and biplot analysis. *Proceedings of the South African Sugar Technologists Association* 84, 345-358.

Ramburan, S., 2011. Interpreting sugarcane varietal adaptability to time of harvest. *Proceedings of the South African Sugar Technologists Association* 84, 375-388.

Ramburan, S., Zhou, M., Labuschagne, M., 2011. Interpretation of genotype x environment interactions of sugarcane: Identifying significant environmental factors. *Field Crops Research* 124, 392-399.

Ramburan, S., Zhou, M., Labuschagne, M., 2012. Investigating test site similarity, trait relations, and causes of genotype x environment interactions in the Midlands region of South Africa. *Field Crops Research* 129, 71-80.

Ramburan, S., Zhou, M., Labuschagne, M.T. 2012. Integrating empirical and analytical approaches to investigate sugarcane genotype x environment interactions. *Crop Science* 52, 2153-2165.

CHAPTER 1

GENERAL INTRODUCTION

Genotype x Environment (G x E) interaction is commonly observed as changes in the performance of cultivars in different environments. It has been a constraint to breeding and selection efforts for decades, as superior cultivar performance in one environment is not necessarily repeated in others. Consequently, breeders and agronomists evaluate cultivar performance across many environments in what is commonly referred to as multi-environment trials (METs). A MET often (not always) consists of the same set of cultivars planted in the same year at different locations (Gauch, 1992). The purpose of MET networks is to evaluate and select promising cultivars for commercial production in a target region. Resource constraints in most MET networks prevent trial siting under all possible production scenarios characterising a target region. It is therefore imperative that selection occurs at sites that are representative of the target region. Additionally, sites within a network should ideally maximize G x E interactions so that valuable resources are not wasted on sites that produce similar genotypic responses. Therefore, the choice of test sites within a MET network is critical to deliver well adapted, high yielding cultivars. Consequently, it is essential that breeders are aware of the nature of G x E interactions as well as the extent of test site similarity within a MET network. This awareness and subsequent corrective action is conventionally achieved through various statistical analyses that have been developed over time. Furthermore, G x E interactions are a result of differential responses of genotypes to variations in biotic and abiotic factors. The identification of such factors is essential (yet an uncommon practice) in determining breeding objectives and optimizing the structure of a MET network.

The sustainable production of sugarcane (*Saccharum* spp.) in South Africa is highly dependent on the continual release of adapted sugarcane cultivars by the South African Sugarcane Research Institute (SASRI). Breeding and selection efforts at SASRI commenced following a period of importation, quarantine, and testing of cultivars from other countries prior to 1945. Crossing and selection programs were well established by 1965, with selection stations operating in the irrigated north and in the southern rainfed regions. After many years of changing selection farms, a total of five selection farms were established for the rainfed

regions by 1998 (Nuss, 1998). In choosing the selection farms, the broad characterisation of the rainfed parts of the industry into the coastal (low altitude), midlands (high altitude) and hinterland (intermediate altitude) regions was taken into consideration, in order to implement region-specific selection (Nuss, 1998). The coastal, hinterland and midlands regions are characterised by typical crop harvest ages of 12, 15-18, and 20-24 months, respectively. Within each region, the selection farms were chosen based on expert knowledge of soil and climatic conditions and variable cropping scenarios. Each selection farm has a selection program (which comprises "on-station" and "off-station" selection sites) linked to it. Figure 1.1 shows the different regions making up the South African sugar industry and the locations of the five selection farms (also aligned with selection programs). Since the establishment of the selection programs there have been limited studies aimed at evaluating and optimizing the efficiencies of the trial networks to improve cultivar delivery to the industry. Most studies that were conducted either focused on the irrigated northern regions (Parfitt, 2000), or were very limited in scope and objectives relative to the entire MET network (Redshaw et al., 2002). At inception of this study, information on the magnitude and significance of G x E interactions; the appropriateness of the broad regional subdivisions for cultivar selection and evaluation; the similarities between test sites within selection programs; the variation associated with different components of G x E interactions; and the performance and stability of released cultivars for the rainfed regions, was severely lacking. Such information is essential to improve the structure of the trial networks and evaluate the efficiencies of current approaches.

In addition to the general lack of information on G x E interactions in the different rainfed regions prior to this study, there was also no information on the overall causes of G x E interactions. The rainfed regions of the industry are characterized by large variations in soil and climatic conditions. Such variability has resulted in the development of different crop management strategies derived from research outputs and grower experience. The effects of specific soil and climatic factors on sugarcane growth and yields in the industry are extensive and well documented. Factors such as crop nutrition, soil type, pests and diseases, seasonal weather, cultivar, and management practices (age at harvest, time of harvest, etc.) have an impact on sugarcane production, and the relative importance of each is unknown. From a sugarcane improvement perspective, however, it is necessary to gain an understanding of the relative importance of these production factors to the G x E interactions. Such understanding will allow for development of selection programs that are specifically designed to

accommodate variations in these important factors, leading to delivery of more adapted cultivars. The relative effects of soil, climate, and management factors on G x E interactions in the rainfed regions of the industry have never been investigated.
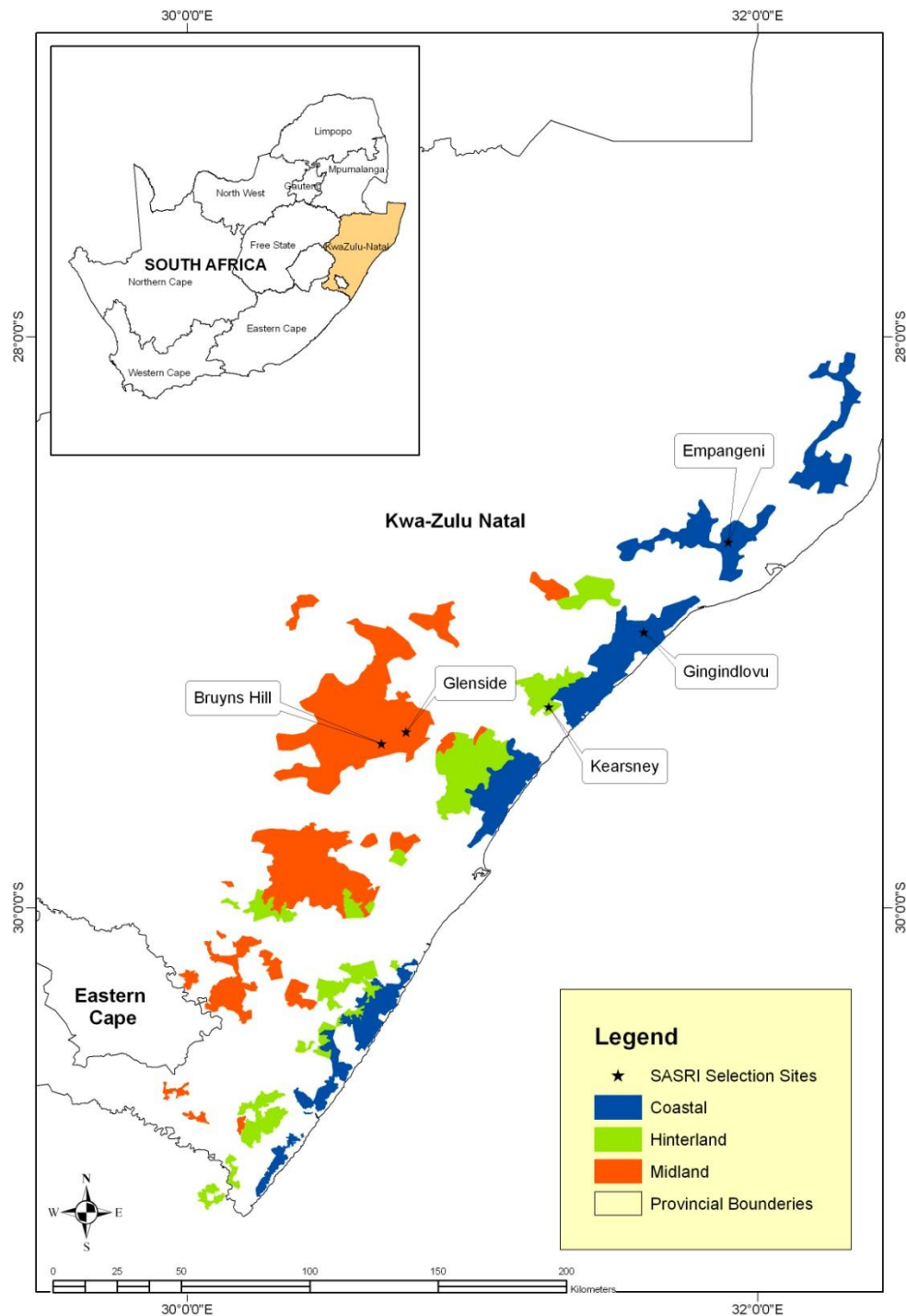
Figure 1.1 The rainfed regions of the South African sugar industry with SASRI selection stations indicated

Multi-environment trials are key tools in any cultivar improvement/evaluation program, where they have contributed to breeding advances and recommendations for decades. Research institutes, government departments, and universities invest massively in sugarcane METs throughout the world annually. The full value of such investments is often not realized because of the empirical approach to data analysis that is often adopted with METs. This is largely due to the lack of appropriate soil, climatic and management data associated with METs, which, if available, can allow for more comprehensive analytical interpretations of MET data. The analysis of soil and climatic data associated with sugarcane METs is limited to a single study (Jackson et al., 1995) and studies documenting possible approaches and methodologies to accomplish these analyses have not been conducted. In particular, the recent advances in crop modelling, and improved quality and accessibility of weather data present opportunities for more extensive characterisations of METs. For example, summarizing climatic data within specific growth phases defined by crop models may assist in determining growth phase sensitivities to various factors. A large-scale characterisation of sugarcane METs across diverse conditions by integrating crop models, climatic data, and soil information has not been attempted previously. Methodologies and approaches to derive and incorporate associated data into sugarcane MET datasets will be illustrated throughout the chapters of this thesis.

The statistical methods of analysing sugarcane G x E interactions have included conventional analysis of variance, regression analysis, additive main effects and multiplicative interaction, and variance components analyses (Crossa, 1990; Kang and Miller, 1984). Newer methods such as GGE (genotype + genotype x environment) biplot analysis have only been evaluated with sugarcane in a limited number of studies (Queme et al., 2007; Glaz and Kang, 2008). Additionally, other statistical methods such as pattern analysis (joint use of ordination and clustering), factorial regression, and other multivariate approaches, which are more suited to analytical/interpretive studies of G x E interactions, have not been used with sugarcane MET data. For example, methods that involve the grouping of sugarcane genotypes based on agronomic traits, and the evaluation of group performance in different environments, may assist in developing sugarcane trait selection strategies. Therefore, in addition to evaluating how associated climatic, soil, and management data can be incorporated into sugarcane MET data analyses, this study will also evaluate the appropriateness of various statistical methods for the analysis and interpretation of sugarcane G x E interactions.

The broad goal of this study was to systematically analyse, identify causes, and explore more comprehensive methods of analysing the G x E interactions of sugarcane, to optimize future MET networks.

The primary objectives of this study were:

1) To investigate the nature of G x E interactions, the components of variation, genotype performance and stability, and the mega-environment constitution of the rainfed regions using GGE biplot and variance components analysis.
2) To investigate the relative influence of soil, climatic, and management factors on the G x E interactions across rainfed regions of the industry.
3) To investigate test site similarities, causes of G x E interactions and opportunities for differential trait selection strategies in the midlands region of the industry.
4) To determine if sugarcane G x E interactions can be more comprehensively explained using environmental covariates summarized within growth phases.
5) To illustrate the integration of soil and climate data with statistical and crop models to comprehensively interpret sugarcane G x E interactions in the coastal region of the industry.

The above objectives were specifically related to chapters in the study. The secondary objectives, which are common themes across all chapters, were:

1) To develop and illustrate methods to enhance the value of sugarcane METs for future G x E studies through use of integrated methods of characterising sugarcane field trials.
2) To evaluate the appropriateness of multivariate statistical analysis techniques for the interpretation of sugarcane G x E interactions.

1.1 References

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

Gauch, H.G., 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam.

Glaz, B., Kang, M.S., 2008. Location contributions determined via GGE Biplot analysis of multi-environment sugarcane genotype-performance trials. *Crop Science* 48, 941-950.

Jackson, P., McRae, T., Hogarth, M., 1995. Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. *Field Crops Research* 43, 119-130.

Kang, M.S., Miller, J.D. 1984. Genotype x environment interactions for cane and sugar yield and their implications in sugarcane breeding. *Crop Science* 24, 435-440.

Nuss, K.J., 1998. Aspects considered in the search for new farms for the experiment station. Proceedings of the South African Sugar Technologists Association 72, 42-45.

Parfitt, R.C., 2000. Genotype by environment interaction among secondary variety trials in the Northern region of the South African sugar industry. *Proceedings of the South African Sugar Technologists Association* 74, 245-248.

Queme, J.L., Crossa, J., Orozco, H., Melgar, M., 2007. Analysis of genotype-by-environment interaction for sugarcane using the sites regression model (SREG). *Proceedings of the International Society of Sugar Cane Technologists* 26, 764-769.

Redshaw, K.A., Govender, N., Smit, M., 2002. Investigating two statistical techniques used in the analysis of sugarcane variety trials. *Proceedings of the South African Sugar Technologists Association* 76, 219-223.

CHAPTER 2

LITERATURE REVIEW

2.1 Sugarcane and the South African industry

Sugarcane is a giant member of the grass family (*Poaceae*), and is cultivated primarily for the extraction of sucrose (sugar) from the plant stalks. More modern end-uses of this tropical crop include the utilization of plant biomass for co-generation (production of energy from combustion of fibre to operate sugar factories) and the production of ethanol from sugar or lignocellulose (Gomez et al., 2008). In South Africa, the crop is cultivated exclusively for sugar, to satisfy both local consumption and export demand. South Africa ranks approximately ninth as the world's largest sugar producing country, with countries like Brazil, India, and China dominating the world markets (Gopinathan, 2010). Sugarcane production in South Africa predominantly occurs along the east coast, extending from approximately 25°33'S to 30°93'S and between 29°92'E and 32°32'E, under a diverse range of conditions. With production occurring the furthest south of the equator for the crop, sugarcane in South Africa is grown in environments that are occasionally not typically conducive to a tropical crop. Nevertheless, the South African sugar industry is still a cost competitive producer of approximately 2.2 million tons of sugar per annum from an estimated 430 000 ha under cultivation (Meyer, 2007). The industry comprises 13 Mill supply areas (MSA), which are each characterized by a single mill owned by a milling company. Each sugar mill receives sugarcane from surrounding commercial and small-scale farms that are located in close proximity (approximately 50 km radius on average). Mill supply areas vary from being fairly homogenous to extremely diverse production regions, depending on factors such as altitude and soil type. Sugarcane milling proceeds for a 9-month period from April to December each year, when sucrose content of the sugarcane stalks is highest.

The crop is vegetatively propagated through the planting of cane setts (segments of the stalk) consisting of approximately three to five viable buds into furrows drawn alongside each other and ranging in spacing from 1 to 1.5 m. In South Africa, planting usually occurs in autumn or spring (preferred due to better soil moisture conditions). Germination, tillering, and stalk elongation rates are highly dependent on genotypic and environmental factors (Smit et al.,

2004). The crop is harvested when sucrose accumulation within the stalks reaches a peak, and the time to maturity also varies depending on genotype and growing conditions. In South Africa, sugarcane is harvested any time between 12 and 24 months of age, depending on temperatures (influenced by altitude). In the northern production area (where sugarcane is grown under irrigated conditions) and along the coastal belt, harvesting generally occurs at 12-months of age. In the hinterland and midlands regions, harvest age ranges from 15 to 24 months.

In most sugarcane industries, including South Africa, harvesting is carried out manually, and involves the cutting of stalks at the base of the stools. As an aid to the harvesting process, the leaf material of the standing crop is usually burnt prior to harvest. Numerous studies have investigated the potential drawbacks of burning compared to the environmentally friendlier practice of cutting "green" cane and returning organic matter to the soil; a practice termed "trashing" (Van Antwerpen and Meyer, 1998). Once stalks are harvested, buds below the ground are released from apical dominance and subsequently germinate to produce a new crop. This regrowth is termed ratooning. Successive ratoons are characterized by reductions in cane yield (ton cane/ha) due to systemic diseases or physical damage to stools, and the number of ratoons obtained from a single harvest also depends on genotypic and environmental factors.

Sugarcane growers are remunerated for sugarcane deliveries to the mill through the implementation of the Recoverable Value (RV) payment system, which takes into consideration quality characteristics of the sugarcane. The RV formula is:

$$RV\% = S - d*N - c*F, \text{ where} \tag{1}$$

S = sucrose % in cane

N = non sucrose % in cane

F = fibre % in cane

d = coefficient to account for losses of sucrose through molasses during processing

c = coefficient to account for losses of sucrose from bagasse during processing

The coefficients d and c are based on milling statistics from the previous three seasons, and therefore vary considerably from season to season. Growers are remunerated based on the

total tons RV (cane yield in tons/ha x RV%) delivered to the mill per season, adjusted for within season variability in quality.

2.2 Sugarcane breeding and evaluation

The successful production of sugarcane in South Africa is attributed to the continuous supply of locally developed, high yielding sugarcane cultivars by SASRI. Modern sugarcane cultivars are complex hybrids of *S. officinarum*, *S. barberi*, *S. sinense* and *S. spontaneum*. Such hybrids are produced in South Africa through the artificial stimulation of flowering in a controlled environment. Natural flowering only occurs under specific climatic conditions and is more widespread in most tropical areas. In South Africa, natural flowering proceeds without the development of viable pollen thereby necessitating artificial interventions (Brett, 1946). Parents are selected from local or imported germplasm according to specific breeding objectives and crosses are made annually. Glasshouse-germinated seedlings are then distributed to one of seven core selection stations located throughout the industry. Five of these selection stations are situated in the rainfed regions (Figure 1.1). Each station represents the general conditions within a region and is characterized by a unique selection program lasting between 12 and 15 years from seedlings to release. The selection stages comprise evaluations at the single plant level through to fully replicated cultivar trials in later stages. Inter-program and off-station evaluation occurs in the later selection stages only (Parfitt, 2005). Selection priorities in South Africa include RV yielding ability and pest and disease resistance. Similar selection procedures and breeding objectives are implemented in other sugar industries as well. There are currently 56 commercial cultivars available for cultivation in the South African industry, however; only about 40% of those cultivars contribute significantly to production (Ramburan et al., 2010a).

Although breeding and selection procedures are similar to other crop breeding programs, the South African industry is unique in its post-release evaluation procedures. Whereas most sugar industries utilize production databases and rely on individual estate evaluations (Ramburan and Van den Berg, 2011), the South African industry has its own post-release cultivar evaluation project, which is coordinated by SASRI. This is a continuous project aimed at increasing the amount of information on released cultivars to strengthen cultivar recommendations. During selection, the range of conditions accross which cultivars are tested are limited and cultivars are released primarily based on their responses to soil and climatic conditions experienced on selection stations. The focus of post-release evaluation is to

identify specific adaptations of released cultivars with respect to environmental and management factors. Post-release evaluation in the sugar industry was initiated in the 1970s, and continues to form an integral part of SASRI's core functions. Although invaluable to the industry in direct recommendations and technology transfer, the scientific applications of such procedures are often overlooked. Post-release evaluation essentially comprises METs that allow for studies of G x E interactions in a similar way to pre-release testing.

2.2 G x E interaction

The term genotype refers to the full complement of genes inherited by an individual that is important for the expression of a trait under investigation, and it is a fixed character that remains constant and unchanged by environmental effects throughout the individual's life. The phenotype refers to the morphological and physiological characteristic of an individual, which changes continually depending on the interaction of the genotype with the environment. The environment refers to the sum total of the effects of physical, chemical and biological factors on an individual other than its genotype (Yan and Kang, 2003).

Crop cultivars are released for commercial production based on their ability to produce high yields (of food, feed, fibre, or fuel) and other essential agronomic characteristics. The yield performance of cultivars is under the control of genetic and environmental influences, and selection attempts to exploit the genetic basis of that phenotype so that released cultivars can continually produce high yields. However, due to the quantitative nature of the trait (controlled by many genes), genes vary in their contribution to yield as environmental conditions change. This introduces a degree of uncertainty when evaluating genotype performance in specific environments, as the actual contribution due to genotype may be influenced (either positively or negatively) by environmental conditions. Consequently, there is uncertainty of the repeatability of genotypic performance in different environments. This is the basis of G x E interaction, which has been a constraint to improvements from selection for decades. Therefore, in its simplest form G x E interaction may be described as changes in the relative performance of genotypes over environments. According to Cooper and Byth (1996), G x E interaction occurs in every aspect of biological science, and as a result, any scientific inference made from research is conditional because of the existence of G x E.

Yan and Kang (2003) described the different types of G x E interactions and highlighted the implications of these in plant breeding and crop production (Figure 2.1). Crossover

interactions (change in rankings of varieties across environments) are of greatest interest to breeders as these directly affect genotype selection in specific environments. Consequently, promising selections in one environment may perform poorly in another. Such crossover interactions often compel breeders to implement multiple selection programs within industries based on the homogeneity of regions, thereby utilizing greater resources. Ignoring significant G x E in favour of resource savings can have detrimental effects. Inaccurate characterization of genotype adaptability may lead to poor productivity in environments that interact negatively with specific genotypes and this has implications on industry sustainability. With regards to genetic gains from selection, large G x E interactions, as components of total phenotypic variance, affect heritability (proportion of total phenotypic variance that is due to genetic variance) negatively. The larger the G x E interaction component, the smaller the heritability estimate; thus, progress from selection would be reduced as well (Yan and Kang, 2003).

The methods employed for the analysis of G x E interactions can be classified into two major groups depending on the nature of the data available and the objectives of the analysis. The classical analysis of G x E interactions involves exploiting yield-based data and evaluating genotypic performance across trials. Alternatively, it is often desirable to describe the reaction of genotypes to environments relative to the biophysical variables that directly affect crop yield i.e. to interpret G x E interactions. Voltas et al. (2005) refer to these approaches as empirical or analytical strategies of G x E analysis.
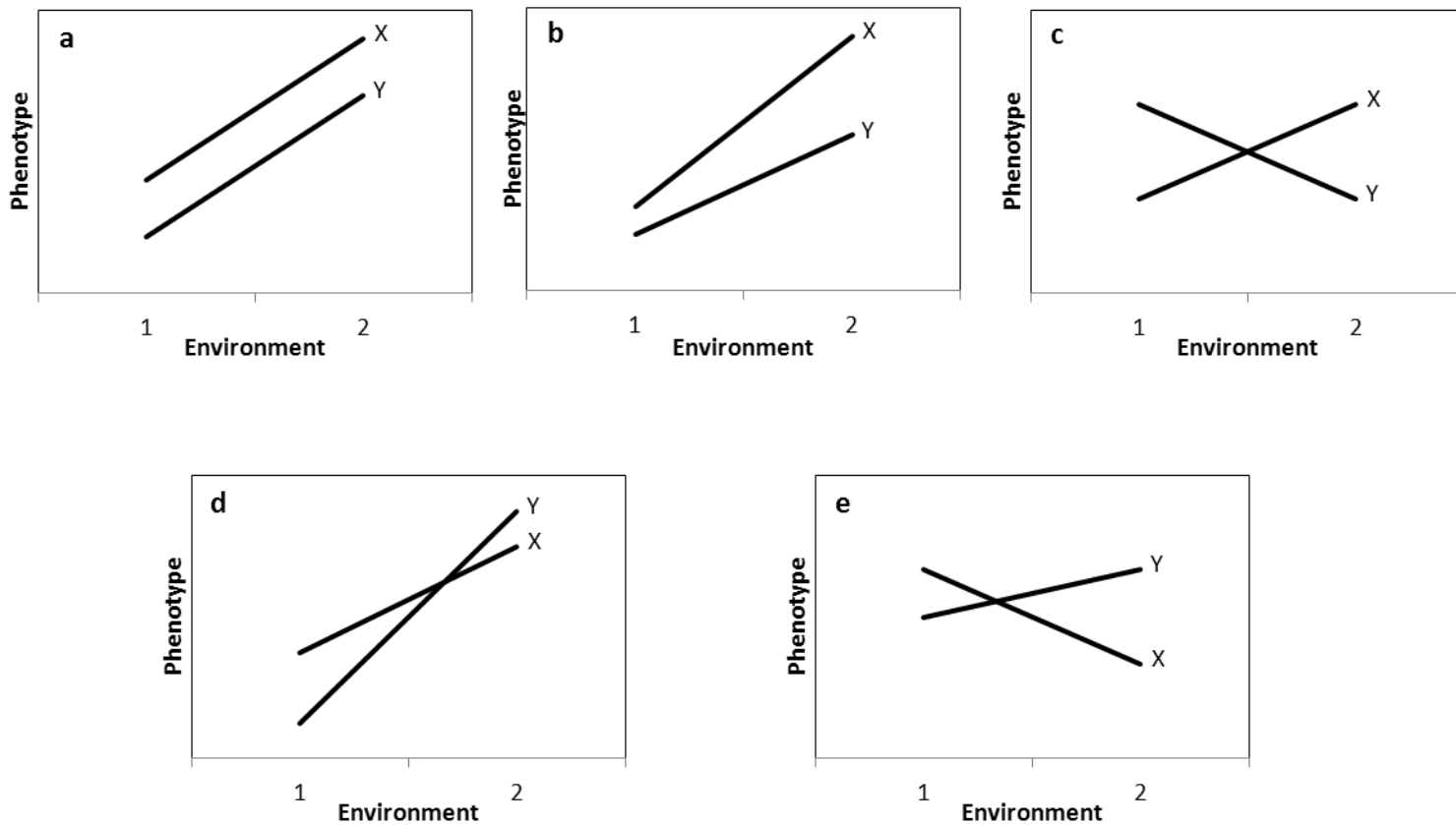
Figure 2.1 Graphical representation of types of G x E interaction: (a) no interaction – X and Y responses parallel in the two environments; (b) non-crossover type interaction – both X and Y increase but unequal inter-genotypic differences in the two environments; (c) crossover interaction – genotypic modification by environment in opposite direction but inter-genotypic difference remains the same; (d) crossover interaction – unequal inter-genotypic difference but both X and Y increase; (e) crossover interaction – unequal inter-genotypic difference in the two environments: X shows a decrease whereas Y shows an increase in environment 2. - adapted from Yan and Kang (2003)

2.3.1 Classical analysis of G x E interactions (Empirical approach)

Genotype x environment interaction has been a focus of plant breeders as early as the 1950's, and there is a wide range of literature outlining examples and methods of dealing with this phenomenon. More recently, statistical advances have given rise to numerous techniques of analyzing yield results from METs, which are the fundamental tools by which G x E interactions are evaluated in most crop industries. According to Gauch (1992), a MET consists of the same set of genotypes planted in the same year at different locations. The primary objective of METs is to identify superior genotypes for a target region, and to

determine if the target region can be subdivided into different mega-environments (Yan et al., 2000). A mega-environment may be defined as a portion of a crop species' growing region with a homogeneous environment that causes some genotypes to perform similarly (Gauch and Zobel, 1997), and is normally identified through analysis of MET data. Currently, there is a wide range of statistical techniques used for the analysis of yield data collected from METs. The most common and contemporary techniques are briefly described below.

2.3.1.1 The analysis of variance (ANOVA)

If one considers a trial in which the yield of G genotypes is measured in E environments, each with R replicates, then the classical model for analyzing the total yield variation contained in the GER observations is the analysis of variance (Fisher, 1925). After removing the replicate effects, the GE observations are partitioned into the additive main effects for genotypes and environments, and the non-additive effects due to the G x E interaction (Crossa, 1990). The ANOVA model of the combined data is then expressed as

$$Y_{ij} = \mu + G_i + E_j + GE_{ij} + e_{ij} \qquad\qquad (2)$$

where Y_{ij} is the expected yield of the ith genotype in the jth environment; μ is the grand mean; G_i, E_j, and GE_{ij} represent the effects of the genotype, environment, and the G x E interaction respectively; and e_{ij} is the error term. In this model, the non-additivity interaction implies that the expected value of the ith genotype in the jth environment depends not only on the levels of G and E separately, but also on the way in which G interacts with E.

The ANOVA model has been applied to MET datasets to estimate components of variation associated with genotypes, environments, and G x E interaction. Knowledge of the size of variance components has been used in MET analyses to obtain estimates of genotype effects, to determine the optimum allocation of resources (number of plots and locations), and to estimate the heritability and predicted gain of a trait under selection (Crossa, 1990). Despite the above uses of the ANOVA model in the analysis of MET datasets, there are certain limitations. One of these limitations is that it does not explore underlying structures within the G x E interaction and does not provide a pattern of response of genotypes and environment.

2.3.1.2 Joint linear regression

The joint linear regression method is a popular technique used to determine the yield stability of genotypes evaluated in METs. The concept of stability has been described in many different ways, and there is a range of other stability statistics that have been utilized over the years (Yan and Kang, 2003). In a broad sense, stability may be referred to as consistency of genotype performance and minimum variation among environments. The linear regression approach, which was popularized by Finlay and Wilkinson (1963), uses the marginal means of the environments as independent variables regressed against genotype yields. In this method, the environmental mean acts as a surrogate for the cumulative effects of soil and climatic factors, and genotype responses to these factors can be interpreted from the slope of the regression curves. The model partitions the G x E interaction into a component due to linear regression (b_i) and a component due to deviations from regression (d_{ij}) so that equation (2) becomes

$$Y_{ij} = \mu + G_i + E_j + (b_iE_j + d_{ij}) + e_{ij} \tag{3}$$

Despite being one of the more popular techniques used in G x E analysis, the linear regression method has limitations. Firstly, the genotype mean is not independent of the marginal means of the environments, and this violates one of the assumptions of regression analysis. Secondly, the analysis assumes a linear relationship, and when violated, the effectiveness of the analysis is reduced. Thirdly, the results of the analysis can be misleading when the data set includes results from a few extremely low or high yielding environments or genotypes (Crossa, 1990). Nevertheless, the method is still a valuable technique that can be used to describe the structures and patterns of G x E interactions as it provides an easily interpretable measure of yield stability (i.e. the slope of the regression line).

Multi-environment trial data vary in their complexity, and the methodology chosen for their analysis will depend on the characteristics intrinsic to specific datasets (DeLacy et al., 1996). In some instances where two or more techniques are compared, the outputs of the analysis are found to be more dependent on the quality of the data rather than on the technique used. Multivariate techniques such as pattern analysis, the additive main effects and multiplicative interaction (AMMI), and GGE (genotype + G x E) biplot analysis have gained popularity due to their ability to produce biplots that allow for rapid visualization of patterns of G x E interactions.

2.3.1.3 Multivariate methods

The main purposes of multivariate analysis are to distinguish systematic from non-systematic variation, to summarize the data, and to reveal a structure within the data (Crossa, 1990). These techniques are appropriate for the analysis of two-way data matrices such as G x E data. Two groups of multivariate techniques have been used to investigate the internal structure of G x E interactions. These include ordination techniques such as principal components analysis (PCA) and classification techniques such as cluster analysis. Pattern analysis involves the joint use of cluster analysis and ordination techniques to explain G x E interactions (McLaren, 1996).

Principal components analysis is a frequently used multivariate method that can be applied to reduce the complexity of a two-way G x E data matrix (many dimensions) into a subspace of fewer dimensions. It is derived from a mathematical theorem called singular value decomposition (SVD). If one considers G points in an E-dimensional space having E axes, then provided that some correlations exist among the variables, the cloud of points would actually have most of its structure in a subspace of fewer than E dimensions. Principal components analysis then defines new coordinate axes that go in the major directions of the cloud and projects the points from the original high-dimensional space into a low-dimensional subspace (Gauch, 1992). The first PCA axis is normally placed in the direction of the cloud that minimizes the sum of squared perpendicular projections (sums of squares) from data points onto the axis. The next PCA axis may be defined perpendicular to the first, thereby accounting for the remaining variation, and this continues for as many PCA axes as desired. Most often, however, only two axes are kept due to the ease of interpretation when illustrated in biplots. The sums of squares (SS) of a PCA axis is termed its "eigenvalue", and its direction relative to the original axis system is termed its "eigenvector". The projection of genotype and environment data points onto the axes then determines their coordinates in a two-dimensional biplot. The PCA model may be written as:

$$Y_{ij} = \mu + \sum \lambda_k \, \alpha_{ik} \, \delta_{jk} \tag{4}$$

where Y_{ij} is the value of the ith genotype in the jth environment; μ is the grand mean; λ_k is the singular value for PC axis k; α_{ik} and δ_{jk} are the PC scores for axis k of the ith genotype and j^{th} environment, respectively.

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 organized into a hierarchy where individuals or groups are fused one at a time to individuals or groups with the most similar patterns accross all environments. In non-hierarchical systems the individuals are organized into a set number of groups in the best possible manner (DeLacy 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 dendrograms that depict composition of groups and the degree of dissimilarity among groups. With respect to analysis of MET data, genotypes or environments that cluster together are expected to have similar contributions to the G x E interaction.

2.3.1.3.1 AMMI

The AMMI method combines the traditional ANOVA and PCA into a single analysis with both additive and multiplicative parameters (Gauch, 1992). The first part of AMMI uses the normal ANOVA procedures to estimate the genotype and environment main effects. The second part involves the PCA of the interaction residuals (residuals after main effects are removed). The AMMI model equation is:

$$Y_{ij} = \mu + G_i + E_j + \sum\lambda_k \, \alpha_{ik} \, \delta_{jk} + R_{ij} + \varepsilon \tag{5}$$

where Y_{ij} is the value of the ith genotype in the jth environment; μ is the grand mean; G_i is the deviation of the ith genotype from the grand mean; E_j is the deviation of the jth environment from the grand mean; λ_k is the singular value for PC axis k; α_{ik} and δ_{jk} are the PC scores for axis k of the ith genotype and j^{th} environment, respectively; R_{ij} is the residual and ε is the error term (Gauch, 1992). The AMMI produces reliable estimates of genotype and environment performance and summarizes the relationships between these components graphically into biplots. The most commonly used biplot is the AMMI1, which plots the interaction principal component (IPCA1) scores against the genotype and environment means. This allows visualization of genotype and environment main effects, the stability of genotypes, the relative adaptability of genotypes to different environments, and the mega-

environments represented by the data. The AMMI has also been shown to improve the accuracy of yield estimates, which was equivalent to increasing the number of replicates by a factor of two to five (Crossa, 1990). The AMMI continues to be a commonly used method for analyzing METs in a range of crop industries presently.

Crossa et al. (1991) showed that AMMI could be used to separate wheat (*Triticum aestivum*) selection environments into homogenous subsets, characterize 18 genotypes based on yield stability, and give more precise estimates of genotype yields than unadjusted means in an analysis of a CIMMYT international MET dataset. Gauch and Zobel (1997) demonstrated that AMMI could be used to separate four mega-environments in a Louisiana maize (*Zea maize*) MET. Sivapalan et al. (2000) used AMMI to identify five groups of wheat genotypes and four groups of Australian environments that discriminated those genotypes similarly and also identified genotypes that could be used as indicators in broad and specific environments. Abamu et al. (1998) utilized AMMI to understand G x E interactions for rice (*Oryza sativa*) reactions to blast (*Pyricularia oryzae*) disease. Ibanez et al. (2001) showed that AMMI explained twice as much of the G x E interaction compared with linear regression analysis in a study involving lovegrass (*Erogrostis curvula*). More recently, Gauch (2006) highlighted the strengths and weaknesses of AMMI and reiterated its superiority as a tool for the analysis of METs.

2.3.1.3.2 GGE biplot

The GGE (genotype + G x E) biplot method (Yan et al., 2000) also makes use of PCA, and differs from AMMI based on how the two-way table of G x E means are treated before performing SVD. The AMMI applies SVD to the data minus the genotype and environment means, while GGE biplot applies SVD to the data minus the environment means only (Gauch, 2006). As a result, conventional AMMI biplots describe only G x E effects, while GGE biplots describe genotype and G x E effects. This is based on the concept that effects of environments are usually large, however, these effects are not relevant to cultivar evaluation and focus should therefore be on the genotype and G x E effects only. The GGE biplot model is:

$$Y_{ij} - E_j = \sum_k \lambda_k \, \alpha_{ik} \, \delta_{jk} + R_{ij} + \varepsilon \qquad\qquad (6)$$

Since the introduction of GGE biplot and the associated user-friendly software (Yan, 2001), there have been numerous applications of the method to MET analyses as well as the analysis

of other types of data with two-way structures. Yan et al. (2000) used the GGE biplot technique to show that winter wheat production environments in Canada should be grouped into two mega-environments, as opposed to a traditional grouping of 4 sub-areas. Yan and Rajcan (2002) employed the GGE biplot technique to soybean (*Glycine max*) MET data and identified a single mega-environment with frequent crossover G x E interactions. In the same study it was demonstrated that GGE biplots could be utilized successfully to investigate genotype x trait data to reveal interrelationships among soybean traits and compare genotypes on the basis of multiple traits. Dehghani et al. (2006) used GGE biplot to identify three barley (*Hordeum vulgare*) mega-environments in Iran. Yan and Kang (2003) demonstrated the application of the GGE biplot technique for the analysis of trait x quantitative trait loci interactions in barley. Sharma et al. (2010) used GGE biplot to determine the performance, stability, and superiority of winter wheat breeding lines in irrigated environments in central and west Asia. The recent popularity of GGE biplot is linked to its versatility and ability to analyse a range of data types with a two-way structure. More recently, however, Yang et al. (2009) reported that GGE biplot is more of a descriptive tool as the statistical significance of any differences between genotypes and environments on the biplot is not indicated, and users should consequently proceed with caution when using the methodology.

2.3.2 Interpretation of G x E interactions (Analytical approach)

Crop growth and potential yield are essentially defined by the availability of resources (light, water, nutrients) and the efficiency of resource use, and the extent to which the attainment of such potential is limited by biological and physical hazards. In this sense, G x E interaction may be explained as the differential use of resources by different genotypes in different environments, as well as differential genotype escape from/tolerance of, environmental hazards in different environments (Bidinger et al., 1996). It therefore follows that an understanding of the causes (resources or hazards) of G x E interactions would allow for more directed approaches to understand the phenomenon. For example, the identification of a predominant limiting factor within a target population of environments in a MET could lead to the development of more efficient breeding strategies. Also, characterization of environments according to a range of environmental factors offers the potential to improve gains from selection through better choice of environments for field trials. In a study involving factors affecting G x E interaction of maize, Romay et al. (2010) indicated that when plant breeders think of breeding for environmental stress tolerance, they often design a breeding program for improving yield under one of the common climatic stresses, without

conducting preliminary studies on which environmental factors actually limit the crop and which genetic parameters are essentially affected. This highlights the importance of interpretive studies of G x E as tools for the design of breeding programs. Another advantage of interpreting G x E interactions is that it can provide insight into the genetic/physiological make-up of genotypes and eventually assist in ideotype design, direct breeding objectives, and identify genotypes suitable for further fundamental physiological research.

Interpretation of G x E essentially involves characterizing plant responses to environmental factors, an objective which can be met through the use of physiological crop models. However, most physiological models are too restrictive to deal with the diversity of interacting factors or require too much investment in resources to gain enough information necessary to explain differences in performance (Cooper and Byth, 1996). Statistical models used in MET analyses are based on actual responses to environments, and if utilized differently, can provide valuable insight into differences in plant growth.

Genotype x environment interactions are commonly interpreted following the characterization of environments in METs. This normally occurs through the gathering of relevant climatic and soil data from trial sites during cropping seasons. These supplementary data are then interpreted in conjunction with yield data to help explain responses. Very often, additional genotypic data are also gathered and interpreted relative to supplementary environmental data to investigate trait responses to environmental drivers. However, genotype responses depend on the influence of many interacting factors that differ in type, intensity, and timing. In addition to the direct effects of such factors on genotype performance, their interactions as well as their indirect associated effects must be considered. As a result of our lack of complete understanding of the effects of interactions between factors, different indirect approaches have been used to characterize environments to facilitate better interpretations.

2.3.2.1 Methods of characterizing environments

Bidinger et al. (1996) suggested a resource-yield approach, where the G x E interaction is interpreted at several levels: resource availability, resource capture, resource use efficiency, and partitioning to yield. In terms of resource availability, it is relatively simple to evaluate performance (conventionally yield) of genotypes by grouping trial sites according to available resources. For example, trials in a MET dataset could be grouped into subsets

according to total rainfall received during the season, and genotype x subset interactions could be used to characterize genotype adaptability to low rainfall conditions. This method can be further extended to include a soil water balance, whereby sites are characterized by the fraction of total potential evapotranspiration experienced during the season. The input data required for such characterizations include daily rainfall, total available soil moisture (TAM), and pan-evaporation, which can easily be accessed from most trial sites. This method can be used to create a range of indices that can be summarized within crop growth phases to help understand crop sensitivities to moisture stress and also group trial sites to interpret G x E interactions.

However, crop growth and yield is more closely related to the amount of resources actually used by the crop than to the amount simply available. Therefore, characterizing trial sites according to resources captured provides more accurate descriptions of G x E interactions. For example, the fraction of incoming radiation intercepted by a crop can be estimated from standard crop coefficients, and sites can be grouped according to this fractional interception. At a more detailed level, methods utilizing radiation use efficiency (RUE) and transpiration efficiency (TE) can also be applied to understand how G x E interactions are influenced by differential radiation and transpiration-driven biomass production. With detailed crop measurements, it is even possible to interpret G x E interactions in relation to differences in partitioning at the genotype and environmental levels. Unfortunately, this level of detail is seldom available in traditional METs, where the extent of information is limited to meteorological and soil data (Bidinger et al., 1996). Nevertheless, with the application of integrated techniques utilized in crop research, novel approaches have been adopted for the characterization of environments in METs.

2.3.2.2 Basic use of meteorological data

Meteorological data recorded at weather stations close to trial sites have been used to explain differential genotypic responses in a range of studies. Baril et al. (1995) used meteorological data to calculate environmental covariates that included the number of frost days in the first and second halves of April, mean temperature over the growing season, and total radiation over the growing season, to interpret G x E interactions in potato (*Solanum tuberosum*) variety trials. Many studies have additionally summarized meteorological data within crop growth phases to identify sensitive periods and their effects on G x E interactions. Van Eeuwijk and Elgersma (1993) defined five developmental periods in perennial ryegrass

(*Lolium perenne*) and summarized the minimum and maximum temperatures, rainfall, relative humidity and wind velocity for each period from METs, to identify factors influencing G x E interactions of seed yield. Vargas et al. (1998) summarized the mean daily maximum and minimum temperatures, monthly total precipitation, and sun hours per day within different months of the growing season (December to March) to interpret G x E interactions in a wheat MET dataset. In a second dataset, the same workers summarized the mean daily maximum and minimum temperatures, and sun hours per day within the vegetative stage, spike growth stage, grain filling stage, and entire growth cycle of another bread wheat MET dataset. It is interesting to note that most studies where covariates were summarized within growth phases was done on determinate grain crops such as wheat, barley, or maize, while very limited similar work has been done on crops like sugarcane where harvesting of vegetative organs is of primary economic concern.

In some instances, knowledge of plant responses to critical environmental limitations has been included in the analyses. For example, Van Eeuwijk et al. (1995) used "the number of days with mean temperature below 10°C in May" as a covariate when analyzing and interpreting a maize MET dataset. More recently, Romay et al. (2010) used "the number of days with maximum temperature above 25°C" and "number of days with average daily temperature above 15°C" as covariates when investigating the climatic and genotypic effects of environmental factors on grain yield of maize. In addition to the direct usage of meteorological and soil data to characterize environments, the recent advances in crop modeling have allowed indirect applications of these data to more effectively describe environments.

2.3.2.3 Use of crop growth models and indices to characterize environments

Meteorological data can also be used to calculate derived indices to describe a season or growth phase, which is then used as a covariate in the analysis of METs. Most studies have focused on calculating indices related to water usage and water stress. Van Oosterom et al. (1996a) calculated a water satisfaction index (WSI) from daily data on rainfall, pan evaporation, and a crop coefficient. The changes in the WSI before and after flowering enabled the estimation of the magnitude of pre- and post-flowering drought stress in pearl millet (*Pennisetum glaucum*). It was also shown that the WSI was three times more effective in explaining yield differences between environments than was actual seasonal rainfall. Similarly, the use of yield to characterize environments is generally less accurate than

applying resource use/availability concepts because yield is determined by both genetic and environmental effects. Voltas et al. (1999) calculated a ratio of rainfall and total evapotranspiration demand as a covariate to investigate G x E interactions for grain filling in barley. Voltas et al. (2005) used a daily water balance from soil characteristics and calculated the total accumulated available water (AW) as well as accumulated daily evapotranspiration demand (ET) from temperature and solar radiation measurements. A drought index (1-AW/ET) was subsequently calculated either pre- or post-anthesis for winter and spring wheat MET datasets to help interpret G x E interactions. Muchow et al. (1996), in a case study of sorghum (*Sorgum bicolor*) growth in water-stressed environments compared three indices of water deficit that varied in complexity and data input requirement. It was found that a simple index based on rainfall and potential evapotranspiration poorly characterized environments, while indices based on a soil water balance and relative transpiration index derived from a sorghum simulation model were more successful in identifying groups of seasons.

As suggested by Muchow et al. (1996), crop growth models are powerful tools that can be used in combination with MET data to help interpret G x E interactions. Their usefulness in this regard seems to lie in their ability to provide better characterizations of environments by estimating parameters such as crop evapotranspiration. However, other applications of crop models in G x E analyses have also been shown. Aggarwal et al. (1996) varied genotypic parameters in a rice simulation model and identified useful physiological traits that could be improved to enhance genotypic yield potentials. In the same study, it was shown that crop simulation models were able to generate G x E interactions similar to those observed in experimental data. This concept of varying input parameters in simulation models to create hypothetical genotypes was also demonstrated by Hammer et al. (1996), who indicated that such techniques could be used to generate hypotheses on plant types most suited to specific environments i.e. the methods can be used in the process of ideotype design. In another application of simulation modelling, Henderson et al. (1996) compared experimental results from METs with model simulations based on the characteristics of the environments to determine the contribution of genotype to the G x E interactions of rice in Thailand. Such successful applications of simulation models and environmental characterizations are in part attributed to an increased understanding of environmental control of crop growth and yield, which allow for more effective descriptions and analyses of environmental effects in MET programs. When used in combination with modern pattern analysis techniques (GGE biplot,

AMMI) and other statistical methods, this knowledge allows for a more effective exploitation of information contained within G x E datasets.

2.3.3 Interpretation of G x E interaction using statistical models

The analytical method of analyzing G x E interactions is generally considered more useful than the straightforward empirical method, as there is greater scope for understanding differential plant responses to environmental drivers. These methods require supplementary data such as meteorological, soil, or genotype trait information from METs in order to be administered. Such data can be obtained using some of the techniques described in the previous sections. Current methods of interpreting G x E interactions make use of factorial regression techniques, as well as various complementary analyses of AMMI and GGE biplot.

2.3.3.1 Interpretation of G x E interactions using Factorial Regression

Factorial regression is a useful tool to describe the G x E interaction in terms of genotypic and environmental covariates (Denis, 1988). The general form of a factorial regression model based on a two-way G x E table with concomitant variables measured in environments is

$$Y_{ij} = \mu + G_i + E_j + \sum \delta_{ik} Z_{jk} \tag{7}$$

where Y_{ij}, μ, G_i, and E_j are the same as in (1); Z_{jk} refers to the value of any environmental variable k for environment j; δ_{ik} represents the sensitivity of genotype i to the explicit environmental variable k. The model can also be extended to accommodate phenotypic descriptors other than yield as described by Vargas et al. (1999). Factorial regression can be used to describe the G x E interaction in terms of differential sensitivity of genotypes to measured environmental variables. After fitting the main effects of genotype and environment, associated variables on the levels of the environmental factor can then be introduced into the model in a multiple regression approach. One of the advantages of factorial regression is that hypotheses about the influence of associated environmental variables on the G x E interaction can be statistically tested and viewed within an ANOVA table.

Many studies have employed factorial regression techniques and successfully interpreted G x E interactions. Baril et al. (1995) applied factorial regression to interpret G x E interactions of potato. In that study it was found that factorial regression was successfully implemented to interpret potato responses to factors such as temperature, frost, and soil type, and it was

shown that soil type was the main determinant of the G x E interaction. Van Eeuwijk et al. (1995), using factorial regression analyses, showed that the total radiation during the growing season, the number of days with mean temperature below 10°C in May, and the mean temperature between July and August accounted for 51.8% of the cultivar x year interaction in a Dutch maize MET. Factorial regression also showed that the incidences of drought and high temperatures during grain filling were responsible for the differential genotypic responses in individual grain weight in a study of G x E interactions for grain filling in barley (Voltas et al., 1999). In an analysis of Spanish winter and spring wheat trials, Voltas et al. (2005) used GGE biplot and factorial regression to identify genotypic and environmental factors causing yield variation. In that study it was shown that pre-flowering thermal time and drought incidence during grain filling caused genotype-dependant responses in grain yield of winter wheat. Other examples of successful applications of factorial regression are given by Ceretta and Van Eeuwijk (2008) and Romay et al. (2010).

As with all linear regression models, factorial regression models become difficult to deal with when there are many explanatory variables that are highly correlated, and Voltas et al. (2005) therefore suggests that the relevant environmental descriptors used should be restricted to a reasonably low number for ease of interpretation. To this end, the use of graphical displays such as biplots derived from AMMI or GGE biplot do offer the visual interpretive advantage when combined with complementary analyses.

2.3.3.2 Interpretation of G x E interactions using AMMI

Although routinely employed in empirical studies of G x E interactions to investigate genotypic stability, environmental grouping, and differential genotypic adaptability, AMMI biplots can also be used analytically to interpret the G x E interaction. Van Eeuwijk and Elgersma (1993) showed that the AMMI IPCA1 and IPCA2 scores of environments can be regressed against measured environmental covariates. Differential positioning of environments on a biplot can then be interpreted relative to environmental covariates if significant correlations between IPCA scores and environmental covariates exist. Van Eeuwijk and Elgersma (1993) found that four environmental covariates (summarized within development periods) showed strong correlations to environment IPCA scores and these corresponded to the covariates identified by factorial regression, in an analysis of G x E interactions of ryegrass. Similarly, Van Oosterom et al. (1996b) related IPCA1 scores to environmental covariates and found that factors such as the mean maximum temperature

during 10 days after flowering and the changes in a water satisfaction index during grain filling were some of the causes of G x E interactions of pearl millet.

The concept of regressing environment IPCA scores on the environmental covariates can be extended to enrich AMMI biplots with environmental covariates included as points (Van Eeuwijk et al., 1995). In this case, the regression coefficients from the regression of the covariates on the scores of both AMMI axes simultaneously serve as coordinates for that covariate on the biplot. The same methodology can be employed to include cultivar covariates on the biplot. A condition for the inclusion of such information on the AMMI biplot is that enough variation in the covariates is explained by the regression on the scores. The enrichment of AMMI biplots using regressions of IPCA scores on environmental covariates has been demonstrated in studies. Vargas et al. (1999) simultaneously enriched AMMI biplots with genotypic and environmental covariates in a study of a wheat MET dataset. In that study, it was found that this technique, together with factorial regression and partial least squares regression (PLS) identified similar cultivar and environmental covariates that explained most of the G x E interaction. Voltas et al. (1999) used this method to identify genotypic and environmental covariates influencing G x E interactions for grain filling in barley. Here, the method identified three covariates, namely; the ratio of rainfall to evapotranspiration demand, maximum temperatures during grain filling, and mean individual grain weight, which were subsequently included into a factorial regression model to interpret G x E interactions. The approach employed in these studies, and others (Voltas et al., 1999) used the enriched AMMI biplots as a preliminary screening of environmental and genotypic covariates before these were tested statistically using factorial regression.

2.3.3.3 Interpretation of G x E interactions using GGE biplots

The PC scores of genotypes and environments obtained from a GGE biplot can also be used to interpret causes of G x E interactions. Yan and Hunt (2001) correlated genotypic and environmental PC1 and PC2 scores from a GGE biplot to genotypic traits and environmental factors measured as covariates. They found that plant height and maturity were the main genotypic causes of G x E interaction, while cold temperatures in winter and hot temperatures in summer were the major environmental causes of G x E interaction of winter wheat yields in Canada. In order to identify specific genotypic traits driving the G x E interaction of barley in North America, Yan and Tinker (2005) constructed a covariate-effect biplot. This type of biplot is based on a trait x environment two-way table of correlation coefficients between the

traits and yield in each of the test environments. The method was used to identify kernel weight, earlier heading, and better lodging resistance as appropriate selection traits for the improvement of barley yields in an eastern mega-environment. The frequency of use of GGE biplot as a tool for the interpretation (analytical) of G x E interaction has not been as high compared to the use of AMMI or factorial regression. However, the use of GGE biplot as a complementary tool to assist with G x E interpretation has been demonstrated (Voltas et al., 2005).

2.3.3.4 Interpretation of G x E interaction using partial least squares regression

In situations where there are many explanatory genotypic or environmental covariates which are highly correlated, the technique of PLS regression may be appropriate to interpret G x E interactions. The PLS regression method is a multivariate regression technique that can be used to compare blocks of variates, or relate one or several response variates to several correlated explanatory variates. The method involves relating X and Y matrices in one single estimation procedure. The Y matrix may contain the G x E yield data as dependant variables, while the X matrix contains the external environmental covariates as explanatory variates (Vargas et al., 1998).

2.4   Analysis of G x E interactions of sugarcane

Despite the wide range of available techniques to analyze G x E interactions, their applications to sugarcane are limited in comparison with other field crops. The majority of studies of G x E interactions of sugarcane fall within the empirical category, where the focus was on genotype stability and identification of homogenous environments within breeding programs (Kang and Miller, 1984; Tai et al., 1982). Jackson et al. (1991) used ANOVA, cluster analysis and PCA to investigate the G x E interactions of three datasets of sugarcane pre-release trials in Australia. In that study significant G x E interactions were observed in all three datasets, and it was concluded that testing across sites was more important than testing across ratoons within sites for identification of superior clones. Rattey and Kimbeng (2001) used variance components analyses to estimate the magnitude of G x E interactions and determine optimal resource allocation in final stage selection trials in the Burdekin district of Australia. They found that genotype x crop (ratoon) contributed more to G x E interactions than genotype x location x ratoon and genotype x location interactions, respectively.

A limited number of sugarcane G x E studies have employed multivariate techniques. Bissessur et al. (2001) showed that AMMI was more effective than ANOVA at identifying significant G x E interactions in a study of final stage selection trials in Mauritius. They found that AMMI was effective at identifying cultivars with broad and specific adaptation and recommended that the technique be routinely used to obtain additional information on clones prior to their commercial cultivation. Similarly, Queme et al. (2001) and Queme et al. (2005) demonstrated that AMMI could be successfully employed to analyse G x E interactions and identify specific cultivar adaptability of sugarcane METs in Guatemala. In India, Srivastava et al. (1999) showed that AMMI was more accurate than the sites regression model (Finlay and Wilkinson, 1963) as it captured a greater percentage of the G x E interaction sums of squares in an analysis of eight sugarcane cultivars in a MET. Redshaw et al. (2002) compared AMMI and restricted maximum likelihood (REML) techniques for the analysis of an unbalanced MET dataset in South Africa. That initial study showed that the dryland regions of the South African sugar industry could be divided into four mega-environments and the authors concluded that both AMMI and REML applications should be investigated further.

Compared to AMMI, relatively fewer studies have employed the GGE biplot method as a tool for the analysis of sugarcane METs. Queme et al. (2007) employed the GGE biplot technique to group sites and identify superior cultivars in each group using MET data from a low production zone in Guatemala. Glaz and Kang (2008) used GGE biplots to identify redundant sites used during selection in the Florida sugarcane industry. They used the technique to successfully identify locations with organic soils that, if replaced by locations with sandy soils, would be least likely to compromise the ability of the selection program to identify superior cultivars for the industry.

The G x E interactions in sugarcane have also been demonstrated using numerous variations of traditional analyses and "non-conventional" methods. Such studies have tended to focus on the crop characteristics of specific industries. For example, Gilbert et al. (2006) employed analyses such as repeated measures ANOVA and identified significant genotype x environment, environment x time of harvest, and genotype x environment x time of harvest interactions in Florida. Jackson et al. (2007) used restricted maximum likelihood (REML) analysis to investigate the magnitude of genotype x region interactions relative to genotype x environment interactions within regions for the Australian sugar industry. This study showed that genotype x region interactions were small compared to genotype x environment

interactions within regions, and that indirect selection in a region different to that being targeted would be only slightly less effective than selection within the targeted region itself. In South Africa, Ramburan et al. (2010b) categorized post-release cultivar evaluation trials based on age and time of harvest and demonstrated highly significant genotype x age and genotype x time of harvest interactions using REML. These results were subsequently used to characterize varieties for inclusion into a decision support system.

Despite the examples of the existence of G x E interactions in sugarcane, attempts to interpret such interactions are very limited. The most recent and only attempt to interpret sugarcane G x E interactions relative to environmental factors was conducted by Jackson et al. (1995). That study used cluster analysis and principal component analysis to relate sugarcane family x site interactions to environmental factors. They found that key climatic factors such as rainfall, temperature and solar radiation did not appear to be important in causing family x site interactions. However, significant correlations were found between soil nutrients (calcium, copper and zinc) and the PC2 scores from PCA analysis, suggesting that these nutrients had effects on the family x site interactions. The scarcity of studies involving interpretation of sugarcane G x E interactions and the relative benefits of such studies to improvements in selection strategies, as demonstrated in other crops, suggest that more priority should be given to this area of sugarcane research. Due to the lack of widespread post-release cultivar evaluation programs in sugarcane (Ramburan and Van den Berg, 2011), most studies on G x E interactions of sugarcane have been conducted within the realms of breeding and selection. Post-release cultivar evaluation programs may also be suitable resources within which analytical (interpretive) studies of G x E interactions are possible. This is due to the wider range of environmental conditions typically evaluated, which is associated with greater representativeness of an industry.

2.5   Summary

The existence of significant crossover G x E interactions can be a serious constraint to crop improvement, leading to reduced genetic gains from selection and a waste of invested resources. Most of the common statistical techniques available to analyse G x E interactions are empirical in nature, focusing on quantification of G x E effects, identifying stable cultivars, and identification of mega-environments. These include techniques such as ANOVA (Fisher, 1925), joint linear regression (Finlay and Wilkinson, 1963), REML (Patterson and Thompson, 1971), AMMI (Gauch, 1992), or GGE biplots (Yan et al., 2000).

Comparatively fewer studies have attempted to interpret G x E interactions in relation to major environmental variables, and in doing gain a more comprehensive understanding of the interactions. Such information can be utilized to improve gains from selection through strategic site selection, identify cultivars suitable for fundamental physiological research, and assist in ideotype design. The techniques used to interpret G x E involves characterization of trial sites according to environmental factors, either using direct measurements, calculated indices, or derived variables from crop growth models. These covariates can then be analyzed in combination with modern multivariate techniques such as pattern analysis, AMMI or GGE biplot to identify patterns of G x E interactions and identify critical factors driving the interactions. Complementary analyses such as factorial regression may then be used to statistically test the significance of specific factor effects on the G x E interactions (Vargas et al., 1999). These techniques, or various modifications thereof, have been demonstrated successfully in a range of other crops (Baril et al., 1995; Van Eeuwijk et al., 1995; Van Oosterom et al., 1996b; Voltas et al., 2005; Yan and Tinker, 2005).

Studies of G x E interactions of sugarcane are limited, and when attempted, have focused in a largely empirical manner on design of breeding and selection networks (Jackson et al., 1991; Rattey and Kimbeng, 2001; Queme et al., 2005; Srivastava et al., 1999). Apart from a single interpretive study of G x E interaction in sugarcane (Jackson et al., 1995), no other attempt has been made to understand sugarcane cultivar responses to environmental factors using a MET approach. This is in spite of numerous examples demonstrating the existence of G x E interaction in various sugarcane industries around the world. The interpretation of G x E interactions of sugarcane through analysis of an extensive MET dataset (with environmental covariates) may provide opportunities to gain further understanding of sugarcane growth and development. Additionally, the use of a post-release cultivar evaluation program as a tool to implement these strategies may be a logical option due to the diversity of conditions evaluated and the general flexibility of such trial networks (Ramburan and Van den Berg, 2011). Such a thrust would represent a first for sugarcane and will shed light on the likely causes of G x E interactions in the crop.

2.6   References

Abamu, F.J., Akinsola, E.A., Alluri, K., 1998. Applying the AMMI models to understand genotype-by-environment (GxE) interactions in rice reaction to blast disease in Africa. *International Journal of Pest Management* 44, 239-245.

Aggarwal, P.K., Kropff, M.J., Matthews, R.B., McLaren, C.G., 1996. Using simulation models to design new plant types and to analyse genotype by environment interactions in rice. p. 403-418. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement . CAB International, Wallingford, UK.

Baril, C.P., Denis, J.B., Wustman, R., Van Eeuwijk F.A., 1995. Analysing genotype by environment interaction in Dutch potato variety trials using factorial regression. *Euphytica* 82, 149-155.

Bidinger, F.R., Hammer, G.L., Muchow, R.C., 1996. The physiological basis of genotype by environment interaction in crop adaptation. p. 329-347. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement. CAB International, Wallingford, UK.

Bissessur, D., Chong, L.C., Ramnawaz, C., Ramdoyal, K., 2001. Analysing G x E interaction in sugar cane using the Additive Main Effects and Multiplicative Interaction (AMMI) Model. *Proceedings of the International Society of Sugar Cane Technologists* 24, 506-511.

Brett, P.C.G., 1946. Seed setting of sugarcane in South Africa. *Nature* 157, 657.

Ceretta, S., Van Eeuwijk, F.A., 2008. Grain yield variation in malting barley cultivars in Uruguay and its consequences for the design of a trials network. *Crop Science* 48, 167-180.

Cooper, M., Byth, D.E., 1996. Understanding plant adaptation to achieve systematic applied crop improvement − A fundamental challenge. p. 5-23. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement . CAB International, Wallingford, UK.

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

Crossa, J., Fox , P.N., Pfeiffer, W.H., Rajaram, S., Gauch, H.G., 1991. AMMI adjustment for statistical analysis of an international wheat yield trial. *Theoretical and Applied Genetics* 81, 27-37.

Dehghani, H., Ebadi, A., Yousefi, A., 2006. Biplot analysis of genotype by environment interaction for Barley yield in Iran. *Agronomy Journal* 98, 388-393.

DeLacy, I.H., Basford, K.E., Cooper, M., Bull, J.K., McLaren, C.G., 1996. Analysis of multi-environment trials − An historical perspective. p. 39-124. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement . CAB International, Wallingford, UK.

Denis, J.B., 1988. Two-way analysis using covariates. *Statistics* 19, 123-132.

Finlay, K.W., Wilkinson, G.N., 1963. The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research* 14, 742-754.

Fisher, R.A., 1925. Statistical methods for research workers. Oliver and Boyd, Edinburgh, Scotland.

Gauch, H.G., 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam.

Gauch, H.G., Zobel, R.W., 1997. Identifying mega-environments and targeting genotypes. *Crop Science* 37, 311-326.

Gauch, H.G., 2006. Statistical analysis of yield trials by AMMI and GGE. *Crop Science* 46, 1488-1500.

Gilbert, R.A., Shine, J.M., Miller, J.D., Rice, R.W., Rainbolt, C.R., 2006. The effect of genotype, environment and time of harvest on sugarcane yields in Florida, USA. *Field Crops Research* 95, 156-170.

Glaz, B., Kang, M.S., 2008. Location contributions determined via GGE Biplot analysis of multi-environment sugarcane genotype-performance trials. *Crop Science* 48, 941-950.

Gomez, L.D., Steele-King, C.G., McQueen-Mason, S.J., 2008. Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist* 178, 473-485.

Gopinathan, M.C., 2010. Research and technology transfer strategies for the next decade – An Indian example. *Proceedings of the International Society of Sugar Cane Technologists* 27, 1-15.

Hammer, G.L., Butler, D.G., Muchow, R.C., Meinke, H., 1996. Integrating physiological understanding and plant breeding via crop modeling and optimization. p. 419-441. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement . CAB International, Wallingford, UK.

Henderson, S.A., Fukai, S., Jongdee, B., Cooper, M., 1996. Comparing simulation and experimental approaches to analyzing genotype by environment interactions for yield in

rainfed lowland rice. p. 443-464. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement . CAB International, Wallingford, UK.

Ibanez, M.A., DI Renzo, M.A., Samame, S.S., Bonamico, N.C., Poverene, M.M., 2001. Genotype-environment interaction of lovegrass forage yield in the semi-arid region of Argentina. *Journal of Agricultural Science, Cambridge* 137, 329-336.

Jackson, P.A., Horsley, D., Foreman, J., Hogarth, D.M., Wood A.W., 1991. Genotype x environment (GE) interactions in sugarcane variety trials in the Herbert. *Proceedings of the Australian Society of Sugar Cane Technologists* 13, 103-109.

Jackson, P., McRae, T., Hogarth, M., 1995. Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. *Field Crops Research* 43, 119-130.

Jackson, P., Chapman, S., Rattey, A., Wei, X., Cox, M., 2007. Genotype x region interactions, and implications for sugarcane breeding programs. *Sugar Cane International* 25, 3-7.

Kang, M.S., Miller, J.D. 1984. Genotype x environment interactions for cane and sugar yield and their implications in sugarcane breeding. *Crop Science* 24, 435-440.

McLaren, C.G., 1996. Methods of data standardization used in pattern analysis and AMMI models for the analysis of international multi-environment variety trials. p. 225-242. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement . CAB International, Wallingford, UK.

Meyer, J., 2007. Advances in field technology and environmental awareness in the South African sugar industry. *Sugar Cane International* 25, 22-28.

Muchow, R.C., Cooper, M., Hammer, G.L., 1996. Characterizing environmental challenges using models. p 349-364. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement . CAB International, Wallingford, UK.

Parfitt, R.C., 2005. Release of sugarcane varieties in South Africa. *Proceedings of the South African Sugar Technologists Association* 79, 63-71.

Patterson, H.D., Thompson, R., 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545–554.

Queme, J.L., Orozco, H., Ovalle, W., Soto, G., Melgar, M., 2001. AMMI analysis of sugarcane cultivar trials in Guatemala. *Proceedings of the International Society of Sugar Cane Technologists* 24, 512-515.

Queme, J.L., Orozco, H., Ovalle, W., Melgar, M., 2005. Analysis of genotype-by-environment interaction for sugarcane based on the AMMI model. *Sugar Cane International* 23, 21-24.

Queme, J.L., Crossa, J., Orozco, H., Melgar, M., 2007. Analysis of genotype-by-environment interaction for sugarcane using the sites regression model (SREG). *Proceedings of the International Society of Sugar Cane Technologists* 26, 764-769.

Ramburan, S., McElligott, D.M., De Haas, O., 2010a. Integrating agronomic and economic principles through applications of variety trial data: The Illovo case study. *Proceedings of the South African Sugar Technologists Association* 83: 106-116.

Ramburan, S., Paraskavopoulos, A., Saville, G., Jones M., 2010b. A decision support system for sugarcane variety selection in South Africa based on genotype-by-environment analysis. *Experimental Agriculture* 46, 243-257.

Ramburan, S., Van den Berg, M., 2011. Review and analysis of post-release variety evaluation of sugarcane: A South African perspective. *International Sugar Journal* 113, 22-29.

Rattey, A.R., Kimbeng, C.A., 2001. Genotype by environment interactions and resource allocation in final stage selection trials in the Burdekin district. *Proceedings of the Australian Society of Sugar Cane Technologists* 23, 136-141.

Redshaw, K.A., Govender, N., Smit, M., 2002. Investigating two statistical techniques used in the analysis of sugarcane variety trials. *Proceedings of the South African Sugar Technologists Association* 76, 219-223.

Romay, C.N., Malvar, R.A., Campo, L., Alvarez, A., Moreno-Gonzalez, J., Ordas, A., Revilla, P., 2010. Climatic and genotypic effects for grain yield in maize under stress conditions. *Crop Science* 50, 51-58.

Sharma, R.C., Morgounov, A.I., Braun, H.J., Akin, B., Keser, M., Bedoshvili, D., Bagei, A., Martius, C., Van Ginkel M., 2010. Identifying high yielding stable winter wheat genotypes for irrigated environments in central and west Asia. *Euphytica* 171, 53-64.

Sivapalan, S., O'Brien, L., Ortiz-Ferrara, G., Hollamby, G.J., Barclay, I., Martin P.J., 2000. An adaptation analysis of Australian and CIMMYT/ICARDA wheat germplasm in Australian production environments. *Australian Journal of Agricultural Research* 51, 903-915.

Smit, M.A., Singels, A., Van Antwerpen, R., 2004. Differences in canopy development of two sugarcane cultivars under conditions of water stress: Preliminary results. *South African Sugar Technologists Association* 78, 149-152.

Srivastava, H.M., Gauch, H.G. Jr., Kulshreshtha, N., Misra, G.P., 1999. Genotype x Environment interaction – Efficiency and use of AMMI Matmodel in Sugarcane. *Proceedings of the International Society of Sugar Cane Technologists* 20, 476-485.

Tai, P.Y.P., Rice, E.R., Chew, V., Miller, J.D., 1982. Phenotypic stability analyses of sugarcane cultivar performance tests. *Crop Science* 22, 1179-1184.

Van Antwerpen, R., Meyer, J., 1998. Soil degradation – II: Effects of trash and inorganic fertilizer application on soil strength. *South African Sugar Technologists Association* 72, 152-158.

Van Eeuwijk, F.A., Elgersma, A., 1993. Incorporating environmental information in an analysis of genotype by environment interaction for seed yield in perennial ryegrass. *Heredity* 70, 447-457.

Van Eeuwijk, F.A., Keizer, L.C.P., Bakker, J.J., 1995. Linear and bilinear models for the analysis of multi-environment trials: II An application to data from the Dutch Maize variety trials. *Euphytica* 84, 9-22.

Van Oosterom, E.J., Bidinger, F.R., Mahalakshmi, V., Rao, K.P., 1996a. Effect of water availability pattern on yield of pearl millet in semi-arid tropical environments. *Euphytica* 89, 165-173.

Van Oosterom, E.J., Mahalakshmi, V., Bidinger, F.R., Rao, K.P., 1996b. Effect of water availability and temperature on the genotype-by-environment interaction of pearl millet in semi-arid tropical environments. *Euphytica* 89, 175-183.

Vargas, M., Crossa, J., Sayre, K., Reynolds, M., Ramirez, M.E., Talbot, M., 1998. Interpreting genotype x environment interaction in wheat by partial least squares regression. *Crop Science* 38, 679-689.

Vargas, M., Crossa, J., Van Eeuwijk, F.A., Ramirez, E., Sayre K., 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype x environment interaction. *Crop Science* 39, 955-967.

Voltas, J., Van Eeuwijk, F.A., Sombrero, A., Lafarga, A., Igartua, E., Romagosa, I., 1999. Integrating statistical and ecophysiological analyses of genotype by environment interaction for grain filling of barley I. Individual grain weight. *Field Crops Research* 62, 63-74.

Voltas, J., Lopez-Corcoles, H., Borras, G., 2005. Use of biplot analysis and factorial regression for the investigation of superior genotypes in multi-environment trials. *European Journal of Agronomy* 22, 309-324.

Yan, W., 2001. GGEbiplot – A Windows application for graphical analysis of multienvironment trial data and other types of two-way data. *Agronomy Journal* 93, 1111-1118.

Yan, W., Hunt, L.A., 2001. Interpretation of genotype x environment interaction for winter wheat in Ontario. *Crop Science* 41, 19-25.

Yan, W., Hunt, L.A., Sheng, Q., Szlavnics, Z., 2000. Cultivar Evaluation and Mega-Environment investigation based on the GGE Biplot. *Crop Science* 40, 597-605.

Yan, W., Kang, M.S., 2003. GGE Biplot Analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL.

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

Yan, W., Tinker, N.A., 2005. An integrated bipot analysis stem for displaying, interpreting, and exploring genotype x environment interaction. *Crop Science* 45, 1004-1016.

Yang, R., Crossa, J., Cornelius, P.L., Burgueno, J., 2009. Biplot analysis of genotype x environment interaction: Proceed with caution. *Crop Science* 49, 1564-1576.

CHAPTER 3

VARIANCE COMPONENTS, MEGA-ENVIRONMENTS, AND GENOTYPE STABILITY ACROSS THE RAINFED REGIONS[1]

3.1 Abstract

The objective of this study was to evaluate the current regional sub-divisions, genotype stability, and magnitude of G x E interactions in the rainfed parts of the industry to direct future selection and evaluation procedures. Cane yield (TCANE), estimated recoverable crystal (ERC) and tons ERC/ha (TERC) data of 15 genotypes from 153 environments (trial x ratoon combinations) harvested between 1999 and 2009 were analysed using variance components and GGE biplot analysis. Differences between trials accounted for the largest proportion of variation in TCANE (35%), ERC (44%), and TERC (47%), while genotype differences were smaller than the G x E interactions for TCANE and TERC, but not for ERC. Larger genotype x trial interactions compared with genotype x ratoon interactions suggested that more emphasis be placed on sampling more trial sites rather than testing more ratoons within trials. The GGE biplots for TERC revealed four mega-environments, each with environments from the coast, hinterland and midlands. Ratoons of the same trials and trials from the same location occasionally grouped into different mega-environments, suggesting large spatial and temporal (seasonal) variability in environmental conditions that can only be managed through extensive trial characterisation. Genotype N31 was identified as most stable and high yielding, suggesting the potential value of the variety as an alternative control to NCo376 for rainfed selection sites. These findings highlighted the need for increased G x E studies to enhance efficiencies of the trial networks in the industry, and illustrated the usefulness of GGE biplot as a tool to evaluate sugarcane multi-environment trials.

3.2 Introduction

Multi-environment trials (METs) are routinely conducted in major cropping industries throughout the world to gain insights into G x E interactions and identify high yielding, stable cultivars that are more adapted to regional agro-climatic conditions. Besides identifying

---

[1]Ramburan, S., Zhou, M., 2011. Investigating sugarcane genotype x environment interactions under rainfed conditions in South Africa using variance components and biplot analysis. *Proceedings of the South African Sugar Technologists Association* 84, 345-358.

superior cultivars suited to particular environments, METs also provide information on the nature of the target population of environments (TPEs), and how they should be subdivided for further selection and evaluation. Such subdivisions into mega-environments (MEs) are necessary for the implementation of region-specific breeding strategies, which can ensure the greatest gains from selection. A ME may be defined as a portion of a crop species' growing region with a homogeneous environment that causes some genotypes to perform similarly (Gauch and Zobel, 1997), and is normally identified through analyses of MET data. Subdivision of the TPE is also important for the development of recommendation domains.

Investigations of MEs are a prerequisite for any meaningful cultivar evaluation and recommendation procedures (Yan and Hunt, 2001). This implies that cultivar improvement and evaluation networks established without preliminary ME investigations could be at risk of reduced genetic gains from selection due to inaccurate regional subdivisions. Another important objective of G x E research is to gain an understanding of the relative contributions of different components of variation on yield variability (Smith et al., 2005). Such information is essential to inform decisions on selection and evaluation, investigate contributions of repeatable and non-repeatable components of G x E interactions, and improve the efficiencies of trial networks (DeLacy et al., 2010).

The rainfed environments of the South African sugar industry are partitioned into three broad regions characterized by different production conditions. The coastal (C) region is characterized by areas with low altitude (< 200 m asl), higher temperatures, and high topographic, soil, and rainfall variability. The midlands (M) region is characterized by higher altitudes (> 600 m asl), lower temperatures, flatter topography, and lower soil variability, while the hinterland (H) region is a transitional region between the midlands and the coast. The large spatial and temporal variation in production conditions, combined with the wide range of planting times (February to December) and harvest ages (10 to 24 months) creates a highly heterogeneous TPE. Cultivars for the rainfed parts of the sugar industry currently originate from five separate selection programs conducted on selection stations run by SASRI. The C region contains two selection stations/programs, broadly representing coastal high and low potential conditions, respectively. The midlands region is also characterized by two selection programs, broadly representing humic and sandy soils, while the hinterland region is represented by a single program. All selection programmes were established within the last 10-15 years.

The choice of stations (farms) for each program was based on expert knowledge of geographic, soil and climatic variability between the different sugarcane producing regions and the representativeness of stations to general growing conditions within the respective regions (Nuss, 1998). The choice of stations proceeded without any formalized preliminary study of their relevance as selection sites for the TPEs, and their effectiveness in this regard has not been evaluated to date. Additionally, no attempts have been made to identify MEs and hence evaluate the accuracy of the broad regional divisions (coastal, hinterland, and midlands). Currently, plant breeding selection trials at the different stations and on selected off-station sites (which are largely permanent) represent only a small proportion of the TPEs, and may not be representative enough for comprehensive ME investigations. Post-release cultivar evaluation trials are routinely conducted throughout the rainfed regions of the industry on commercial farms with different combinations of soil types, harvest age, harvest season, and husbandry practices, and are more representative of industry conditions. More importantly, the vast spatial distribution of this trial network suggests that it is a better sample of the TPE's compared with selection trials only.

Methods of analyzing MET data have included the conventional analysis of variance, linear regression (Finlay and Wilkinson, 1963), stability variance (Shukla, 1972), and AMMI (Gauch, 1992). The GGE biplot method developed by Yan et al. (2000) has recently gained popularity as a tool to investigate various aspects of G x E interactions. The method has been successfully applied to crops such as wheat (Sharma et al., 2010), barley (Dehghani et al., 2006), and soybean (Yan and Rajcan, 2002). Studies with sugarcane in Florida, USA (Glaz and Kang, 2008) and Guatemala (Queme et al., 2001) have also employed GGE biplot analysis; however, the methodology has not been attempted in South Africa. Variance components analyses have also been used widely to evaluate the relative contributions of relevant terms to variability and to assess the precision of trial networks to optimize resources (Ceretta and Van Eeuwijk, 2008). For example, De la Vega and Chapman (2010) used linear mixed models to estimate variance components attributed to trials, hybrids, replicates, and incomplete blocks, and tracked genetic progress for yield of commercial sunflower in Argentina.

Despite the availability of appropriate data and methods of analysis, no attempts have been made to improve sugarcane selection and evaluation procedures in South Africa using contemporary G x E analysis techniques. A systematic analysis of a series of METs has not

been conducted for the rainfed production regions, and information on ME constitution, industry-wide genotypic performance and stability, and components of variation are lacking. Recent concerns around the relevance of selection sites and differential performance of varieties in the TPE compared with performance during selection have prompted such an analysis. As the first step of a systematic analysis of G x E interactions, the objectives of this study were (i) to investigate the magnitude of G x E interactions, (ii) to evaluate variance components, and (iii) to identify MEs within the rainfed parts of the South African sugar industry.

3.3 Materials and methods

3.3.1 Trial datasets

The MET dataset used for this study comprised 43 trials conducted in 18 different locations, and harvested across one to six ratoons during the period 1999-2009. The majority of the trials (32) formed part of SASRI's post-release cultivar evaluation project; while other (11) late stage plant breeding selection trials were included as well. Most trials were established on commercial fields across the three regions and allowed to run for as many ratoons as the commercial plantings surrounding the trials. Plant breeding trials that were included in the dataset only ran for a maximum of three ratoons (plant crop plus two ratoons), and were conducted on SASRI selection stations. For the purpose of this study, an environment was defined as a trial x ratoon combination i.e. each crop that was harvested constituted an environment. As with most studies of this nature, the 'location' was a loose spatial reference identified by a town name, and trials at different sites in each location were numbered in the order of establishment. Table 3.1 describes the environments, which are abbreviated according to the region, location, site, and ratoon number, respectively. For example, CEM21 refers to a first ratoon crop, of the second trial conducted at Empangeni, in the coastal region. Cultivar (will be used interchangeably with the term genotype) numbers per trial varied from five to 12, and not all cultivars were included in each trial. Consequently, only the 15 most commonly tested cultivars were included in the analyses, with the majority of the trials containing the commercial control NCo376.

All trials were conducted as randomised complete block designs with four to six replicates. Trial plots consisted of five or six rows that were between 8 and 10 m long, spaced 1.0 to 1.2 m apart. Weed and fertilizer management proceeded as per commercial farm practice. At harvest, three or four net rows were cut and bundled by hand and weighed using a hydraulic

grab apparatus equipped with a load cell to determine cane yield in ton cane/ha (TCANE). A 12-stalk sucrose sample was taken from each plot to determine estimated recoverable crystal percentage (ERC), which is an estimate of the RV described in formula (1) in Chapter 2. For each trial plot, the tons ERC/ha (TERC) was calculated as the product of TCANE and ERC.

Table 3.1 Definitions of the environments making up the multi-environment trial dataset. Trials were conducted at various locations in the coastal (C), hinterland (H) and midlands (M) regions and were coded as environments defined by the respective site and ratoon number

Ref	Region	Location	Trial (within location)	Ratoon	Environ code	Ref	Region	Location	Trial (within location)	Ratoon	Environ code
1	C	Amatikulu (AK)	1	0	CAK10	80	H	Doringkop (DK)	1	0	HDK10
2	C	Amatikulu (AK)	1	1	CAK11	81	H	Doringkop (DK)	1	1	HDK11
3	C	Amatikulu (AK)	2	0	CAK20	82	H	Doringkop (DK)	1	2	HDK12
4	C	Amatikulu (AK)	3	0	CAK30	83	H	Doringkop (DK)	1	3	HDK13
5	C	Amatikulu (AK)	3	1	CAK31	84	H	Doringkop (DK)	1	4	HDK14
6	C	Amatikulu (AK)	3	2	CAK32	85	H	Doringkop (DK)	2	1	HDK21
7	C	Amatikulu (AK)	3	3	CAK33	86	H	Doringkop (DK)	2	2	HDK22
8	C	Amatikulu (AK)	4	0	CAK40	87	H	Doringkop (DK)	2	3	HDK23
9	C	Amatikulu (AK)	4	1	CAK41	88	H	Doringkop (DK)	3	0	HDK30
10	C	Amatikulu (AK)	4	2	CAK42	89	H	Doringkop (DK)	3	1	HDK31
11	C	Amatikulu (AK)	4	3	CAK43	90	H	Doringkop (DK)	3	2	HDK32
12	C	Amatikulu (AK)	4	4	CAK44	91	H	Eshowe (ES)	1	0	HES10
13	C	Amatikulu (AK)	4	5	CAK45	92	H	Eshowe (ES)	1	1	HES11
14	C	Empangeni (EM)	1	0	CEM10	93	H	Eshowe (ES)	1	2	HES12
15	C	Empangeni (EM)	2	0	CEM20	94	H	Eshowe (ES)	1	3	HES13
16	C	Empangeni (EM)	2	1	CEM21	95	H	Kearsney (KS)	1	1	HKS11
17	C	Empangeni (EM)	2	2	CEM22	96	H	Kearsney (KS)	1	2	HKS12
18	C	Empangeni (EM)	2	3	CEM23	97	H	Kearsney (KS)	2	1	HKS21
19	C	Empangeni (EM)	2	4	CEM24	98	H	Kearsney (KS)	2	2	HKS22
20	C	Empangeni (EM)	2	5	CEM25	99	H	Kearsney (KS)	3	0	HKS30
21	C	Empangeni (EM)	3	0	CEM30	100	H	Kearsney (KS)	3	1	HKS31
22	C	Empangeni (EM)	3	1	CEM31	101	H	Kearsney (KS)	3	2	HKS32
23	C	Empangeni (EM)	3	2	CEM32	102	H	Kearsney (KS)	4	0	HKS40
24	C	Empangeni (EM)	3	3	CEM33	103	H	Kearsney (KS)	4	1	HKS41
25	C	Empangeni (EM)	3	4	CEM34	104	H	Kearsney (KS)	4	2	HKS42
26	C	Empangeni (EM)	3	5	CEM35	105	H	Kearsney (KS)	4	3	HKS43
27	C	Empangeni (EM)	3	6	CEM36	106	H	Kearsney (KS)	4	4	HKS44
28	C	Empangeni (EM)	4	0	CEM40	107	H	Kearsney (KS)	4	5	HKS45
29	C	Empangeni (EM)	4	1	CEM41	108	H	Kearsney (KS)	5	0	HKS50
30	C	Empangeni (EM)	4	2	CEM42	109	H	Kearsney (KS)	5	1	HKS51
31	C	Empangeni (EM)	4	3	CEM43	110	H	Kearsney (KS)	5	2	HKS52
32	C	Empangeni (EM)	4	4	CEM44	111	H	Kearsney (KS)	5	3	HKS53
33	C	Empangeni (EM)	4	5	CEM45	112	H	Kearsney (KS)	5	4	HKS54
34	C	Empangeni (EM)	4	6	CEM46	113	H	Melmoth (MM)	1	0	HMM10
35	C	Empangeni (EM)	5	0	CEM50	114	H	Melmoth (MM)	1	1	HMM11
36	C	Empangeni (EM)	6	0	CEM60	115	H	Melmoth (MM)	1	2	HMM12
37	C	Empangeni (EM)	6	1	CEM61	116	H	Paddock (PD)	1	0	HPD10
38	C	Empangeni (EM)	6	2	CEM62	117	H	Paddock (PD)	2	0	HPD20
39	C	Empangeni (EM)	7	0	CEM70	118	H	Paddock (PD)	2	2	HPD22

40	C	Empangeni (EM)	7	1	CEM71	119	H	Paddock (PD)	2	3	HPD23
41	C	Empangeni (EM)	7	2	CEM72	120	H	Paddock (PD)	2	4	HPD24
42	C	Hibberdene (HB)	1	0	CHB10	121	M	Bruyns Hill (BH)	1	0	MBH10
43	C	Hibberdene (HB)	1	1	CHB11	122	M	Bruyns Hill (BH)	1	1	MBH11
44	C	Hibberdene (HB)	1	2	CHB12	123	M	Bruyns Hill (BH)	1	2	MBH12
45	C	Hibberdene (HB)	1	3	CHB13	124	M	Bruyns Hill (BH)	2	0	MBH20
46	C	Scottburgh (SC)	1	0	CSC10	125	M	Bruyns Hill (BH)	2	1	MBH21
47	C	Scottburgh (SC)	1	1	CSC11	126	M	Bruyns Hill (BH)	2	2	MBH22
48	C	Scottburgh (SC)	1	2	CSC12	127	M	Bruyns Hill (BH)	3	0	MBH30
49	C	Scottburgh (SC)	1	3	CSC13	128	M	Bruyns Hill (BH)	3	1	MBH31
50	C	Scottburgh (SC)	1	4	CSC14	129	M	Bruyns Hill (BH)	3	2	MBH32
51	C	Scottburgh (SC)	1	5	CSC15	130	M	Bruyns Hill (BH)	4	0	MBH40
52	C	Scottburgh (SC)	1	6	CSC16	131	M	Bruyns Hill (BH)	4	1	MBH41
53	C	Scottburgh (SC)	2	1	CSC21	132	M	Bruyns Hill (BH)	4	2	MBH42
54	C	Scottburgh (SC)	2	2	CSC22	133	M	Dalton (DL)	1	1	MDL11
55	C	Scottburgh (SC)	2	3	CSC23	134	M	Dalton (DL)	1	2	MDL12
56	C	Scottburgh (SC)	2	4	CSC24	135	M	Eston (ES)	1	0	MES10
57	C	Umfolozi (UF)	1	0	CUF10	136	M	Eston (ES)	1	1	MES11
58	C	Umfolozi (UF)	1	1	CUF11	137	M	Glenside (GS)	1	1	MGS11
59	C	Umfolozi (UF)	1	2	CUF12	138	M	Glenside (GS)	1	2	MGS12
60	C	Umfolozi (UF)	1	3	CUF13	139	M	Glenside (GS)	1	3	MGS13
61	C	Umfolozi (UF)	1	4	CUF14	140	M	Glenside (GS)	2	0	MGS20
62	C	Umfolozi (UF)	1	5	CUF15	141	M	Glenside (GS)	2	1	MGS21
63	C	Umfolozi (UF)	2	0	CUF20	142	M	Glenside (GS)	3	0	MGS30
64	C	Umfolozi (UF)	3	0	CUF30	143	M	Glenside (GS)	3	1	MGS31
65	C	Umfolozi (UF)	3	1	CUF31	144	M	Glenside (GS)	3	2	MGS32
66	C	Umfolozi (UF)	3	2	CUF32	145	M	Mid Illovo (MI)	1	0	MMI10
67	C	Umfolozi (UF)	3	3	CUF33	146	M	Mid Illovo (MI)	1	1	MMI11
68	C	Umfolozi (UF)	4	0	CUF40	147	M	Mid Illovo (MI)	1	3	MMI13
69	C	Umfolozi (UF)	4	1	CUF41	148	M	York (YK)	1	0	MYK10
70	C	Umfolozi (UF)	4	2	CUF42	149	M	York (YK)	1	1	MYK11
71	C	Umfolozi (UF)	4	3	CUF43	150	M	York (YK)	1	2	MYK12
72	C	Verulam (VM)	1	0	CVM10	151	M	York (YK)	2	0	MYK20
73	C	Verulam (VM)	1	1	CVM11	152	M	York (YK)	2	1	MYK21
74	C	Verulam (VM)	1	2	CVM12	153	M	York (YK)	2	2	MYK22
75	C	Zinkwazi (ZK)	1	0	CZK10						
76	C	Zinkwazi (ZK)	1	1	CZK11						
77	C	Zinkwazi (ZK)	1	2	CZK12						
78	C	Zinkwazi (ZK)	1	3	CZK13						
79	C	Zinkwazi (ZK)	1	4	CZK14						

3.3.2 Variance components analysis

To evaluate the variance contributions of different components, the following random effect model was fitted to the plot data:

$$Y_{ijkl} = \mu + T_j + B(T)_{l(j)} + G_i + GT_{ij} + GB(T)_{il(j)} + R_k + RT_{jk} + RB(T)_{kl(j)} + RG_{ik} + RGT_{ijk} + e$$

where Y_{ijkl} is the phenotypic performance of observation l of genotype i in ratoon k of trial j, μ is the grand mean, T_j is the trial main effect, $B(T)_{l(j)}$ is the effect of block within trial, G_i is the genotype main effect, GT_{ij} is the effect of the genotype x trial interaction, $GB(T)_{il(j)}$ is the

effect of genotype x block (within trial), R_k is the main effect of ratoon, RT_{jk} is the effect of the ratoon x trial interaction, $RB(T)_{kl(j)}$ is the effect of the ratoon x block (within trial) interaction, RG_{ik} is the effect of the ratoon x genotype interaction, RGT_{ijk} is the effect of the genotype x trial x ratoon interaction, and e is the error. Restricted maximum likelihood (Patterson and Thompson, 1971) using the sparse Average Information Algorithm (Gilmour et al. 1995) was used to estimate the variance components, as implemented in GenStat 12.0 (Anon, 2009).

3.3.3 GGE biplot analysis

The two-way tables of variety x environment data for TCANE, ERC and TERC were analysed using GGE biplot software (Yan, 2001). The GGE biplot analysis produces biplots derived from principal components analysis of the environment-centered data (data minus the environment means), which therefore represents the genotype main effect and the G x E interaction. The genotypes and environments are represented by points on a two-dimensional plot of principal component (PC) scores (conventionally PC1 and PC2). The distance of an environment from the biplot origin is a measure of its ability to discriminate between genotypes, the distance between two environments measures their dissimilarity in discriminating the genotypes, and the angle between the environments represents their correlation. Acute angles represent positive correlations, obtuse angles represent negative correlations, and right angles represent no correlations between environments (Yan and Tinker, 2005). In general, the proximity of environments on the biplots is an indication of their similarity, and the proximity of genotypes to environments is an indication of the degree of positive interaction with the environment. The GGE biplot provides a range of viewing options to investigate relationships between environments and genotypes, identify mega-environments, examine the representativeness of test environments as selection sites, rank genotypes based on performance in single environments, and much more. Extensive descriptions of user options are given by Yan and Tinker (2005) and some of these will be described later.

3.4 Results

3.4.1 Variance components

The model is defined to split all variance components so that they are separated as much as possible, thereby allowing for an evaluation of the variance components of interest. In this model, terms such as 'ratoon x block (within trial)' and 'genotype x block (within trial)' are

not of interest, as there would be no need to interpret replicate effects in a trial. These terms are simply included to remove the variability in the trial associated with the design.

The estimates of variance components are presented in Table 3.2. The largest proportion of the total variation in TERC (47%) was accounted for by the main effect of trials followed by the environment (trial x ratoon combination) component (24%). In terms of the components related to the G x E interaction (genotype, genotype x trial, genotype x ratoon, and genotype x trial x ratoon), the genotype x trial and genotype x trial x ratoon accounted for the largest variation. This highlights the importance of trial site contribution to the G x E interaction, as well as the influence of trial site on the variation in ratoon yields. The main effect of genotype accounted for 1.2% of the total variation compared with the 7.2% accounted for by the G x E interaction (as expressed by the sum of the genotype x trial, genotype x ratoon, and genotype x trial x ratoon interactions). This suggests that only a small proportion of the total variation was due to the mean differences between genotypes, and that the G x E interaction was more pronounced. The variation accounted for by the main effect of ratoon was less than the variation accounted for by the individual ratoon interaction components. The genotype x ratoon interaction accounted for a minor proportion (0.25%) of the total variation relative to the trial x ratoon interaction (24.4%), suggesting that variation in ratoon yields is influenced to a larger extent by site differences compared with genotype differences.

The proportion of variation accounted for by the different terms for TCANE was similar to that of TERC. Most of the variation was accounted for by the main effect of trials (35%), followed by the trial x ratoon interaction component (33%). The G x E interaction (as expressed by the sum of the genotype interaction variance components) accounted for more variation (6.7%) than the main effect of genotype (4.2%), suggesting that TCANE, together with TERC, is influenced to a larger extent by G x E interactions than by genotypic main effects. However, when only considering terms related to the G x E interaction (genotype, genotype x trial, genotype x ratoon, and genotype x trial x ratoon), the genotype main effect accounted for most of the variation, followed by the genotype x trial, genotype x trial x ratoon, and genotype x ratoon interactions, respectively. The small variation accounted for by the genotype x ratoon interactions relative to the genotype x trial interaction once again highlights the greater influence of site factors than genotype on ratoon yield variation. The strong genotype x trial x ratoon interaction suggests that cultivar cane yields and their ratooning ability were trial specific.

Similarly, for ERC, the largest proportion of the variation was accounted for by the main effect of trials (44.0%) followed by the trial x ratoon interaction (i.e. environment main effect; 28.6%). In contrast to TCANE and TERC, the main effect of genotype accounted for a higher proportion (5.5%) of the variation than the combined G x E interaction components (5.1%). This suggests that ERC is more strongly influenced by genetic composition than G x E interactions. Once again, the variation accounted for by the genotype x ratoon interaction was minor (0.1%) compared with the contribution of the trial x ratoon interaction (28.0%). When only considering terms related to the G x E interaction (genotype, genotype x trial, genotype x ratoon, and genotype x trial x ratoon), the genotype main effect was once again the largest. This was followed by the genotype x trial x ratoon effects, indicating that genotype variability for ERC across ratoons was largely controlled by trial location. The genotype x trial component was the third largest, followed by the genotype x ratoon interaction. This suggests that ERC among genotypes was least influenced by ratoons.

3.4.2 Biplot analysis

3.4.2.1 Mega-environment analysis

Figure 3.1 represents the polygon view of GGE biplots for TCANE, ERC and TERC based on the two-way table of G x E means. The polygon is drawn on vertex genotypes that were the most responsive (i.e. the best or poorest performers at some or all of the environments). Perpendicular lines are drawn from each side of the polygon to the origin, such that the biplot is divided into sectors representing different mega-environments. Environments within a ME sector have the same effects on variety performance and should be considered as a homogenous group.

For TERC, the biplot explained a total of 42% of the total variation of the two-way data matrix (Figure 3.1a), highlighting the complexity of the G x E structure in the dataset. The TERC biplot showed four distinct mega-environments represented by the N31, N12, N29, and N36 sectors, respectively. The majority of the midlands (M) environments were located within the N31 and N12 sectors, while the N29 and N36 sectors contained both coastal (C) and hinterland (H) environments. The seven M environments that were separated from the larger M cluster were actually from trials conducted in frost pockets and harvested on a 12-month cutting cycle compared with the conventional 18-24 month cycle. In general, the ratoons within trials clustered together strongly, followed by a clear separation of the M

environments from C and H environments. The MEs identified here are not in keeping with the geographical regional divisions utilised currently for the rainfed parts of the industry (midlands, coast, hinterland). For example, if the current regional separations were appropriate, the biplot would consist of three MEs corresponding to the C, H and M regions. The MEs identified here may not necessarily group according to any geographical zoning, and attempts should be made to identify the main environmental factors characterising these MEs.

Table 3.2 Variance components estimates and percentage of total phenotypic variance for tons cane/ha (TCANE), estimated recoverable crystal percentage (ERC %), and tons estimated recoverable crystal/ha (TERC). Standard errors of estimates are in brackets

	TERC		TCANE		ERC%	
	Variance	**%**	**Variance**	**%**[*]	**Variance**	**%**
Trial	8.41 (2.22)	47.34	312 (92.4)	35.0	2.54 (0.69)	44.2
Block (within trial)	0.28 (0.06)	1.60	20.2 (4.4)	2.3	0.03 (0.009)	0.5
Genotype	0.22 (0.22)	1.24	37.5 (15.8)	4.2	0.31 (0.12)	5.5
Genotype x trial	0.62 (0.61)	3.48	33.5 (5.7)	3.8	0.11 (0.02)	1.9
Genotype x block (within trial)	0.79 (0.79)	4.47	53.4 (4)	6.0	0.02 (0.01)	0.3
Ratoon	0.08 (0.08)	0.47	0	0	0	0
Ratoon x trial	4.34 (0.64)	24.43	294 (42.6)	32.9	1.64 (0.23)	28.6
Ratoon x block (within trial)	0.23 (0.03)	1.30	14.2 (2.1)	1.6	0.04 (0.01)	0.7
Genotype x ratoon	0.05 (0.02)	0.25	2.6 (1.2)	0.3	0.01 (0.007)	0.1
Genotype x trial x ratoon	0.61 (0.06)	3.45	23.2 (2.7)	2.6	0.18 (0.02)	3.1
Error	2.13 (0.05)	11.98	104 (2.9)	11.7	0.87 (0.02)	15.2

*Percentge of total phenotypic variance

The TCANE biplot (Figure 3.1b) also accounted for approximately 42% of the variation in the G x E table of means and was very similar to the biplot produced for TERC. However, this biplot exhibited three MEs instead of four, as the N12 ME observed in the TERC biplot now fell within the N31 sector. The majority of the M environments fell within the N31 sector, which also contained some C and H environments. This implies that N31 was the highest yielding cultivar in the environments grouped into that ME. The N36 sector contained both C and H environments, as well as four M environments that were different from the other M environments. The N29 sector contained C and H environments only.
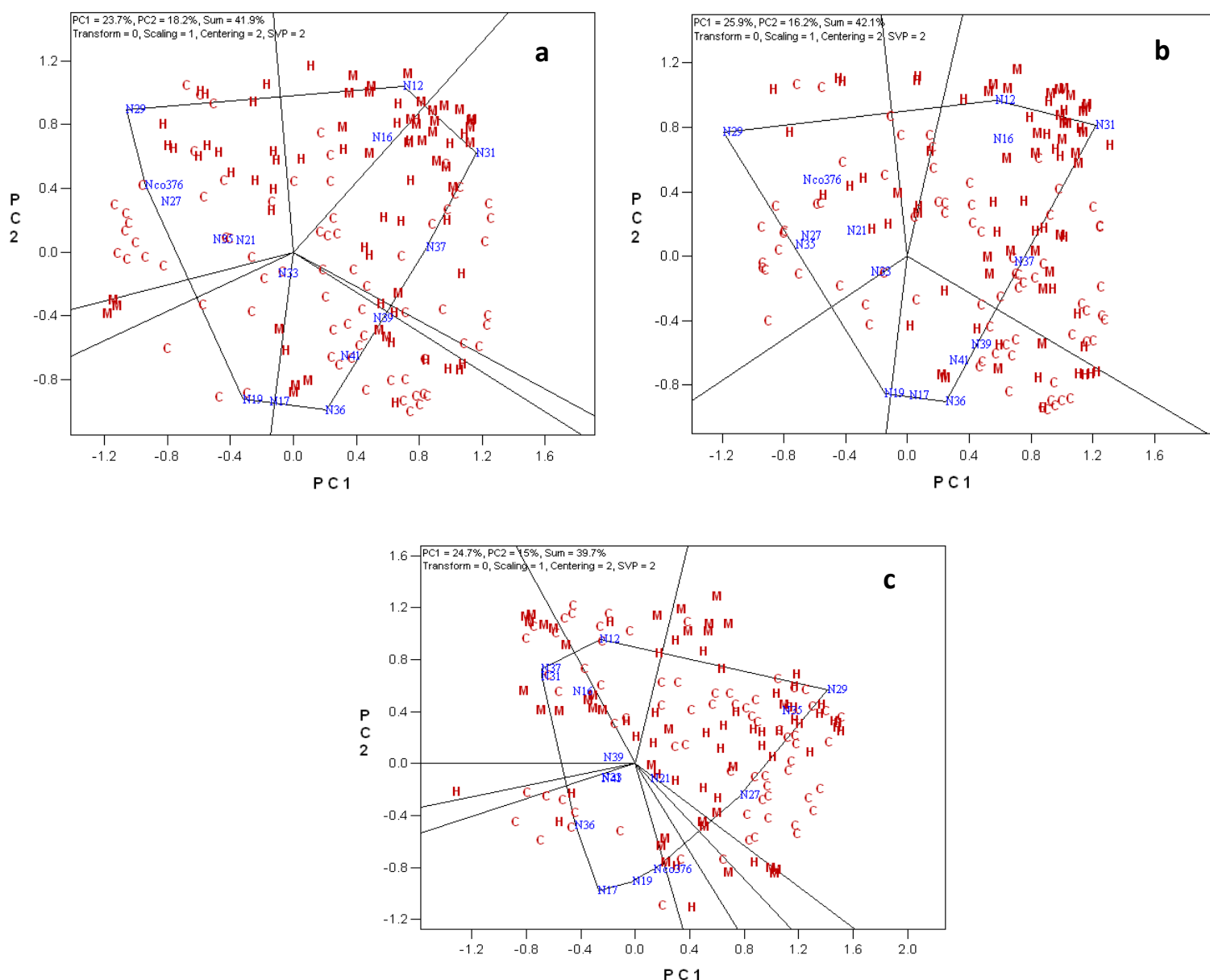
Figure 3.1 Polygon views of GGE biplots for TERC (a), TCANE (b) and ERC (c) based on data of 15 genotypes tested accross 153 environments. For better visualization of regional overlapping and separation, the environments are abbreviated by their first letters only (in red), while the genotypes are indicated in blue

The ERC biplot, which accounted for 40% of the variation in the G x E table showed that the majority of the C and H environments fell within a ME characterized by the vertex genotype N29 (Figure 3.1c). This genotype was released to the industry for its characteristically high ERC%. The M environments were scattered throughout the biplot, forming indistinct clusters. However, the vertex genotype N37 formed a ME composed primarily of M environments in

the top left hand corner of the biplot. The formation of a dominant ME for ERC may be linked to the larger ratio of G to G x E observed from variance components analyses (Table 3.2), which indicates that ERC is influenced to a larger extent by genetic effects than by the G x E interaction effects.

Although environments tended to cluster according to the ratoons harvested within a trial, frequently, a given trial showed diverse responses between ratoons. These patterns are illustrated in Figure 3.2, which shows the biplot for TERC (same as Figure 3.1a, but with full names of the environments indicated). Environments CEM30, CEM32, CEM34 and CEM36 were located within the N31 ME while the other ratoons from this site (CEM31, CEM33, CEM35) were located in different MEs. Genetic correlations between these environments confirmed that CEM30, CEM32, CEM34 and CEM36 were positively correlated while CEM31, CEM33 and CEM35 were negatively correlated to the former ratoons (Table 3.3). Similarly, environments HES10, HES11, and HES12 formed part of the N29 ME, while the next crop from this trial (HES13) was located in the N36 ME. Table 3.3 data confirms the negative correlations between HES13 and the other ratoons from that particular trial. Other examples may be identified through further interrogation of Figure 3.2. Similar observations were also made in the TCANE and ERC biplots (not shown). This differential grouping of ratoons from a trial site demonstrates the trial x ratoon interaction, which was also observed to be highly influential from the variance components analysis (Table 3.2). In addition to differential responses of ratoons within trials, it was also frequently observed that trials from a particular location grouped into different mega-environments (Figure 3.2). For example, ratoons from CAK1, CAK2 and CAK3 were located in different ME sectors. Table 3.3 data shows positive correlations between ratoons of CAK1 and CAK2, while correlations between those trials and ratoons of CAK3 were negative.
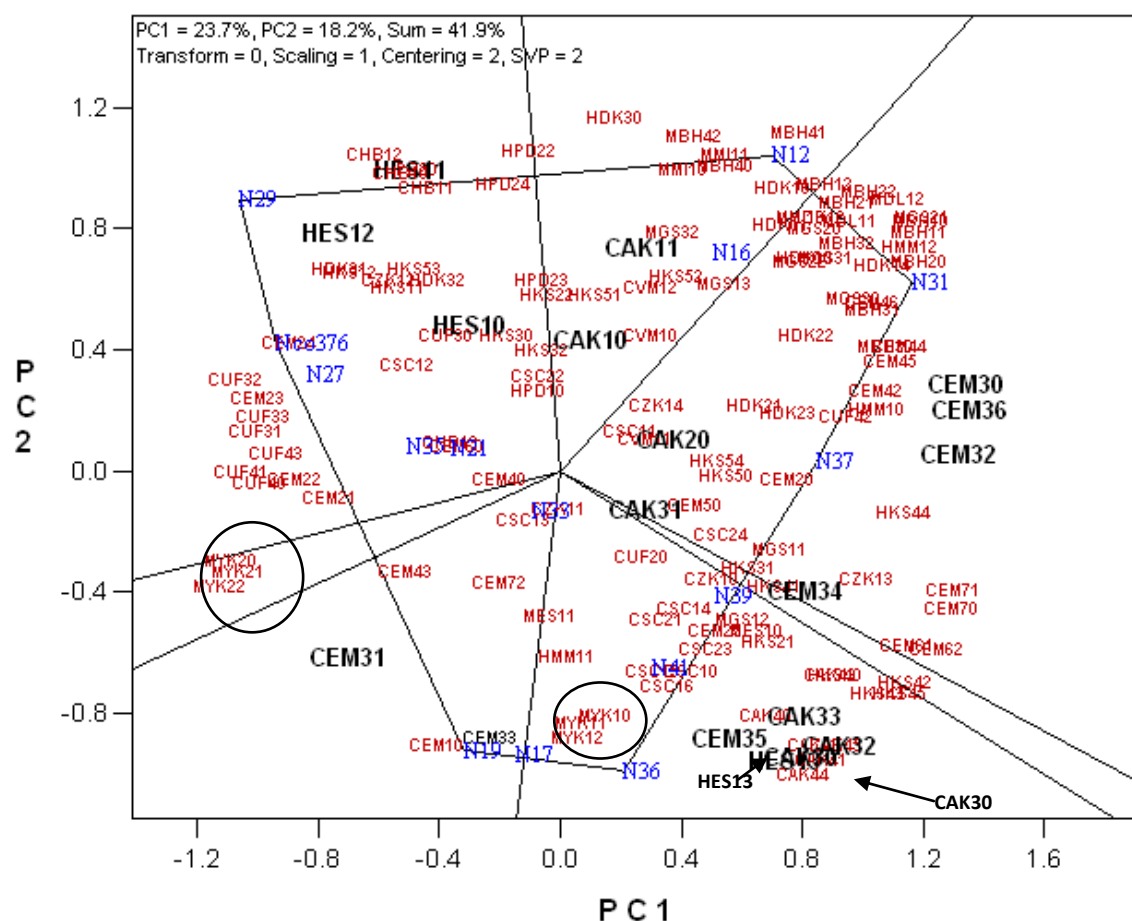
Figure 3.2 Polygon view of a GGE biplot for TERC based on data of 15 genotypes tested over 153 environments. Environments are shown in red and genotypes are shown in blue. Environments cited in text are shown in bold, black font. The six M environments are circled

Table 3.3 Phenotypic correlations between selected environments showing the variable relationships between ratoons of the same trials and between trials within the same location

	CEM30	CEM31	CEM32	CEM33	CEM34	CEM35	CEM36
CEM30	1						
CEM31	-0.75**	1					
CEM32	0.91**	-0.46	1				
CEM33	-0.56*	0.79**	-0.22	1			
CEM34	0.37	0.11	0.61*	0.23	1		
CEM35	-0.04	0.44	0.24	0.61*	0.68**	1	
CEM36	0.98**	-0.65**	0.96**	-0.44	0.47	0.11	1
	HES10	HES11	HES12	HES13			
HES10	1						
HES11	0.61*	1					
HES12	0.32	0.73**	1				
HES13	-0.39	-0.75**	-0.52*	1			
	CAK10	CAK11	CAK20	CAK30	CAK31	CAK32	
CAK10	1						
CAK11	0.54*	1					
CAK20	0.79**	0.55*	1				
CAK30	-0.27	-0.55*	-0.01	1			
CAK31	-0.76**	-0.41	-0.72**	0.22	1		
CAK32	-0.51*	-0.57*	-0.25	0.91**	0.51*	1	
CAK33	-0.59*	-0.51*	-0.31	0.84**	0.49*	0.89**	

*P≤0.05, **P≤0.01

3.4.2.2 Genotype performance and stability

Figure 3.3a is a different view of the TERC biplot highlighting the average performance and stability of the different genotypes. In this view, the average TERC of the genotypes are indicated by the projections of their markers onto a derived axis (red line with single arrowhead). When perpendicular lines are drawn from the genotype markers to this axis, the TERC rankings of genotypes can be seen. When comparing the biplot results with the mean TERC of the genotypes, the biplot approximated the ranking of the genotypes very well (correlation = 0.78). Thus, genotype N31 had the highest mean TERC, followed by N12, and then N16, while N19 had the lowest mean TERC. The stability of the genotypes is measured by the projections of their markers on the other derived axis (blue line with double arrowhead). Perpendicular lines are drawn from the genotype markers to this axis and the points at which they intersect this axis then define their stability. The greater the absolute

distance of the intersections from the origin, the less stable a genotype is. Thus, genotypes like N33, N31 and N21 were the most stable, while genotypes such as N29, N36 and N17 were the most responsive. The center of the concentric circles in the biplot represents the location of a hypothetical 'ideal' genotype (i.e. highest yield and highest stability). Genotype N31 was located closest to this point, suggesting that it may be the ideal genotype to utilize as a control when selecting for TERC in selection programs.

Performance and stability of genotypes based on TCANE can be seen in Figure 3.3b. The biplot and genotype rankings are very similar to the TERC biplot, with genotype N31 showing the highest TCANE and exhibiting stability, i.e closest to the ideal genotype. The correlation between biplot ranks and genotype means was 0.88, suggesting that the projections of the genotype markers were a good approximation of the average genotype performance. Once again, N31 was located closest to the ideal genotype, suggesting that it would be a good candidate as a control during selection for TCANE.

Regarding ERC (Figure 3.3c), genotype N29 was ranked as the top performer, followed by N35 and N27, while the lowest ranked genotype was N17 (correlation between biplot ranks and means was 0.88). Genotypes N41, N36, N39 and N21 were the most stable for ERC. Genotype N31, which was stable in terms of TCANE (Figure 3.3a), was responsive in terms of ERC. In contrast, genotype N36 was responsive for TCANE (Figure 3.3a) and stable for ERC (Figure 3.3b). Genotypes N33 and N21 were stable for both TCANE and ERC; however, their mean performance was not impressive for both traits. Genotype N29 was the closest to the ideal genotype for ERC, suggesting that it would be a good candidate for a control genotype when selecting specifically for ERC.
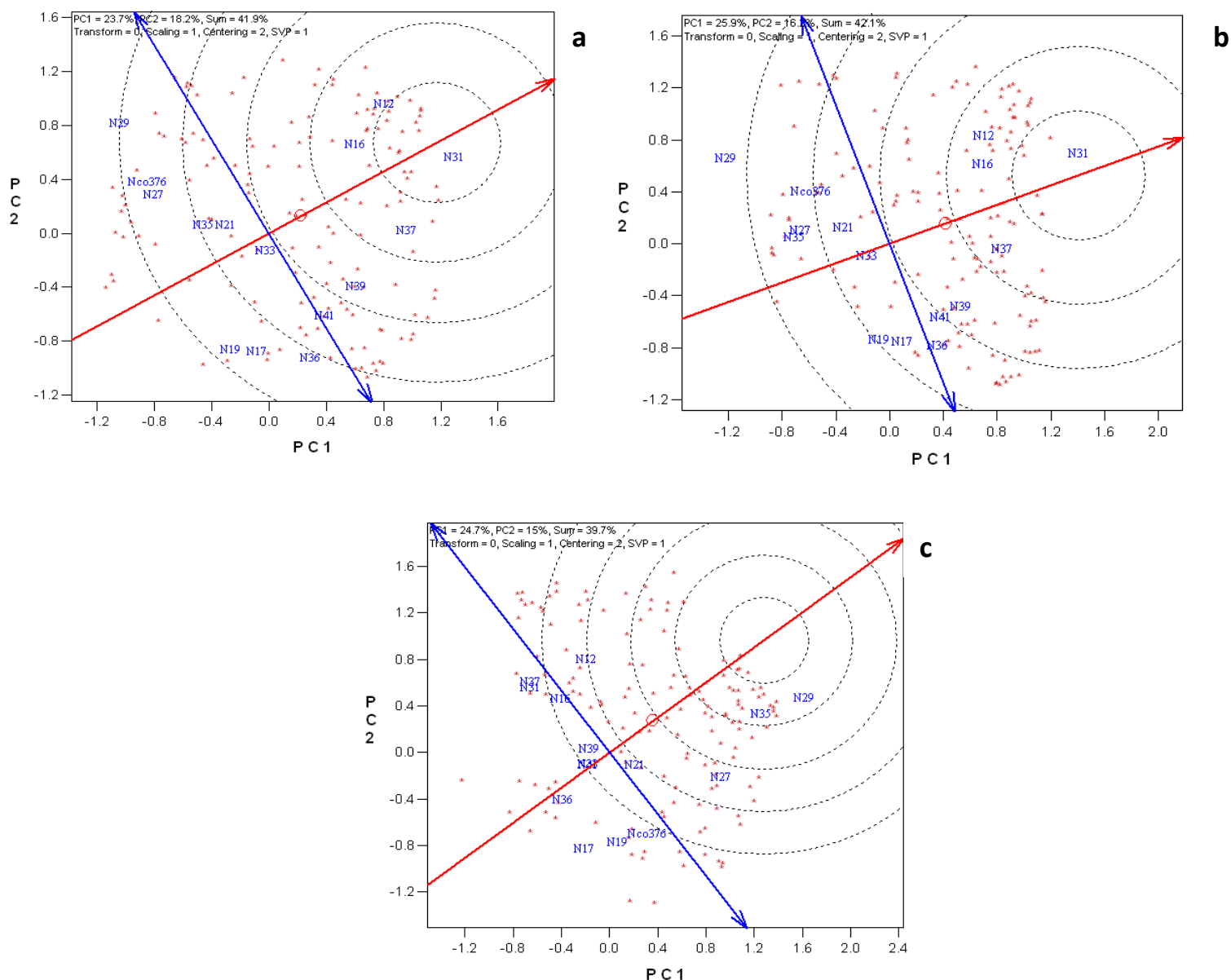
Figure 3.3 The 'means vs. stability' view of GGE biplots for TERC (a), TCANE (b), and ERC (c) based on data from 15 genotypes tested over 153 environments. Environments are indicated by an asterisk, while genotypes are indicated by their names

3.5 Discussion

The results of this study have provided valuable insights into the nature of the G x E interactions characterising the rainfed parts of the industry. The large component of variation accounted for by the genotype x trial interaction relative to the genotype x ratoon interaction highlights that there may not be much value in extending trials to longer crops. It follows

51

that, to increase responses to selection, greater emphasis should be placed on sampling more trial sites than on testing ratoons within a trial site. This confirms results obtained in other studies on sugarcane in Australia (Jackson et al., 1991; Mirzawan et al., 1994). This trend was also evident in the biplots where, in general, environments that clustered close together were ratoons of the same trial. However, frequently, ratoons of the same trial also showed diverse responses (Figure 3.2), showing that seasonal/ratoon differences could have an effect on the categorization of a particular environment into a ME. The grouping of ratoons from a trial into different ME sectors suggests the existence of influential climatic (soil characteristics remain constant), biotic, or management factors that can produce genotypic rank changes between ratoons. Further analyses of the factors characterizing such environments may help explain the reasons for these differential ratoon responses within trials. The existence of 'subsets of environments' or 'environment types' that are characterized by similar biotic and abiotic stresses may be a possibility, and these should be investigated for future improvements of the trial networks.

The occasional differential grouping of trials from the same location into different MEs suggests rapid changes in the values of relevant environmental factors for environments that were geographically close. This may be due to the use of different experimental fields within a location, where factors such as cropping history, management practices (the majority of trials were conducted on commercial farms), and variations in soil physical and chemical properties brought about by topographic position have large confounding effects. These are further complicated by the effects of different times of harvest and harvest ages, thereby giving rise to large and complex G x E interactions that can only be interpreted through proper characterization of trial sites. Other studies have also addressed the issue of such genotype x management interactions for grain yield of wheat (Cooper et al., 2001). Furthermore, in this study the effect of season (year) *per se* was not taken into consideration, and was largely analysed through the effects of the ratoon factor, with which it was confounded. The thinking behind this was that crops of the same ratoon age (i.e 1st, 2nd,3rd ratoon etc) that were harvested in a particular season, would have probably started at different times of the year and would be at different ages, thereby making the interpretation of the season effect very difficult. This confounding effect of ratoon with season is the reason for the large trial x ratoon interaction observed with all three traits (Table 3.2).

Biplot analysis has shown that selection strategies for TERC in the rainfed parts of the sugar industry should be targeted at four MEs, which do not necessarily conform to the traditional regional sub-divisions currently in use. Mega-environments consisted of a mix of C, H, and M environments, which discriminated genotypes similarly, implying that certain environments, although geographically distinct, had a similar effect on genotypic responses. This was particularly true of environments belonging to the C and H regions, as the M environments tended to cluster fairly tightly, with only few outliers. The six M environments that were outliers (Figure 3.2) were actually trials conducted in frost pockets in the midlands that were harvested on a 12-month cutting cycle as opposed to the conventional 18-24 month cycle for that region. This suggests a possible dominant role of harvest age in determining environmental groupings. Further analysis of the biplot also showed that environments representative of the five selection programmes (coastal high potential, coastal low potential, hinterland, midlands humic, midlands sandy) did not cluster together to form five separate mega-environments. This may imply the need for possible adaptation of the current selection strategy to target the MEs identified in this study. Such adaptations may only be possible after detailed characterizations of the MEs, together with adequate demonstration of the repeatability of the observed responses.

In addition to the implications on selection, the results of this study have provided insight into the redundancy of trial sites for the purposes of post-release evaluation. Environments tightly clustered within a ME should be similar with regard to genotypic performance (Yan and Tinker, 2005), suggesting that one or more environments could be removed without much loss of information. In this study, a range of M environments clustered tightly within the N31 sector in Figure 3.1a. These environments included those representing the 'sandy' and 'humic' conditions, which are currently evaluated in the midlands and for which two separate selection programs currently exist. The appropriateness of the sandy vs. humic subdivision in the midlands has never been tested and the results of this study suggest the need for such an evaluation to optimize resources and improve the efficiencies of both selection and evaluation (this will be addressed in Chapter 5). Another example is the tight clustering of ratoons from the trials CUF3, CUF4 and CEM3 (see the left hand side of Figure 3.2), which suggests that at least one of those trial sites could be removed for post-release evaluation purposes.

The analysis of genotype performance and stability identified N31 as the ideal genotype for TERC and TCANE, while N29 was identified as the ideal genotype for ERC. These results

suggest that these genotypes may be ideal candidates for controls in selection trials to improve genetic gains. Currently, however, genotype NCo376 is used as a control in all plant breeding selection trials. The poor relative performance (TERC, TCANE and ERC) and stability of NCo376 is an indication of successful selection gains achieved through the release of newer genotypes. In order to sustain future gains, the use of alternate, higher yielding and stable controls such as N31 must be considered. The variance components ratio of genotype to G x E was higher for ERC than for TERC and TCANE, suggesting that gains from selection are likely to be more rapid for ERC relative to the other two traits. This has indeed been the case in the South African sugar industry, where the average ERC of new commercial releases have increased substantially relative to TCANE. Additionally, adoption of new cultivars in the South African sugar industry has not occurred at the same rate as cultivar release, leading to large areas dominated by single (older) cultivars such as NCo376. The results of this study demonstrate the potential benefits of adopting newer cultivars.

This study has highlighted a number of issues deserving of consideration for future G x E studies in the industry. However, it is acknowledged that this study is characterized by many shortcomings due primarily to the nature of the data used. Only highly selected released cultivars were analysed, thereby reducing the potential genotypic variability in the dataset. This was due to the lack of balanced datasets across the different post-release trials. Only a few plant breeding selection trials were included, based on the commercial controls which were common to the post-release trials. Future studies should consider the use of a mixed model approach, with the inclusion of "region" as an additional factor in the model. Despite these limitations, the study provided useful insights into the complexity of G x E interactions in the industry, which is dealt with in subsequent chapters.

3.6 Conclusions

This study was the first to evaluate the variance components and MEs associated with METs conducted in the rainfed regions of the South African sugar industry. The results have demonstrated the existence of large and complex G x E interactions that may be exploited for improved selection and evaluation. Important findings of the study that must be confirmed with further detailed investigations were:

- The identification of larger trial x ratoon interactions relative to genotype x ratoon interactions, which have direct impact on the design of the selection and evaluation network.
- The existence of four MEs that do not correspond to the current regional sub-divisions used in the rainfed parts of the industry.
- A discrepancy between the current five selection programs and the target environments comprising the rainfed regions.
- The existence of differential ME groupings of ratoons within trials and trials within locations, highlighting the need to characterize trials according to environmental factors.
- The evaluation of relative performance and stability of genotypes for use as controls when selecting for different yield traits.

Future studies should involve a complete characterization of environments used for selection relative to climatic variables, soil characteristics, and management/crop factors (age and time of harvest). Such studies will also allow for the analytical interpretation of G x E interactions which has implications for developing more targeted breeding programmes, better choice of environments for selection and evaluation, and will provide insight into the genetic and physiological makeup of genotypes for further physiological studies.

3.7 References

Anonymous., 2009. GenStat® Executable Release 12.1. Lawes Agricultural Trust, Rothamstead Experimetnal Station, Harpenden, UK. Clarendon Press, London, UK.

Ceretta, S., Van Eeuwijk, F.A., 2008. Grain yield variation in malting barley cultivars in Uruguay and its consequences for the design of a trials network. *Crop Science* 48, 167-180.

Cooper, M., Woodruff, D.R., Phillips, I.G., Basford, K.E., Gilmour, A.R., 2001. Genotype-by-mangement interactions for grain yield and grain protein concentration of wheat. *Field Crops Research* 69, 47-67.

Dehghani, H., Ebadi, A., Yousefi, A., 2006. Biplot analysis of genotype by environment interaction for barley yield in Iran. *Agronomy Journal* 98, 388-393.

DeLacy, I.H., Kaul, S., Rana, B.S., Cooper, M., 2010. Genotypic variation for grain and stover yield of dryland (rabi) sorghum in India: 1. Magnitude of genotype x environment interactions. *Field Crops Research*. 118, 228-235.

De la Vega, A.J., Chapman, S.C., 2010. Mega-environment differences affecting genetic progress for yield and relative value of component traits. *Crop Science* 50, 574-583.

Finlay, K.W., Wilkinson, G.N., 1963. The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research* 14, 742-754.

Gauch, H.G. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam.

Gauch, H.G., Zobel, R.W., 1997. Identifying mega-environments and targeting genotypes. *Crop Science* 37, 311-326.

Gilmour, A.R., Thompson, R., Cullis, B.R., 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51, 1440-1450.

Glaz, B., Kang, M.S., 2008. Location contributions determined via GGE biplot analysis of multienvironment sugarcane genotype-performance trials. *Crop Science* 48, 941-950.

Jackson, P.A., Horsley, D., Foreman, J., Hogarth, D.M., Wood, A.W., 1991. Genotype x environment (GE) interactions in sugarcane variety trials in the Herbert. *Proceedings of the Australian Society of Sugarcane Technologists* 13, 103-109.

Mirzawan, P.D.N., Cooper, M., DeLacy, I.H., Hogarth, D.M., 1994. Retrospective analysis of the relationship among the test environments of the southern Queensland sugarcane breeding programme. *Theoretical and Applied Genetics* 88, 707-716.

Nuss, K.J., 1998. Aspects considered in the search for new farms for the experiment station. *Proceedings of the South African Sugar Technologists Association* 72, 42-45.

Patterson, H.D., Thompson, R., 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545-554.

Queme, J.L., Orozco, H., Ovalle, W., Soto, G., Melgar, M., 2001. AMMI analysis of sugarcane cultivar trials in Guatemala. *Proceedings of the International Society of Sugar Cane Technologists* 24, 512-515.

Sharma, R.C., Morgounov, A.I., Braun, H.J., Akin, B., Keser, M., Bedoshvili, D., Bagei, A., Martius, C., Van Ginkel M., 2010. Identifying high yielding stable winter wheat genotypes for irrigated environments in central and west Asia. *Euphytica* 171, 53-64.

Shukla, G.K., 1972. Some statiscal aspects of partitioning genotype-environment components of variability. *Heredity* 29, 237-245.

Smith, A.B., Cullis, B.R., 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.

Yan, W., 2001. GGEbiplot – A Windows application for graphical analysis of multienvironment trial data and other types of two-way data. *Agronomy Journal* 93, 1111-1118.

Yan, W., Hunt, L.A., 2001. Interpretation of genotype x environment interaction for winter wheat in Ontario. *Crop Science* 41, 19-25.

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

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

Yan, W., Tinker, N.A., 2005. Biplot analysis of multi-environment trial data: Principles and applications. *Canadian Journal of Plant Science* 86, 623-645.

CHAPTER 4

INTERPRETATION OF GENOTYPE X ENVIRONMENT INTERACTIONS OF SUGARCANE: IDENTIFYING SIGNIFICANT ENVIRONMENTAL FACTORS[2]

4.1 Abstract

An understanding of the causes of G x E interactions is essential for the implementation of efficient selection and evaluation networks. Currently, studies involving the interpretation of sugarcane G x E interactions are limited. The objective of this study was to investigate the relative influence of environmental factors on the G x E interactions in sugarcane under rainfed conditions in South Africa through a comprehensive analysis of a MET dataset. Fifteen commercial cultivars were evaluated across 147 environments (trial x ratoon combinations) across the coastal (C), hinterland (H) and midlands (M) regions of the sugar industry. Environments were characterized according to five site covariates (soil depth, clay percentage, organic matter percentage, nitrogen mineralization category, and total available moisture) and nine seasonal covariates (time of harvest, age at harvest, average daily heat units, solar radiation, rainfall, evaporation, and three derived water stress indices).

AMMI2 biplots for TCANE, ERC and TERC revealed overlapping of C and H environments, while M environments formed unique clusters characterized by specific cultivar adaptabilities. The PCA allowed visualization of the covariates determining the regional separation patterns. The AMMI interaction principal components axes (IPCA) 1 and 2 scores were correlated to the covariates and showed that harvest age, temperature, and water stress were mainly responsible for separation of M environments from C and H environments on the TCANE and TERC biplots. Time of harvest was identified as an important covariate influencing ERC G x E patterns in the C and H regions. The third water stress index (based on a ratio of observed yields to simulated irrigated yields) was a dominant factor influencing G x E patterns within the C and H regions and was identified as a superior indicator of water deficient environments for future studies. The M trials were characterized by shallower soils with lower total available moisture and greater variability in this regard compared with the C

---

[2]Ramburan, S., Zhou, M., Labuschagne, M., 2011. Interpretation of genotype x environment interactions of sugarcane: Identifying significant environmental factors. *Field Crops Research* 124, 392-399.

and H trials. Nitrogen mineralization category, organic matter percent, and clay percent were not significantly correlated to IPCA scores, while soil depth was identified as a major site selection criterion in the M region. The M region should be treated as a single mega-environment, while the C and H regions could be combined for future interpretive studies, where covariates should be summarized within growth phases. The results of this study will assist in restructuring the MET network through exploitation and targeting of the relevant environmental factors within the different regions.

4.2 Introduction

The analysis of the MET dataset in Chapter 3 illustrated the large and complex G x E interactions in the rainfed regions of the industry. The study, which employed GGE biplot analysis (Yan et al., 2000) showed differential ME groupings of ratoons from the same trials and trials from the same locations, suggesting large spatial and temporal fluctuations in major yield determining environmental factors. That study highlighted the need to investigate the nature of such G x E interactions for exploitation within the breeding and evaluation networks.

Genotype x environment interaction has long been a barrier to efficient plant improvement through its impairment of efforts to select superior genotypes and its negative impact on heritability. For sugarcane, the majority of G x E studies have been empirical in nature, focusing mainly on quantifying G x E interactions and genotype stability (Tai et al., 1982; Queme et al., 2005; Jackson et al., 2007), identifying mega-environments (Jackson et al., 1991; Mirzawan et al., 1994; Queme et al., 2007), and quantifying sources of variation for resource allocation (Bissessur et al., 2010, Rattey and Kimbeng, 2001). In comparison, fewer sugarcane studies have focused on the interpretation of G x E interactions to understand the nature and causes thereof. Such an understanding may benefit sugarcane improvement initiatives and allow for the exploitation of G x E interactions rather than its avoidance or acceptance. An understanding of the factors responsible for G x E interactions can be used to establish breeding objectives, formulate recommendation domains, contribute to ideotype design, and identify ideal test conditions (Yan and Hunt, 2001).

Approaches currently used to gain an understanding of the causes of G x E interactions can be categorized into two strategies. The first involves the use of factorial regression models based on two-way G x E tables with concomitant variables which could either be environmental factors, genotypic traits, or combinations thereof (Baril et al., 1995). The

second strategy involves the correlation of genotypic or environmental scores derived from AMMI analysis to genotypic or environmental covariates (Van Eeuwijk et al., 1995). Both strategies, although different in approaches, have been shown to produce similar results (Vargas et al., 1999). With a large number of environmental covariates (that may or may not be correlated) factorial regression becomes difficult to implement.

Consequently, many studies have utilized the second strategy due to the production of PCA-based biplots, which allow for easy visualization of responses and relationships. In addition, biplots produced from AMMI analysis can be enriched with covariates to allow for easier interpretation, provided there are strong correlations between AMMI IPCA scores and the covariates. Voltas et al. (1999) used this method to identify genotypic and environmental covariates influencing G x E interactions for grain filling in barley. The three covariates identified were subsequently included into a factorial regression model to interpret G x E interactions. Van Oosterom et al. (1996) related AMMI IPCA scores to environmental covariates and found that factors such as the mean maximum temperature 10 days after flowering and the changes in a water satisfaction index during grain filling were some of the causes of G x E interactions of pearl millet. Van Eeuwijk and Elgersma (1993) found that four environmental covariates showed strong correlations to AMMI environment IPCA scores and these corresponded to the covariates identified by factorial regression, in an analysis of G x E interactions of ryegrass. De Vita et al. (2010) used this approach to investigate environmental and genotypic causes of yield variation of wheat in Italy, and identified water availability as a key factor. Vargas et al. (1999) showed that this approach produced results that were comparable with those of factorial regression and partial least squares regression in their analysis of two wheat datasets. The methodology can also be adapted to understand genetic causes of G x E interactions. Annicchiarico et al. (2010) investigated the adaptation of lupin (*Lupinus albus*) landraces by correlating AMMI IPCA scores of different germplasm pools to mean values of different morphophysiological traits.

Interpretive studies of sugarcane G x E interactions are currently limited to a single analytical study that investigated causes of G x E interactions in Australian selection trials (Jackson et al., 1995). Subsequent interpretive studies of sugarcane G x E interactions involved the straightforward characterization of METs (in a single-factor approach) according to soil type (Del Blanco et al., 2010; Glaz and Kang, 2008), age at harvest (Ramburan and Sewpersad, 2009), regional characteristics (Bissessur et al., 2010), and time of harvest (Gilbert et al., 2006). However, since Jackson et al. (1995), no further attempts have been made to

investigate the nature of sugarcane G x E interactions within a MET framework through combined analysis of the major environmental drivers of crop growth. This approach should allow for a better understanding of crop performance in different environments and can only be achieved through a detailed characterization of the soil and climatic factors associated with METs. The complexity of sugarcane G x E interactions in South Africa necessitates an investigation into the nature thereof, leading to more informed breeding objectives, better test site selection, development of appropriate recommendation domains, and sugarcane ideotype design. The objective of this study was to investigate the relative influence of different environmental factors on the G x E interactions of sugarcane under rainfed conditions through an interpretive analysis of the MET dataset from Chapter 3.

4.3 Materials and methods

4.3.1 MET dataset

The trial dataset used in this study comprised 43 trials (32 post-release cultivar evaluation trials and 11 advanced plant breeding selection trials) grown in 18 different locations, harvested during the period 1999-2009. Trial details, distribution, and abbreviations were described in detail in Chapter 3. Genotypic data for cane yield (TCANE), estimated recoverable crystal percentage (ERC) and tons ERC/ha (TERC) of the 15 cultivars most common to all trials were analyzed.

4.3.2 Environment characterization

For each trial in the dataset, stable site characteristics were determined through soil profile analyses and conventional soil sampling procedures to develop five *site covariates*. A minimum of three soil pits were dug at random points within each trial (sites where older trials were established were identified from farm records) to determine a range of characteristics. These included the effective rooting depth (Depth), clay percent (Clay), organic matter percent (OM), and N mineralization category (Ncat) (1=low, 2=medium, 3=high, 4=very high) (Anonymous, 2000). The total available moisture (TAM) for each site was subsequently estimated from Clay and Depth, from a formula developed by Van Antwerpen et al. (1994).

Apart from the stable site covariates, nine other covariates that varied seasonally (hence termed *seasonal covariates*) were also developed. For each environment (trial x ratoon combination) weather data were summarized from weather stations that were closest to the trial sites, and used to calculate the following covariates: average daily thermal time (base

10°C) in heat units (TT); average daily solar radiation (RAD) in MJ/m^2/sec, average daily rainfall (RAIN) in mm, average daily evaporation (EVP) in mm, and three water stress indices (WSI1, WSI2, and WSI3) that ranged from 0 to 1. The WSI1 was calculated by using the TAM, crop start/ratoon dates, and relevant weather station data as inputs to run the Canesim crop growth model (Singels and Donaldson, 2000), which estimated the actual crop evapotranspiration (EVT$_{act}$) for the respective crop cycle. The same model scenarios were also run for the same cycle as if fully irrigated to simulate unstressed crop growth and estimate the potential evapotranspiration (EVT$_{pot}$) had there been irrigation. The WSI1 was subsequently calculated as the ratio of EVT$_{act}$: EVT$_{pot}$, where a high WSI1 represented no moisture stress, while values closer to zero represented higher levels of stress. The second WSI (WSI2), was based on cane yields, and was calculated as a ratio of simulated rainfed cane yield: simulated irrigated cane yield. The third WSI (WSI3) was calculated as a ratio of observed trial mean yield: simulated irrigated yield. For each environment, the age at which the crop was harvested (AGE) and the time of harvest (TOH) were also recorded as additional covariates. Details of the five site and nine seasonal covariates, together with their means and ranges are indicated in Table 4.1.

Table 4.1 Descriptions of nine seasonal and five site covariates measured over 43 trials and 147 environments

Covariate	Description	Mean	Range
	Seasonal covariates		
TT	Thermal time in heat units using a base temperature of 10°C (Heat units/day)	10.78	6.74-13.97
RAD	Average daily solar radiation (MJ/m^2/sec/day)	15.77	12.99-17.89
RAIN	Average daily rainfall (mm/day)	2.56	0.98-4.57
EVP	Average daily A-pan evaporation (mm/day)	4.37	2.83-8.33
WSI1	Water stress index 1. Simulated rainfed evapotranspiration vs. simulated irrigated evapotranspiration	0.48	0.15-0.98
WSI2	Water stress index 2. Simulated rainfed cane yield vs. simulated irrigated cane yield	0.43	0.08-0.99
WSI3	Water stress index 3. Observed trial mean yield vs. simulated irrigated cane yield	0.61	0.18-1.32
AGE	Age at harvest (months)	15.69	10.72-25.97
TOH	Time of harvest (1-9). Coded as month 1 (April) to month 9 (December)	N/A	N/A
	Site covariates		
Depth	Effective rooting depth (mm)	667.44	300-1200
Clay	Clay percentage (%)	30.02	5 - 65
OM	Organic matter percentage (%)	2.16	0.2-4.54
Ncat	Nitrogen mineralization category (1-4)	N/A	N/A
TAM	Total available moisture (mm)	74.35	30-192

4.3.3 Statistical analyses

4.3.3.1 G x E interactions

The two-way tables of G x E means for TERC, TCANE and ERC were analyzed using AMMI, as described by model 5 in Chapter 2. The AMMI2 biplots were produced to visualize the G x E interactions for TERC, TCANE and ERC. The AMMI2 biplots are characterized by the projection of the genotype and environment IPCA1 and IPCA2 scores onto a two-dimensional biplot. The interpretations of G x E interactions from an AMMI2

biplot are similar to those from other forms of biplot analysis. Genotypes and environments are represented by points. The distance from the origin of the biplot to these points represents the amount of interaction that is exhibited by the respective genotype or environment. The angles between two genotype vectors correspond to their correlation. In general, acute angles represent positive correlations, right angles represent no correlation, and obtuse angles represent negative correlations (Voltas et al., 1999). Similar interpretations hold true for environments. The magnitude of interactions between genotypes and environments can be interpreted by their respective vector direction, where similar directions indicate positive interactions and vice versa (Gauch, 1992).

In order to study the correlation structures of the environments in relation to the nine seasonal covariates, a PCA was conducted on the environment x covariate two-way table. The biplot of the first two PC axes was used to assess the main differences between the environments, identify the covariates driving such differences, and to reveal relationships and redundancies between the covariates. The PCA provided an overview of the environmental factors characterizing the different regions. However, in order to investigate the effects of the environmental factors on the G x E interactions, the covariates were further correlated to the AMMI IPCA1 and IPCA2 environment scores. Significant correlations suggested that any separation of the environments on the AMMI2 biplots was attributed to the relevant covariate, thereby highlighting the importance of that covariate to the G x E interactions. Hence, the results of the correlation analysis, together with the differential grouping of environments in the biplots allowed for biological interpretations of the factors driving G x E interactions.

4.3.3.2 G x trial interactions

In the above analysis, an environment was considered a trial x ratoon combination characterized by varying seasonal covariates. Each trial, however, was also characterized by stable/fixed/immovable site covariates usually related to soil factors. Therefore, the effects of the five site covariates could not be evaluated using the above procedures, as ratoons of the same trial would be characterized by the same site covariates, resulting in overlapping of these environments on the PCA biplot. Consequently, a second approach to the analysis was adopted, where the phenotypic performance of a genotype was averaged across ratoons of a trial, and the AMMI analysis was conducted on the G x trial two-way table of means. In order to study the correlation structures of the trials in relation to environmental factors, the nine seasonal covariates (averaged across ratoons, within a trial), together with the five site

64

covariates were analyzed using PCA, as above. This PCA allowed for the comparison of the relative effects of site versus seasonal covariates on the separation of the trials used in the study. The effects of the covariates on the G x trial interactions were subsequently evaluated through correlation analysis, as above. This approach represented an evaluation of the repeatable component of G x E, while the former approach represented an analysis of a mixture of repeatable and non-repeatable G x E. All statistical analyses were conducted using Genstat® Version 12.1 statistical software (Anonymous, 2009) and biplots were produced using Canoco® for Windows Version 4.51 (Ter Braak and Smilauer, 2003).

4.4 Results

4.4.1 G x E interactions

4.4.1.1 AMMI analysis

The AMMI analysis of variance results are presented in Table 4.2. The analysis was conducted using genotypic means for each environment averaged across replicates and the pooled error was determined from the individual trial errors. The results indicate that the genotype and environment main effects were significant for all three variables. As a percentage of total sums of squares, the environment accounted for 39.9, 40.4 and 42.8% of variation, while genotype accounted for 3.7, 6.1, and 3.8% of variation in TERC, TCANE, and ERC, respectively. The G x E interaction accounted for 56.0, 53.3, and 53.1% of total variation for TERC, TCANE and ERC, respectively.

Table 4.2 Analysis of variance for cane yield (TCANE), estimated recoverable crystal (ERC), and ton ERC/ha (TERC) of 15 sugarcane genotypes evaluated in 147 environments

Souce of variation	df	TERC		TCANE		ERC	
		SS	MS	SS	MS	SS	MS
Genotype (G)	14	2647	189[**]	231862	79161[**]	920	65.7[**]
Environment (E)	146	28766	197[**]	1521952	49826[**]	10129	69.4[**]
G x E	2044	40518	16.9[**]	2007361	812[**]	12545	5.3[**]
Error (pooled)		91.6	3.2	4684	172	29.7	0.9

df Degrees of freedom, *SS* sums of squares, *MS* mean squares, [**] P<0.001

Figure 4.1a represents the AMMI2 biplot for TERC, which accounted for 52.6% of the G x E interaction and showed considerable separation of the M environments (i.e. environments in the M region) from the C and H environments. There was substantial overlap between the C

and H environments, suggesting that trials within these two regions may discriminate cultivars similarly. Isolated groups of M environments, which were located in the upper and lower left quadrant of the biplot represented trials conducted in frost pockets and harvested at 12-months of age. The C and H environments exhibited substantial deviation along the IPCA2 axis, while the M environments were spread more explicitly along the first axis. Three groups of cultivars were apparent: N12, N16, N31 and N37 grouped together and showed better adaptability to the M environments in general; N19, N17, N41, N36, N39 and N33 grouped together in the upper left quadrant and did not appear to have any specific regional affinity; and N21, N35, N27, NCo376 and N29 grouped together in the bottom left quadrant and were correlated to some of the C and H environments. The four cultivars that grouped together with the M environments are typically harvested older than 15 months of age. This suggests that the regional grouping may be due to the harvest age effect, rather than a true positive interaction.

The biplot for TCANE depicted in Figure 4.1b, which accounted for 53.8% of the G x E interaction, revealed substantial clustering of the environments closer to the biplot origin, suggesting that environments were less interactive for TCANE than for TERC (Figure 4.1a). However, the regional groupings were similar to that of the TERC biplot, with the C and H environments showing considerable overlap, separate from most of the M environments. The H and M environments were spread more explicitly along the IPCA2 axis, while the C environments showed greater separation along the first axis. Cultivar groupings were also similar to the TERC biplot.

The ERC biplot (52% of G x E interaction) in Figure 4.1c displayed considerable overlapping of the C, H, and M environments, which tended to cluster close to the biplot origin. Three ratoons from the trial in a frost pocket in the Midlands formed a cluster on the right side of the biplot. No other environments showed significant deviations along the IPCA1 axis. Four C environments, and one H environment exhibited high IPCA2 scores, however, in general there was tight clustering of environments along the IPCA2 axis as well. The cultivar groupings differed slightly from those observed in the TCANE and TERC biplots: N37, N39, N41, and N36 grouped together in the top left quadrant; N17 and N19 formed another group in the top right quadrant; NCo376, N35, N29 and N27 grouped in the bottom right quadrant; while the remaining cultivars formed a group in the bottom left quadrant. There was no clear affinity of cultivar groups to specific regions, as observed with TERC (Figure 4.1a) and TCANE (Figure 4.1b).
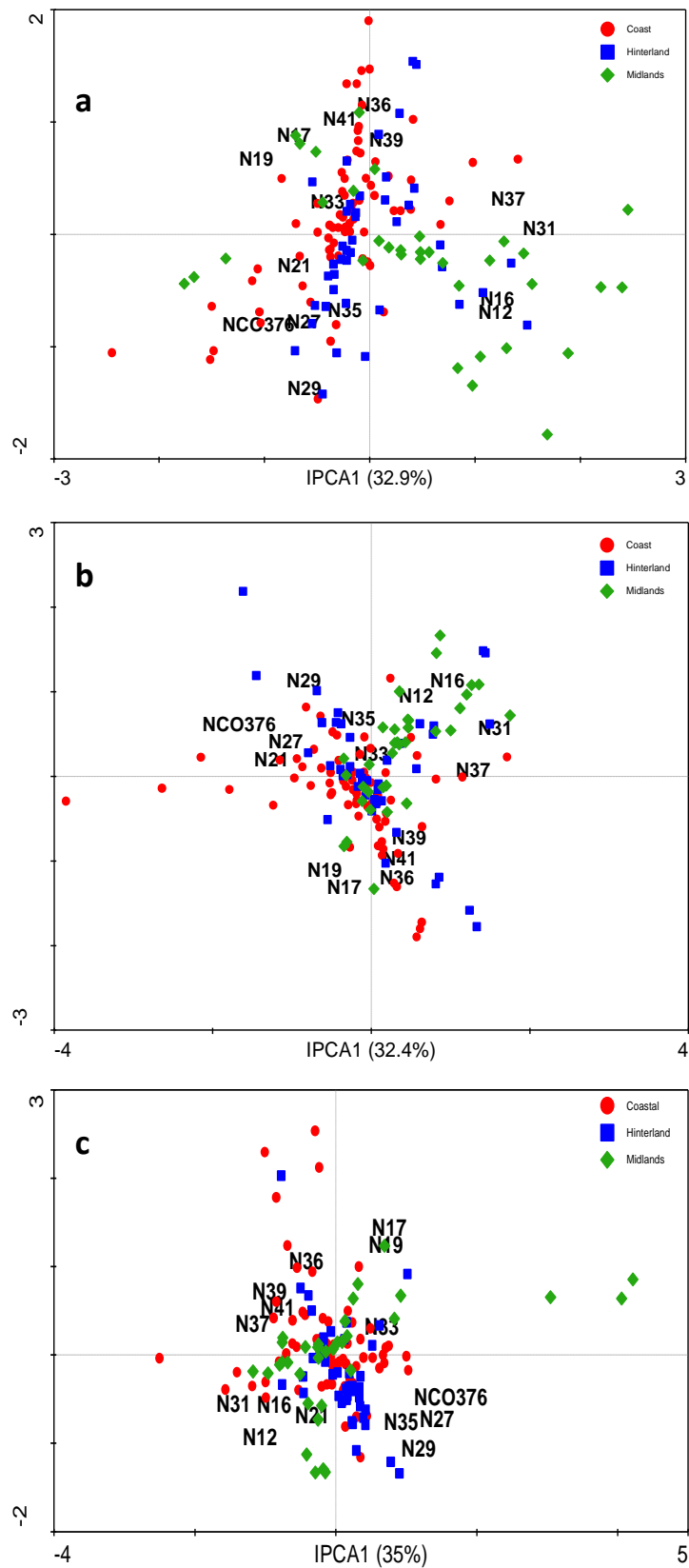
Figure 4.1 AMMI2 biplots for TERC (a), TCANE (b) and ERC (c) of 15 genotypes evaluated across 147 environments

4.4.1.2 Relationship between environments based on PCA

Figure 4.2 represents the PCA based on the environment x covariate two-way table, which accounted for 51.2% of the total variation. Separation of the C, H, and M regions were observed along the second PC axis, with most of the M environments forming a cluster at the bottom, while most of the C and H environments overlapped towards the middle and upper sector of the biplot. Some H environments formed part of the cluster of M environments in the bottom of the biplot, and were positively correlated to AGE. The covariates that had the largest effects on environmental separation (as indicated by the lengths of their vectors) included AGE, TT, WSI1, and WSI2. This was followed by RAIN, which was negatively correlated to EVP and RAD, while TOH and WSI3 showed minor effects on environment separation. The TT, RAD and EVP covariates were positively correlated and were associated to a larger extent with the C and H environments in general. The WSI1 (based on simulated rainfed vs. simulated irrigated evapotranspiration) and WSI2 (based on simulated rainfed vs. simulated irrigated cane yields) were positively correlated, as the Canesim model calculates cane yield as a function of evapotranspiration. Consequently, either one of these covariates could be used in future studies. Although also correlated to the other two water stress indices, the WSI3 (based on actual trial yield vs. simulated irrigated yield) showed a slightly different response, and was closely associated with RAIN. The M environments were characterized by higher AGE as well as lower TT, RAD, EVP and RAIN. Additionally, the negative correlation of the M environments to TOH suggests greater bias towards harvesting such trials early in the season.
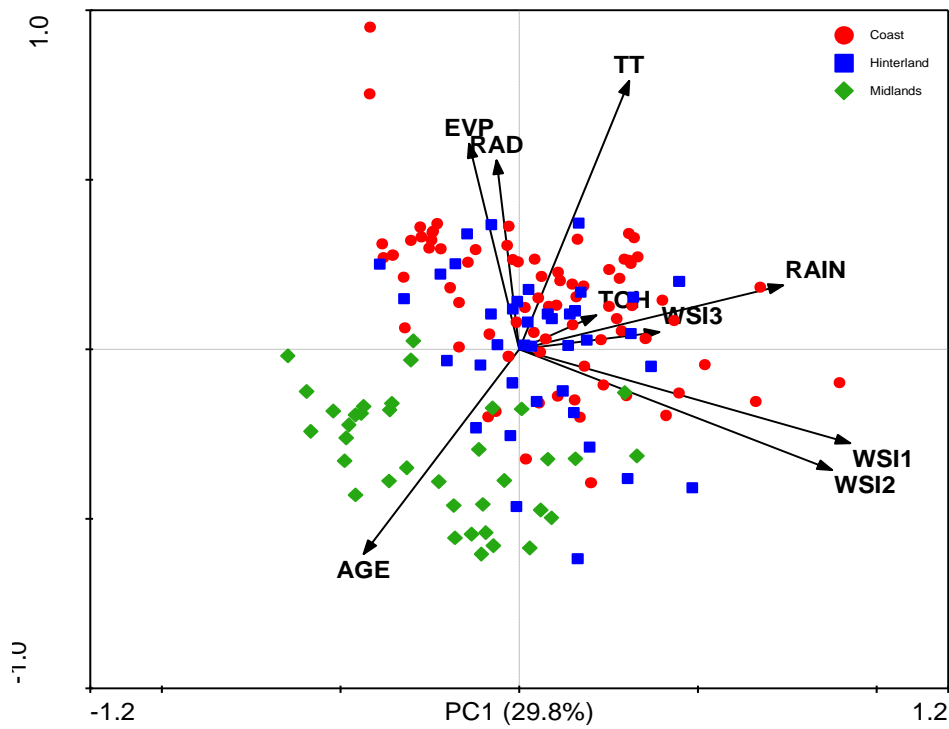
Figure 4.2 PCA biplot summarizing the relationships among 147 environments and nine environmental covariates

4.4.1.3 Interpretation of G x E interactions

The correlations between the AMMI IPCA scores and each of the nine seasonal covariates are indicated in Table 4.3. These correlations must be interpreted jointly with the environmental separation patterns observed in the AMMI biplots in Figure 4.1. The effects of the covariates on the G x E interactions are analyzed by interpreting the positions of environments (IPCA1 and IPCA2 scores) on the AMMI biplots and relating those positions/patterns to any observed correlations to the covariates. Most of the significant correlations were observed between the covariates and the IPCA1 scores. The AGE, TT, WSI1 and WSI3 covariates showed significant correlations with IPCA1 scores for all three variables. The TERC and TCANE AMMI biplots in Figures 4.1a and 4.1b showed that most of the M environments were characterized by higher IPCA1 scores compared with the C and H environments. The positive correlations of AGE to the TERC (0.58) and TCANE (0.37) IPCA1 scores suggest that the M environments were characterized by greater AGE. This is confirmed by the direction of the AGE vector in the PCA biplot in Figure 4.2. Similarly, the significant negative correlations of RAIN, TT, WSI1, WSI2, and WSI3 to the TERC and TCANE IPCA1 scores suggest that the M environments were characterized by lower daily rainfall and temperature, and higher water stress (low WSI values) compared with the C and

69

H environments. The three M environments with low IPCA1 scores in Figure 4.1a represent ratoons of a trial in a frost pocket (valley bottom), usually characterized by deeper soils with better water holding capacity and less water stress (high WSI values). Figure 4.1a showed a greater spread of the C and H environments compared with the M environments along the IPCA2 axis. Consequently, the positive correlation (0.55) between these IPCA2 scores and WSI3 shown in Table 4.3 suggest large variability in water stress within these regions, highlighting its importance for selection and evaluation. The significant correlations of AGE, RAD, TOH, TT and WSI3 to the TCANE IPCA2 scores, and the substantial deviation of the H environments along this axis in Figure 4.1b, highlights the importance of these covariates as discriminating factors within this region.

The ERC AMMI biplot did not reveal clear separation of the C, H, and M regions, thereby preventing clear associations between the environment groups and the covariates. There was greater deviation of environments along the IPCA2 axis compared with the IPCA1 axis (ignoring the three frost pocket environments). The WSI3 showed a weak (0.28) but significant (P<0.001) correlation to the ERC IPCA1 scores, while TOH was the only covariate that showed a significant correlation to the ERC IPCA2 scores. Therefore, the deviation of the C and H environments along the IPCA2 axis in Figure 4.1c is an indication of the importance of TOH as an ERC determinant in these two regions. Although EVP and RAD showed moderate effects on the separation of the M environments from the C and H environments in the PCA (Figure 4.2), these covariates showed no significant correlations to the AMMI IPCA1 scores, and only a single significant (P<0.05) correlation to IPCA2 scores for TCANE (Table 4.3).

Table 4.3 Correlation coefficients between nine seasonal and five site covariates and interaction principal component 1 (IPCA1) and 2 (IPCA2) environment scores derived from AMMI analysis of cane yield (TCANE), estimated recoverable crystal (ERC) and tons ERC/ha (TERC)

Covariate	G x E analysis						G x Trial analysis					
	IPCA1			IPCA2			IPCA1			IPCA2		
	TERC	TCANE	ERC	TERC	TCANE	ERC	TERC	TCANE	ERC	TERC	TCANE	ERC
AGE	0.58^{***}	0.37^{***}	-0.20^{**}	0.16^{ns}	0.33^{***}	0.02^{ns}	0.65^{***}	-0.51^{***}	-0.29^{ns}	-0.06^{ns}	0.08^{ns}	-0.34^{*}
EVP	-0.12^{ns}	-0.03^{ns}	0.07^{ns}	-0.08^{ns}	-0.09^{ns}	-0.00^{ns}	-0.21^{ns}	0.17^{ns}	0.04^{ns}	-0.10^{ns}	-0.18^{ns}	0.09^{ns}
RAD	0.09^{ns}	-0.01^{ns}	-0.14^{ns}	0.11^{ns}	0.18^{*}	-0.06^{ns}	0.12^{ns}	0.02^{ns}	-0.19^{ns}	0.09^{ns}	0.09^{ns}	-0.09^{ns}
RAIN	-0.29^{***}	-0.25^{**}	-0.08^{ns}	0.02^{ns}	-0.02^{ns}	-0.10^{ns}	-0.44^{**}	0.42^{**}	0.18^{ns}	0.23^{ns}	0.13^{ns}	0.19^{ns}
TOH	-0.14^{ns}	0.01^{ns}	0.19^{*}	0.00^{ns}	-0.22^{**}	0.36^{***}	-0.15^{ns}	-0.01^{ns}	0.19^{ns}	0.09^{ns}	-0.00^{ns}	-0.01^{ns}
TT	-0.24^{**}	-0.25^{**}	-0.23^{**}	-0.19^{*}	-0.19^{*}	-0.07^{ns}	-0.23^{ns}	0.35^{*}	-0.23^{ns}	-0.08^{ns}	-0.09^{ns}	0.12^{ns}
WSI1	-0.25^{**}	-0.25^{**}	0.17^{*}	0.16^{ns}	0.09^{ns}	-0.07^{ns}	-0.34^{*}	0.29^{ns}	0.31^{*}	0.32^{*}	0.26^{ns}	0.14^{ns}
WSI2	-0.18^{*}	-0.22^{**}	0.09^{ns}	0.15^{ns}	0.12^{ns}	-0.11^{ns}	-0.24^{ns}	0.22^{ns}	0.22^{ns}	0.28^{ns}	0.26^{ns}	0.14^{ns}
WSI3	-0.31^{***}	-0.44^{***}	0.28^{***}	0.55^{***}	0.40^{***}	-0.04^{ns}	-0.27^{ns}	0.28^{ns}	0.45^{**}	0.58^{***}	0.42^{**}	-0.06^{ns}
Clay	-	-	-	-	-	-	-0.16^{ns}	0.08^{ns}	0.23^{ns}	0.1^{ns}	0.14^{ns}	0.28^{ns}
Depth	-	-	-	-	-	-	-0.33^{*}	0.30^{*}	0.32^{*}	0.04^{ns}	-0.17^{ns}	-0.06^{ns}
Ncat	-	-	-	-	-	-	0.12^{ns}	-0.12^{ns}	0.05^{ns}	0.26^{ns}	0.34^{*}	0.04^{ns}
OM	-	-	-	-	-	-	0.24^{ns}	-0.28^{ns}	0.04^{ns}	0.25^{ns}	0.37^{*}	-0.02^{ns}
TAM	-	-	-	-	-	-	-0.32^{*}	0.34^{*}	0.24^{ns}	0.19^{ns}	0.07^{ns}	0.07^{ns}

* P<0.05, **P<0.01, *** P<0.001, ns = not significant

4.4.2 G x Trial interactions

In this approach, the phenotypic data of the genotypes were averaged across ratoons within a trial and the two-way tables of G x trials were analyzed using AMMI. The AMMI biplots for the G x trial interactions (not shown) revealed similar patterns of responses observed in the G x E biplots. Genotypic groupings mimicked the G x E biplots, while trials within regions showed similar deviations as the environments (within regions) in the previous analysis (G x E analysis). The second PCA analysis i.e. with site covariates included and seasonal covariates averaged across ratoons, accounted for 44.4% of the variation and revealed slightly different patterns to the first PCA analysis (Figure 4.3). Once again, most of the M trials formed a group and were positively correlated to AGE. All site covariates were positively correlated to each other, with Clay and TAM accounting for most of the variability (longest vectors). Either site covariate could therefore be used as a group representative for further evaluations. Additionally, WSI3 (actual trial mean yield vs. simulated irrigated yield) was

more strongly associated with the site covariates than the other two water stress indices. This suggests its possible superiority as a representative index to characterize water availability, which is essentially defined by the site covariates. The site covariates appeared to be negatively associated with a large proportion of the H trials, suggesting that these trial sites are dominated by poorer soils. Also, RAIN, TT, and TOH were once again correlated and negatively associated with the M environments, while EVP and RAD accounted for very little variation.

The correlations of the G x Trial AMMI IPCA scores to the site and season covariates are also shown in Table 4.3. The AGE covariate showed strong significant correlations to the IPCA1 scores for TERC (0.65) and TCANE (-0.51), once again highlighting its important role in discriminating regions and trials. The RAIN covariate showed stronger correlations with the G x T IPCA1 scores for TERC and TCANE compared with the G x E IPCA1 scores for the same variables, while TT only showed a significant correlation with the TCANE IPCA1 score. The WSI3 showed stronger correlations than WSI1 for the ERC IPCA1 scores, TERC IPCA2 scores, and TCANE IPCA2 scores. These stronger correlations again highlight the superiority of WSI3 for use as an indicator of water deficient environments in this study. From all site covariates, DEPTH showed significant correlations with IPCA1 scores for the three variates and interpretation of this correlation with the AMMI biplots suggested that most of the M trials were characterized by shallower soils. The biplots also showed that most of the deviation along the IPCA1 axis was observed by the M trials, suggesting that sites in this region were the most variable in soil depth. The Ncat and OM showed significant (P<0.05) correlations with the TCANE IPCA2 scores, suggesting that deviations of trials along this axis, which were evident within all three regions, were due to differences in Ncat and OM. The TAM caused separation of trials along the TCANE and TERC IPCA1 axis, and interpretations from the biplots suggested that M trials were characterized by lower TAM.
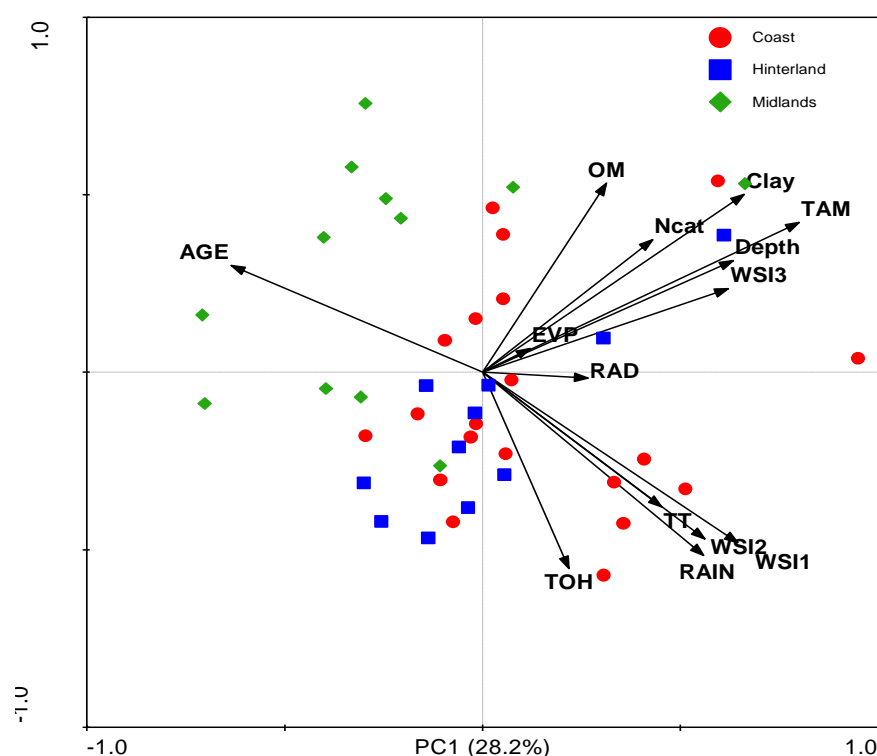
Figure 4.3 PCA biplot summarizing the relationships among 43 trials, nine seasonal covariates, and five site covariates

4.5 Discussion and conclusions

The results of this study showed that the M environments/trials were unique in their ability to discriminate cultivar performance in TERC and TCANE. The M environments regularly formed distinct clusters, separate from the C and H environments, on both the AMMI and PCA biplots. The overlap of the C and H environments on the AMMI biplots suggests that these regions discriminated cultivars similarly, while the PCA biplots showed that the environments in these regions were also characterized by similar climatic and soil features. These regions should be considered a single mega-environment within the rainfed selection program and for any future studies involving the interpretation of G x E interactions. The similarities between environments in the C and H regions were also highlighted in Chapter 3, where GGE biplot analysis was performed on the same dataset. The biplots produced from the GGE biplot analyses differ slightly from those of AMMI analyses. The discrepancy lies in the treatment of the G x E matrix prior to performing singular value decomposition. The GGE biplot technique utilizes environment centered data (matrix minus environment means), while AMMI uses the matrix of residuals (matrix minus genotype and environment means). Despite

this difference, the techniques allowed for similar conclusions to be drawn. The AMMI was selected as the analysis of choice in this study due to the demonstrated successes of interpreting environmental covariates through its use (De Vita et al., 2010; Van Eeuwijk et al., 1995; Van Oosterom et al., 1996; Voltas et al., 1999).

The approach used in this study identified factors that separated rainfed environments. However, the correlations between AMMI IPCA scores and covariates were generally weak (strongest correlation of 0.65). This may be due to the large size of the dataset, which led to weak but significant correlations. This limited the further use of the technique to enrich the AMMI biplots with the covariates (Van Eeuwijk and Elgersma, 1993), which could have allowed for easier interpretations. Nevertheless, the useful inferences made about the nature of the different environments and regions will inform future site selection for breeding and evaluation. The variability and significance ($P<0.01$) of water stress within the C and H regions suggest that improved genetic tolerance to limited water availability may be achieved through selection in contrasting environments within these regions. The results also suggested that the M region was characterized by greater variability in soil depth. This is in contrast to local knowledge of regional soil characteristics, as greater topographic variability characteristic of the C and H regions was expected to produce greater variability in soil depth. The discrepancy may be linked to biased trial site selection in the C and H regions, where the most homogeneous, accessible, high potential fields were conventionally chosen. Consequently, more emphasis should be placed on selecting trial sites with contrasting soil characteristics within the C and H regions.

The PCA of the environment x covariate matrix identified AGE as a dominant factor influencing the separation of the M region from other regions in the rainfed parts of the industry. The M region is characterized by higher altitude, which is normally associated with lower temperatures and slower growth rates, thereby requiring longer crop cycles. The negative correlation between AGE and TT is depicted in Figure 4.2, which also shows that the M environments are characterized by lower daily RAD, RAIN, and EVP. These are typical indications of an overall lower climatic potential of cropping environments. However, this is not the case with the M environments, where lower climatic potential is compensated for by the longer cutting cycle (AGE). Other studies have demonstrated higher yields in M environments. The present study has shown that these higher yields are related to increased harvest age rather than superior climatic potential of the region.

In this study, the influence of climatic factors and water stress was captured using the average daily values for the entire crop cycles (plant/ratoon to harvest). These averages may not be appropriate (and may be misleading), especially when considering the within-season weather variations and their influence on particular growth phases. For example, high water stress during the early growth phase of tillering may negatively affect stalk population and final cane yield. Favourable moisture conditions in later growth phases will improve the seasonal average; however, the effects of early water stress may go undetected. Radiation, evaporation and rainfall did not show clear, consistent associations with IPCA scores. This lack of correlation may have also been due to the use of seasonal averages, as high levels during one growth phase may have been offset by lower levels during another phase. A more accurate approach would be to characterize climatic and water stress covariates within individual growth phases to determine their specific effects on G x E interactions. This approach has been used in other crops such as barley (Voltas et al., 1999), ryegrass (Van Eeuwijk and Elgersma, 1993) and wheat (Voltas et al., 2005). No attempts have been made to interpret sugarcane G x E interactions using this approach, and the following chapters will address this issue.

The G x E interactions may be categorized as being repeatable or non-repeatable, depending on the likelihood of particular G x E interaction events being observed again (DeLacy et al., 2010). The first approach to the analysis in this study (G x E) represented an investigation into the combined influence of repeatable and non-repeatable G x E interactions, while the second approach to the analysis (G x Trial) represented an evaluation of the repeatable component of G x E interactions. The second approach was not successful at identifying site factors that correlated with IPCA scores, as most covariates were not significant. This was unexpected as Clay, in particular, is generally considered an important site factor causing variation in cultivar responses in the sugar industry. The lack of frequent correlations between IPCA scores and covariates such as Clay, Ncat, and OM suggest that these may not be as influential on G x E interactions as previously thought. Stronger correlations were observed between IPCA scores and the covariates relating to soil physical characteristics i.e. Depth and TAM. Based on their influence on the repeatable G x Trial interaction, these covariates should be given priority when selecting contrasting sites for selection and evaluation. Furthermore, their mutual correlation shown in Figure 4.3, suggests that either

one may be used as a representative covariate for future studies (TAM was derived from Depth measurements).

With respect to water stress indices, WSI3 was identified as the most promising at detecting genotypic responses to environmental differences in water stress, due to its correlations with IPCA scores, as well as its correlations with the site covariates in Figure 4.3. In this respect (correlation with site covariates), WSI3 best captured the combined influence of the site covariates, suggesting its potential use in future studies as an integrated covariate that represents the ability of soil to provide moisture. The WSI1 and WSI2, which were based on simulated evapotranspiration and cane yields, respectively, didn't correlate as strongly with IPCA scores as WSI3. This is likely due to WSI3 being derived from observed trial mean yields, while the other two indices were based completely on simulated responses and the inaccuracies usually associated with such simulations.

In this study, the categorization of environments into the C, H, and M regions were pre-determined and based on industry knowledge. The joint use of AMMI and PCA successfully separated the regions (the M region specifically) according to their environmental characteristics and the genotypic responses, thereby confirming local knowledge. The patterns of genotypic responses to environments in the AMMI analyses were explained by the environmental characterization based on seasonal and site covariates. This suggests that the methodology could have potential benefits to sugarcane G x E studies, particularly in areas where little is known about the mega-environment constitution of a crop area. This is in contrast to Ceretta and Van Eeuwijk (2008), who found that differential barley cultivar responses between two environments (using GGE biplot) were not fully explained by an environmental characterization based on PCA of meteorological variables. Additionally, this study highlighted the value of environmental characterization of sugarcane field trials and demonstrated the potential for integrated use of crop models to interpret G x E interactions. Future initiatives will focus on optimizing the use of crop models in G x E interpretive studies through evaluations of effectiveness of different weather data sources, comparisons of simulated vs. observed yields, and using simulated yields as hypothetical probe genotypes (commonly studied genotypes for which detailed physiological and yield parameters are known).

In addition to the effects of some covariates on regional separation, substantial variation was also observed within regions i.e. environments within regions showed deviation along IPCA

axes. Attempts will be made, within regions, to group environments based on either seasonal covariates or G x E patterns. These analyses will also involve the use of genotypic covariates (traits) to help investigate the responses of sugarcane phenotypic groups to environmental factors. Such an approach will complement this study and provide a clearer understanding of variation within regions to inform trial site selection and region-specific breeding objectives.

4.6 References

Anon., 2000. Information sheet 7.1: Recommendations for nitrogen. The South African Sugar Association Experiment Station, Mount Edgecombe, Durban, South Africa, pp. 2-3.

Anon., 2009. GenStat® Executable Release 12.1. Lawes Agricultural Trust, Rothamstead Experimental Station, Harpenden, U.K. Clarendon Press, London, U.K.

Annicchiarico, P., Harzic, N., Carroni, A.M., 2010. Adaptation, diversity, and exploitation of global white lupin (*Lupinus albus* L.) landrace genetic resources. *Field Crops Research*. 119, 114-124.

Baril, C.P., Denis, J.B., Wustman, R., Van Eeuwijk, F.A., 1995. Analysing genotype by environment interaction in Dutch potato variety trials using factorial regression. *Euphytica* 82, 149-155.

Bissessur, D., Ramnawaz, C., Ramdoyal, K., 2010. Use of historical multi-location sugar cane variety trials data to identify relationships among environments in terms of genotype responses. *Proceedings of the International Society of Sugar Cane Technologists* 27, 1-8.

Ceretta, S., Van Eeuwijk, F.A., 2008. Grain yield variation in malting barley cultivars in Uruguay and its consequences for the design of a trials network. *Crop Science* 48, 167-180.

DeLacy, I.H., Kaul, S., Rana, B.S., Cooper, M., 2010. Genotypic variation for grain and stover yield of dryland (rabi) sorghum in India: 1. Magnitude of genotype x environment interactions. *Field Crops Research* 118, 228-235.

Del Blanco, I.A., Glaz, B., Edme, S.J., 2010. Improving efficiency of sugarcane genotype selection in Florida. *Crop Science* 50, 1744-1753.

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

Gauch, H.G., 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam.

Gilbert, R.A., Shine, J.M., Miller, J.D., Rice, R.W., Rainbolt, C.R., 2006. The effect of genotype, environment and time of harvest on sugarcane yields in Florida, USA. *Field Crops Research* 95, 156-170.

Glaz, B., Kang, M.S., 2008. Location contributions determined via GGE biplot analysis of multi-environment sugarcane genotype-performance trials. *Crop Science* 48, 941-950.

Jackson, P., Horsley, D., Foreman, J., Hogarth, D.M., Wood, A.W., 1991. Genotype x environment (GE) interactions in sugarcane variety trials in the Herbert. *Proceedings of the Australian Society of Sugarcane Technologists* 13, 103-109.

Jackson, P., McRae, T., Hogarth, M., 1995. Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. *Field Crops Research* 43, 119-130.

Jackson, P., Chapman, S., Rattey, A., Wei, X., Cox, M., 2007. Genotype x region interactions, and implications for sugarcane breeding programs. *Sugar Cane International* 25, 3-7.

Mirzawan, P.D.N., Cooper, M., DeLacy, I.H., Hogarth, D.M., 1994. Retrospective analysis of the relationships among test environments of the southern Queensland sugarcane breeding programme. *Theoretical and Applied Genetics* 88, 707-716.

Queme, J.L., Orozco, H., Ovalle, W., Melgar, M., 2005. Analysis of genotype-by-environment interaction for sugarcane based on the AMMI model. *Sugar Cane International* 23, 21-24.

Queme, J.L., Crossa, J., Orozco, H., Melgar, M., 2007. Analysis of genotype-by-environment interaction for sugarcane using the sites regression model (SREG). *Proceedings of the International Society of Sugar Cane Technologists* 26, 764-769.

Ramburan, S., Sewpersad, C., 2009. Investigating cultivar-by-cutting cycle interactions for coastal sugarcane production in South Africa. *Sugar Cane International* 27 164-168.

Rattey, A.R., Kimbeng, C.A., 2001. Genotype by environment interactions and resource allocation in final stage selection trials in the Burdekin district. *Proceedings of the Australian Society of Sugarcane Technologists* 23, 136-141.

Singels, A., Donaldson, R.A., 2000. A simple model of unstressed sugarcane canopy development. *Proceedings of the South African Sugar Technologists Association* 74, 151-154.

Tai, P.Y.P., Rice, E.R., Chew, V., Miller, J.D., 1982. Phenotypic stability analyses of sugarcane cultivar performance tests. *Crop Science* 22, 1179-1184.

Ter Braak, C.J.F., Smilauer, P., 2003. Canoco® for Windows version 4.51. Biomteris Plant Research International, Wageningen, The Netherlands.

Van Antwerpen, R., Meyer, J.H., Johnston, M.A., 1994. Estimating water retention of some Natal sugar belt soils in relation to clay content. *Proceedings of the South African Sugar Technologists Association* 68, 75-79.

Van Eeuwijk, F.A., Elgersma, A., 1993. Incorporating environmental information in an analysis of genotype by environment interaction for seed yield in perennial ryegrass. *Heredity* 70, 447-457.

Van Eeuwijk, F.A., Keizer, L.C.P., Bakker, J.J., 1995. Linear and bilinear models for the analysis of multi-environment trials: II An application to data from the Dutch maize variety trials. *Euphytica* 84, 9-22.

Van Oosterom, E.J., Mahalakshmi, V., Bidinger, F.R., Rao, K.P., 1996. Effect of water availability and temperature on the genotype-by-environment interaction of pearl millet in semi-arid tropical environments. *Euphytica* 89, 175-183.

Vargas, M., Crossa, J., Van Eeuwijk, F.A., Ramirez, E., Sayre, K., 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype x environment interaction. *Crop Science* 39, 955-967.

Voltas, J., Van Eeuwijk, F.A., Sombrero, A., Lafarga, A., Igartua, E., Romagosa, I., 1999. Integrating statistical and ecophysiological analyses of genotype by environment interaction for grain filling of barley I. Individual grain weight. *Field Crops Research* 62, 63-74.

Voltas, J., Lopez-Corcoles, H., Borras, G., 2005. Use of biplot analysis and factorial regression for the investigation of superior genotypes in multi-environment trials. *European Journal of Agronomy* 22, 309-324.

Yan, W., Hunt, L.A., 2001. Interpretation of genotype x environment interaction for winter wheat in Ontario. *Crop Science* 41, 19-2.

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

CHAPTER 5

TEST SITE SIMILARITY, TRAIT RELATIONS AND CAUSES OF GENOTYPE X ENVIRONMENT INTERACTIONS IN SUGARCANE IN THE MIDLANDS REGION OF SOUTH AFRICA[3]

5.1 Abstract

Commercial genotypes are selected for the midlands region of the South African sugar industry from six testing sites. The similarities between these sites in genotype performance have never been evaluated, and factors causing observed G x E interactions in the midlands are unclear. The objectives of this study were (i) to investigate the mega-environment (ME) patterns of the test sites, (ii) to investigate causes of G x E interactions, and (iii) to investigate whether sugarcane G x E patterns could be explained by relevant environmental factors and different genotypic trait relations. Estimated recoverable crystal (ERC) yield data from eight series (1999 to 2006) of multi-environment trials (METs) were analyzed using GGE biplot analysis. Data from post-release evaluation trials were incorporated into the multi-series dataset to identify overall MEs for the region. Pattern analysis was used to formulate environment groups and genotype sets based on G x E patterns and trait relations, respectively. Average daily temperature, rainfall, evaporation, and radiation; age and time of harvest, and a derived water stress index were calculated for each trial. Soil characteristics including effective rooting depth, clay and organic matter percentages, nitrogen mineralization category, and total available moisture were also determined.

The testing sites formed a single, complex ME with frequent crossover G x E interaction. Post-release trials in frost pockets formed a unique ME, suggesting that the network should be re-designed to incorporate such a testing site. Site S3 was identified as redundant and should be removed from the trial network, while site B3 was flagged as another possible redundant site. Pattern analysis identified six environment groups and five genotype sets, the characteristics of which were interpreted using known production scenarios in the region. The clustering of sites on the biplots was related to similarities in soil characteristics, while the

---

[3]Ramburan, S., Zhou, M., Labuschagne, M., 2012. Investigating test site similarity, trait relations, and causes of genotype x environment interactions in the Midlands region of South Africa. *Field Crops Research* 129, 71-80.

climatic covariates had negligible effects, and were not associated with G x E patterns. Trait relations between genotype sets did not differ drastically among environment groups, however, useful trait relations were established to aid with future selection. The approaches and principles described in this study can be used in other sugarcane studies to enhance the value of METs and gain an understanding of causes of G x E interactions.

5.2 Introduction

G x E interaction complicates the identification of superior genotypes in MET networks. Consequently, MET networks must include a range of locations, years, and management regimes. This ensures that genotypes are selected under wide variation in biotic and abiotic factors, with the expectation of superior performance being repeated after release. Such an empirical approach is the basis of most crop improvement programs, which are consequently highly dependent on the representativeness of test sites. However, for reasons such as limited resources, logistics, politics, convenience, and difficulties with land acquisition, the selection of test sites for breeding programs is often not ideal (Hamblin et al., 1980). Long-established programs conventionally comprise sites that were subjectively designated according to factors such as soil type, management regime, and geographic zoning. Resource limitations, financial constraints, and the increasing demand for high yielding genotypes necessitate the continual evaluation of the appropriateness of selection sites for the improvement of network efficiencies.

In addition to identifying superior genotypes for specific environments, MET networks can also provide information on the nature of the target population of environments. The GGE biplot analysis (Yan et al., 2000) has gained popularity as a tool to investigate G x E interactions, and in particular, to identify MEs within a MET network. The method evaluates the similarity among test sites in their ability to discriminate genotypes, and identifies the most interactive or redundant sites within a MET network. Yan et al. (2000) used the GGE biplot technique to show that winter wheat production environments in Canada should be grouped into two mega-environments, as opposed to a traditional grouping of four sub-areas. Yan and Rajcan (2002) employed the GGE biplot technique to soybean MET data and identified a single mega-environment with frequent crossover G x E interactions. Malla et al. (2010) used GGE biplots to divide the South Dakota winter wheat testing environments into homogenous groups. Sharma et al. (2010) used GGE biplots to determine the performance, stability, and superiority of winter wheat breeding lines in irrigated environments in central

and west Asia. Several other studies have successfully applied GGE biplot analysis to evaluate similarities between test environments (Dehghani et al., 2006; Setimela et al., 2010; Zhe et al., 2010). The application of GGE biplot analysis in sugarcane is limited to studies conducted in Florida, USA (Glaz and Kang, 2008), and Guatemala (Queme et al., 2010).

Although historic breeding advance has been achieved empirically, it is generally accepted that an understanding of the nature and causes of G x E interactions will further accelerate genetic gains from breeding. A lack of knowledge of the causes and repeatability of G x E interactions reduces confidence in the effectiveness of selection (Mirzawan et al., 1994). Knowledge of the factors responsible for both predictable (influenced by site and management factors) and unpredictable (influenced by season or crop year) G x E interactions is critical for the structuring of effective MET networks and for the development of recommendation domains. For example, the identification of overriding factors causing repeatable G x E interaction may indicate the need for dedicated selection sites with contrasting levels of such factors e.g. moisture stress. Furthermore, an understanding of the variability in crop trait relations in response to fluctuations in G x E dependant factors may lead to more targeted trait selection strategies for different MEs within a target region. However, because of the lack of associated climatic, soil, and genotypic trait data from METs, such interpretive studies of G x E interactions are limited in comparison to traditional empirical studies. When associated environmental data are available, interpretive studies of G x E interactions have conventionally involved the correlation of climatic data to environment scores derived from PCA – based methods,, such as AMMI (Gauch, 1992). This approach has been used in interpretive studies on crops such as barley (Voltas et al., 1999), pearl millet (Van Oosterom et al., 1996), ryegrass (Van Eeuwijk and Elgersma, 1993), wheat (De Vita et al., 2010) and others. The interpretation of sugarcane G x E interactions using this approach is limited to a single study (Jackson et al., 1995).

The G x E interactions are conventionally analyzed using a combination of PCA and classification methods, such as cluster analysis. The combined use of these techniques is commonly referred to as pattern analysis (DeLacy et al., 1996). In this approach, individuals (genotypes or environments) are classified into homogenous groups depending on relative genotype performance (in the case of environments) or phenotypic trait similarities (in the case of genotypes). The PCA is often used to visualize G x E patterns in biplots, while cluster analysis allows for the grouping of environments that discriminate genotypes similarly. In some cases, cluster analysis is used to group genotypes based on their trait values (De la

Vega and Chapman, 2010). Clustering and PCA (ordination) are powerful tools, which, when combined with associated environmental covariates can allow for the interpretation of causes of G x E interactions. DeLacy et al. (2010) recently demonstrated the use of pattern analysis methodology in a study of sorghum in India, where trait relations of genotypes were studied in five different environment groups derived from cluster analysis of G x E yield data. Given the large diversity of genotypes expressing different traits in a range of environments typical of conventional METs, any methodology involving the grouping of genotypes and environments will greatly assist in interpreting G x E interactions. For example, the performance of genotype groups in different types of environments will imply the need for differential trait selection strategies within a testing network.

The previous two chapters of this study have illustrated the complex G x E interactions present in the rainfed regions of the South African sugar industry. Those chapters showed that the midlands region is unique in the way it discriminates genotype performance (Chapter 3) and with regard to the environmental factors driving those differences (Chapter 4). Production in the midlands region occurs at high altitudes (>600 m above sea level) with low temperatures leading to occasional frost in low lying fields, and with harvest ages that vary between 12 to 24 months. The region is divided into northern and southern production areas, and all genotype selection occurs in the northern areas only. Commercial sugarcane genotypes are selected and released for this region from one of two selection programs. The Bruyns Hill (B) and Glenside (S) selection programs run concurrently to identify promising genotypes suited to humic (high potential) and sandy (low potential) soils, respectively. Each program is represented by three permanent testing sites located either on SASRI owned research farms (B1 and S1), or on grower co-operator farms (B2, B3, S2, S3) in the region. The so-called "off-station" sites B2, B3, S2 and S3, on co-operator farms, represent growing conditions that are slightly different from those experienced on the research farms. The criteria for choosing the research farms and the off-station sites varied widely, and included factors such as perceived/expected differences in soil types, geographical differences, politics, and convenience. The cultivar information gathered from the selection programs is supplemented by post-release evaluation trials conducted on commercial farms. The objective of post-release trials is to gather additional information on the performance of cultivars under more diverse conditions not represented by the selection trials, primarily for recommendation purposes (Ramburan and Van den Berg, 2011).

Since the inception of the above selection sites in 1997 (Nuss, 1998) and the *ad hoc* establishment of post-release trials since the 1980s, no study has been conducted to evaluate the relationships between the test sites in genotypic performance. Although G x E interactions are regularly observed between the B1 and S1 sites (on research farms), very often, similar genotypic performance has been observed between the off-station sites, and their similarities in this regard need to be evaluated. Despite the availability of appropriate data and a wide range of statistical tools, there has not been a systematic analysis of the MET network for the midlands region. The objectives of this study were (i) to investigate the ME groupings of the test sites used for the midlands region to determine the need for network re-structuring, (ii) to investigate causes of G x E interactions that can inform future test site location within the region, and (iii) to investigate whether sugarcane G x E patterns could be explained by relevant environmental factors and different genotypic trait structures for differential trait selection strategies in future.

5.3 Materials and methods

5.3.1 The MET dataset

The dataset used for this study comprised advanced plant breeding selection trials established between 1999 and 2006 at six locations (B1, B2, B3, S1, S2, and S3) in the midlands region. Each year, a set of 32 to 41 promising sugarcane genotypes that had progressed through the selection stages were tested in a MET, together with three to six commercial checks. Trials established in each planting year (series) were harvested as plant, first and second ratoon crops. However, some trials were not harvested across all ratoons. The genotype x trial dataset was balanced within a series, but unbalanced across different series (except for common control genotypes). In this analysis, the trial series are named according to the year of planting i.e. 1999, 2000, 2001, 2002, 2003, 2004, 2005 and 2006. Trials were harvested between April and December and varied in harvest age between 12 and 24 months. The trials were planted as randomized complete-block designs with three to four replicates. Experimental units comprised five sugarcane rows that were 10 m long, and spaced 1 m apart. At each harvest event, three net rows were hand-harvested and weighed using a digital scale mounted on a tractor operated hydraulic boom, and used to determine cane yield (TCANE) in ton/ha. A 12-stalk sample was taken from each plot to determine individual stalk mass (STKMS) and a range of milling characteristics, including estimated recoverable crystal (ERC), dry matter (DM), and fibre (FIB) percentages. The TERC was calculated as a product of TCANE and ERC. Other important traits measured at harvest included stalk population

(POP) in stalks/ha and stalk diameter (DIAM) in mm. Visual ratings (1 to 9) of lodging (LOD), canopy formation (CAN) and stalk bending (SKEW) were also done in each trial at appropriate growth stages.

For each trial in the dataset, stable site characteristics were determined through soil profile analyses and conventional soil sampling procedures to develop five *site covariates*. Apart from the stable site covariates, seven other covariates that varied seasonally with each ratoon (hence termed *seasonal covariates*) were also determined. The descriptions of the site and seasonal covariates are given in Chapter 4. The water stress index (WSI) in this study was calculated as a ratio of simulated rainfed evapotranspiration to simulated irrigated evapotranspiration. Further details of the covariates, together with their means and ranges are indicated in Table 5.1.

Table 5.1 Descriptions of seven seasonal and five site covariates measured over 57 trials and 119 environments (trial x ratoon combinations)

Covariate	Description	Mean	Range
	Seasonal covariates		
TT	Thermal time in heat units using a base temperature of 10°C (Heat units/day)	8.05	7.10-9.36
RAD	Average daily solar radiation (MJ/m^2/sec/day)	16.39	14.34-17.58
RAIN	Average daily rainfall (mm/day)	2.16	1.22-3.24
EVP	Average daily A-pan evaporation (mm/day)	3.85	2.93-5.64
WSI	Water stress index. Simulated rainfed evapotranspiration vs. simulated irrigated evapotranspiration	0.49	0.26-0.71
AGE	Age at harvest (months)	21.18	7.39-28.19
TOH	Time of harvest (1-9). Coded as month 1 (April) to month 9 (December)	3.58	1-9
	Site covariates		
ERD	Estimated rooting depth (cm)	62.53	35-91
Clay	Clay percentage (%)	19.54	6 - 44
OM	Organic matter percentage (%)	2.46	0.79-6.57
Ncat	Nitrogen mineralization category (1-4)	N/A	N/A
TAM	Total available moisture (mm)	60.43	23.44-104.34

5.3.2 Empirical analysis of G x E interactions

A trial x ratoon combination was considered an environment. Within each series, the two-way array of genotype x environment TERC means (derived from restricted maximum

likelihood analysis) was analyzed using GGE biplot software (Yan, 2001). A detailed description of this software is given in Chapter 3. In this study, the polygon viewing option was used to investigate the ME constitution of the midlands. In this view, a polygon is drawn by connecting genotypes that are located furthest away from the biplot origin. These vertex genotypes were the most responsive i.e. they were either the best or poorest performers in some or all of the environments. A perpendicular line is drawn from each side of the polygon through the biplot origin such that the biplot is divided into sectors. The vertex genotype in each sector is the best genotype in all environments whose markers fall within that sector. Environments within the same sector share the same winning genotype, and environments in different sectors have different winning genotypes. Thus, the polygon view indicates the presence or absence of crossover G x E interaction, and is suggestive of the existence or absence of different mega-environments among the test sites (Yan and Rajcan, 2002). In addition to selection trials, post-release evaluation trials conducted on co-operator farms during the period 1999 to 2006 were also analyzed. These trials contained a subset of six commercial genotypes that were also routinely used as controls in the selection trials. The data from these trials were subsequently combined with the selection trial dataset, using only the six commercial genotypes. This allowed for a combined biplot analysis of the selection and post-release trials across trial series, to evaluate the overall ME constitution of the midlands region.

5.3.3 Interpretation of G x E interactions

The approach used to interpret the G x E interactions was to create genotype sets and environment groups based on phenotypic trait similarities and G x E patterns, respectively. Differential genotype-set responses were then interpreted in terms of environmental characteristics associated with each environment group. The environment groups were formed by applying pattern analysis methodology (DeLacy et al., 1996) to the combined selection and post-release trial dataset that contained the subset of six commercial genotypes, as described above. The G x E matrix of TERC means was subjected to GGE biplot analysis, and cluster analysis was performed on the first four PC environment scores that were derived. The cluster analysis employed hierarchical agglomerative clustering with squared Euclidean distance as the dissimilarity measure and average linkage as the fusion criterion, as implemented in Genstat 12.0 (Anonymous, 2009). The classifications were portrayed by a dendrogram, and resulted in the production of six environment groups. The average covariate values within each group were then compared to each other to determine if the environmental

groupings determined from G x E patterns could be explained by differences in environmental factors.

The genotype sets were formed by fitting the following random model for each measured trait across all trial series:

$$Y_{ijkl} = \mu + T_j + B(T)_{l(j)} + G_i + GT_{ij} + GB(T)_{il(j)} + R_k + RT_{jk} + RB(T)_{kl(j)} + RG_{ik} + RGT_{ijk} + e$$

where Y_{ijkl} is the observation l of genotype i in ratoon k of trial j, μ is the grand mean, T_j is the trial main effect, $B(T)_{l(j)}$ is the effect of block within trial, G_i is the genotype main effect, GT_{ij} is the effect of the genotype x trial interaction, $GB(T)_{il(j)}$ is the effect of genotype x block (within trial), R_k is the main effect of ratoon, RT_{jk} is the effect of the ratoon x trial interaction, $RB(T)_{kl(j)}$ is the effect of the ratoon x block (within trial) interaction, RG_{ik} is the effect of the ratoon x genotype interaction, RGT_{ijk} is the effect of the genotype x trial x ratoon interaction, and e is the error. Restricted maximum likelihood (Patterson and Thompson, 1971) using the sparse average information algorithm (Gilmour et al., 1995) was used, as implemented in GenStat 12.0 (Anonymous, 2009). The associations among the trait best linear unbiased predictors (BLUPS) were examined through PCA and cluster analysis. Because of the different measurement scales, the data for the 11 traits were trait-standardized by subtracting the mean and dividing by the standard deviation prior to PCA and clustering. The genotype sets were formed through cluster analysis based on the first four PC scores (Gowda et al., 2011), which resulted in the formation of five sets portrayed by a dendrogram. The trait relations among the genotype sets were then examined within environment groups using PCA to determine if trait relations remained constant across groups.

5.4 Results

5.4.1 Empirical approach

5.4.1.1 G x E interactions within series

The GGE biplots for each series (1999 to 2006) are shown in Figure 5.1. Depending on the trial series, the biplots explained between 76.9 and 89.5% of the variation attributable to genotype and the G x E interaction, respectively. In general, ratoons of the same trial clustered tightly first, before clustering with other trials. Therefore, to assist with biplot interpretations, the genotypic yields were averaged across ratoons in each trial. None of the sites consistently responded as a distinct ME across all series. The 2000, 2001, 2002, 2003, 2004, and 2005 series biplots showed sites separating into two distinct MEs, which did not

necessarily separate according to the B and S selection programs. In fact, the B and S sites frequently grouped into the same MEs, except in the 2004 series, where the two MEs were represented by all B and S sites, respectively. The 1999 series showed most sites falling into one ME with the same winning genotype (N44).   However, this pattern was not repeated in other series.

Site S1 only formed a unique ME in the 2000 series, but otherwise clustered tightly with other sites in the rest of the biplots. Sites S2 and S3 showed a similar response, forming unique MEs in the 2005 and 2006 series, respectively, but clustering with other sites in the rest of the biplots. Of the three S sites, S2 generally showed greater separation from the other S and B sites across the different series, particularly in 1999, 2002, 2003, 2004, and 2005. In general, S2 was located further away from the B sites than S1 and S3, suggesting that it was more effective than S1 and S3 at discriminating genotypes between sandy and humic soils. Site S1 grouped together with S2 and/or S3 in four out of eight series, while sites S2 and S3 grouped together in six out of eight series. This suggests that S2 and S3 are very similar in their ability to discriminate genotype performance and that one of these sites may in fact be redundant. There was frequent correlation between sites S3 and B2, which either fell in the same ME (1999, 2000, 2001, 2003, 2005) or were in very close proximity to each other (2002, 2004, 2006). This implies that site S3 could be dropped from the selection program without much loss of information. This is supported by the fact that S3 did not produce a unique response in seven out of eight series, it clustered frequently with site S2, and showed frequent similarity with B2 in genotype discrimination.
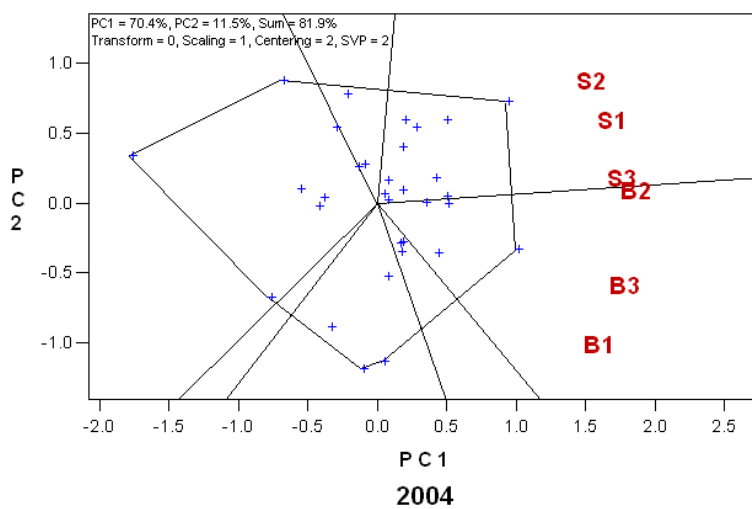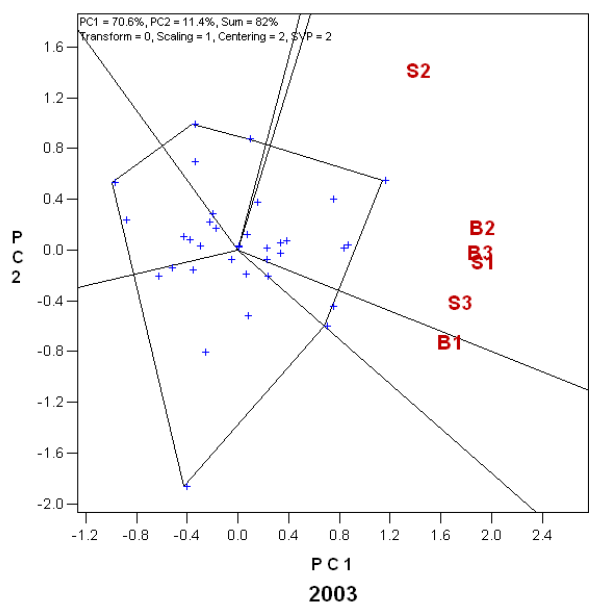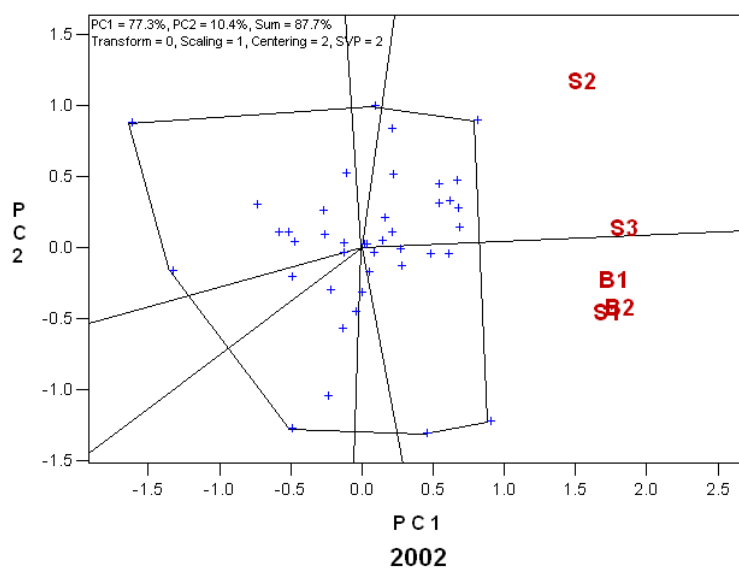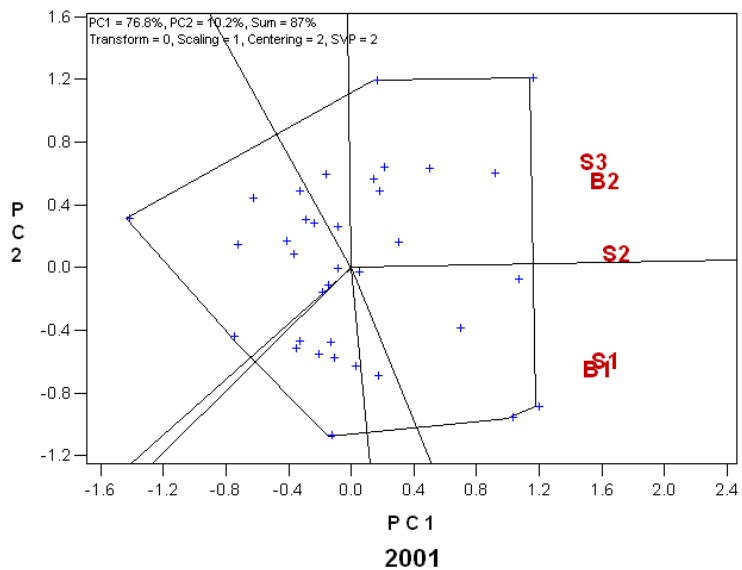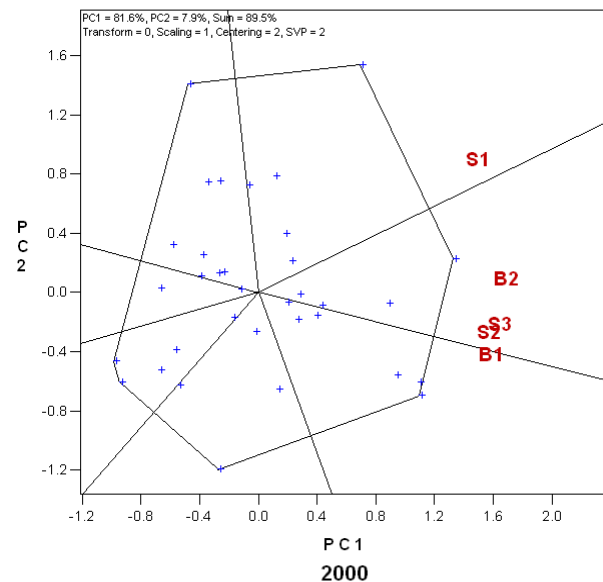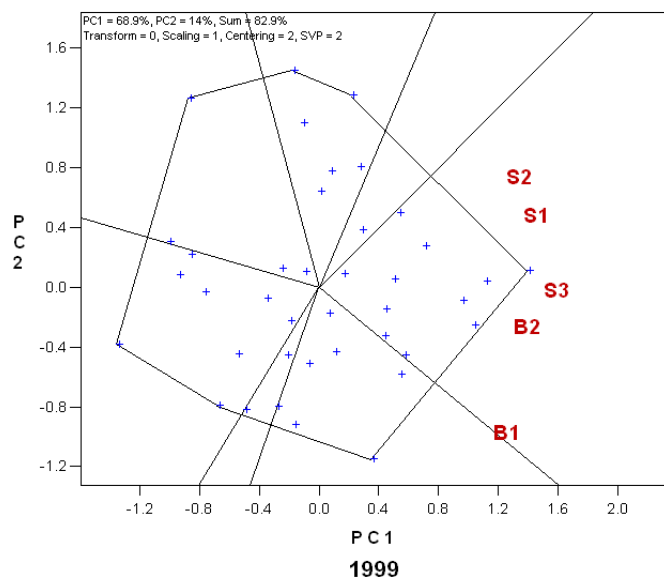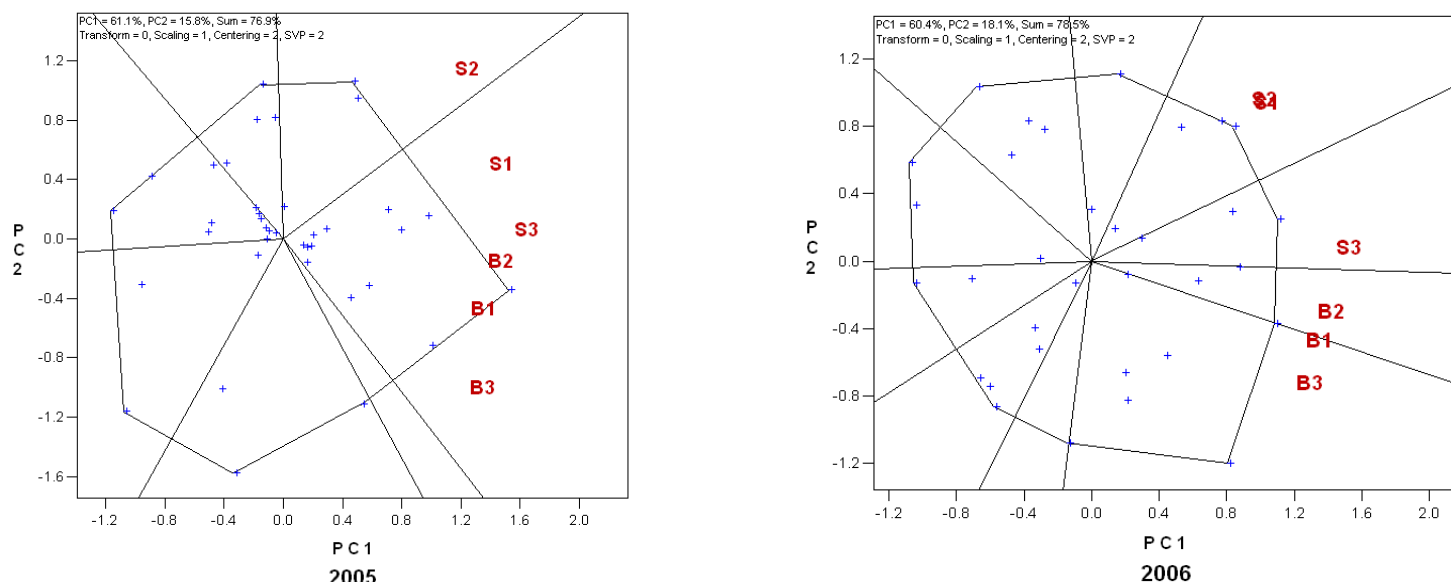
Figure 5.1   GGE biplots for estimated recoverable crystal yield from plant breeding trials established in the 1999-2006 series. Genotypes are depicted by "+" and trial sites are shown in bold red font

In general, the B sites did not show frequent separations from the S sites, and except for the 2004 series, frequently fell within the same ME as some of the S sites. Of the three B sites, site B1 generally showed greater separation from the other B and S sites across the different series, particularly in 1999, 2000, 2001, 2003, and 2004. Sites B1 and B2 fell into the same MEs in four out of eight series and fell into different MEs in the other four series. Although site B3 was only tested in four series, it always grouped with either B1 or B2, and never formed a unique ME. This suggests that B3 may in fact be another redundant site, which could be removed from the program without much loss of information, as it does not provide any unique information on genotype performance.

5.4.1.2 G x E interactions across series

The GGE biplot based on the subset of six commercial genotypes common across all series and post-release evaluation trials (highlighted in bold font), which were averaged across ratoons, showed that the midlands region can be represented by four ME's (Figure 5.2). This biplot showed major overlap between the S and B sites, particularly in the sectors defined by genotypes N44 and N31. Trials F1, F2 and F3, which were post-release trials aimed at identifying genotypes suited to frost pockets, clustered together, forming a unique ME. Consequently, future restructuring of the MET network should consider including these as permanent sites for testing. There was no clear or consistent separation of the S and B sites (which is in keeping with the within-series analysis), however, the majority of the S sites

91

clustered in the N44 and N31 sectors, while a considerable number of the B sites grouped together in the N16 sector. A post-release evaluation trial conducted in the midlands south region (represented by E1), grouped with a number of B environments in the N16 sector, suggesting that this site may not provide further information on variety performances.

In series 1999, 2000, 2001 and 2006, sites S2 and S3 grouped into the N44 sector, while in series 2002, 2003, 2004, and 2005 sites S2 and S3 grouped into the N31 sector (Figure 5.2). This suggests that sites S2 and S3 showed the same switch in winning genotypes when environmental conditions (seasons) changed from one series to the next, and when genotypes changed from one series to the next. This once again highlights the similarities between these two sites which were observed in the within-series analysis, further confirming the site redundancy. Sites B2 and S3 grouped together in the N44 ME in series 1999, 2001 and 2006, and also grouped together in the N31 ME in series 2002 and 2003. This is also in agreement with the within-series analysis, which suggested that site S3 could be dropped without much loss of information. Additionally, the trials in the biplot did not cluster according to the series within which they were planted, suggesting that the effects of season (trials planted in the same series would be subjected to the same seasons) was not as pronounced as the site differences. In general, this biplot depicted the large changes in temporal and spatial environmental conditions across the trial network, which needed to be investigated more intensively.

Figure 5.2 GGE biplot for estimated recoverable crystal (ERC) yield from plant breeding trials established from 1999-2006 across all planting series. The genotypes are depicted by "+" and the trial sites are shown in red font. The trial sites are depicted by S1, S2, S3, B1, B2, and B3, and are preceded by the corresponding series year (9=1999, 0=2000, 1=2001, 2=2002 etc.). Post-release evaluation trials are highlighted in bold black font

5.4.2 Analytical approach

5.4.2.1 Site soil differences

Figure 5.3 represents the average ERD, Clay %, Ncat and OM % of the six selection sites. Site B1 was characterized by fields with higher ERD and Clay % than most of the other sites. Sites B1 and B3 were characterized by similar clay % and NCat, however, B3 showed significantly higher OM % than all sites. The ERD of B2 was not significantly different from B3, however, B2 did show significantly lower levels of Clay %, Ncat and OM% than the other B sites. Site S2 was characterized by the lowest ERD, Clay %, Ncat, and OM% of all sites. Sites B2 and S3 showed no significant differences in any of the parameters measured.

Sites B1 and S1 (the SASRI selection farms) were contrasting in the measured parameters, suggesting that the soil characteristics on these farms are indeed different.
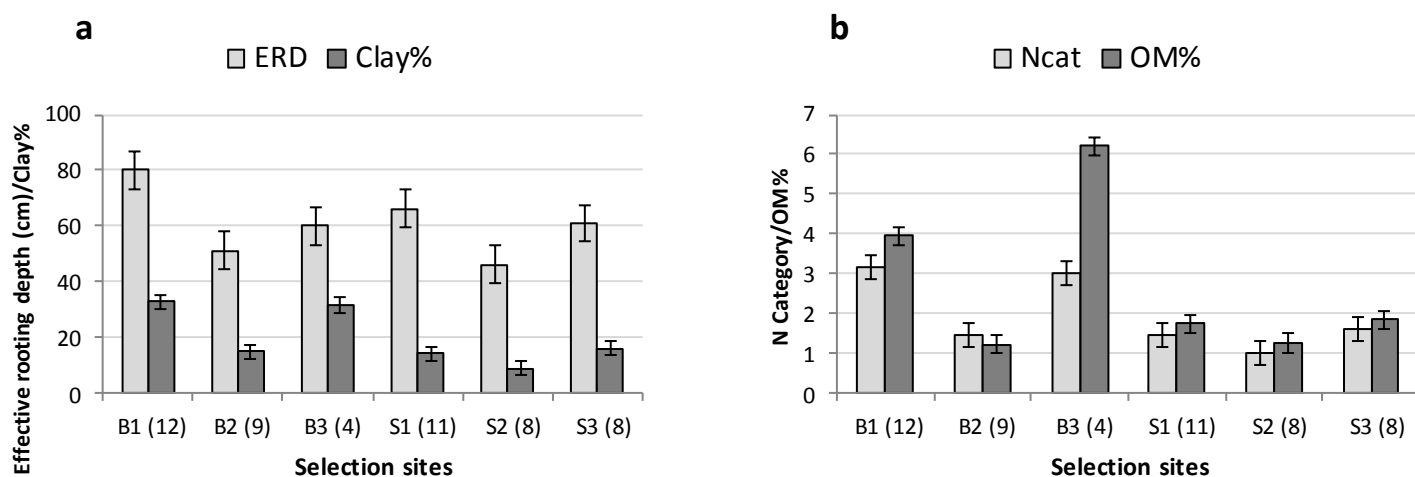
Figure 5.3 Average estimated rooting depth (ERD) and clay percentage (a), and nitrogen mineralization category and organic matter percentage (b) for fields at different selection sites in the midlands region. The number of fields used for trials at each site are indicated in parentheses. Significant differences (P<0.05) are depicted by error bars

5.4.2.2 Genotype sets

The clustering of all 223 genotypes derived from PCA of the trait BLUPs resulted in five genotype sets (dendrograms not shown). In some trials, commercial genotypes were duplicated as treatments, and these duplications were detected by the cluster analysis. For example, genotypes N16, N16a, and N16b clustered together in set A; N37, N37a, and N37b clustered in set B; N31, N31a and N31b clustered in set D; and NCo376, NCo376a, and NCo376b clustered in set E. Despite the duplicates being treated as different varieties in all analyses, the cluster analysis grouped them together appropriately, adding confidence to the clustering strategy used. Based on the mean trait values for each set (Table 5.2) it was deduced that set D was characterized mainly by genotypes with poor quality (ERC), high TCANE and FIB, and that typically produced a sparse canopy, but a high population of thinner, lighter stalks that were prone to stalk bending and lodging. Characteristics of genotypes in set B were generally contrasting to those of set D. Genotypes in sets A and E were generally characterized by moderate to low cane yields with quality characteristics

slightly better than set D genotypes. These genotypes formed denser canopies and produced lower populations of thicker, heavier stalks. Set C genotypes were characterized by contrasting traits to those in sets A and E.

Table 5.2 Mean values for cane yield (TCANE), estimated recoverable crystal percentage (ERC), dry matter percent (DM), fibre percent (FIB), stalk diameter (DIAM), stalk population (POP), stalk mass (STKMS), canopy rating (CAN), skewness rating (SKEW), and lodge rating (LOD) of five genotype sets identified by cluster analysis

Genotype sets	TCANE (t /ha)	ERC (%)	DM (%)	FIB (%)	DIAM (mm)	POP (stalks/ha)	STKMS (grams)	CAN (1-9)	SKEW (1-9)	LOD (1-9)
A	92.68^{a*}	13.09^b	28.29^b	12.88^c	6.59^a	93180^c	830.33^a	2.75^b	3.77^c	4.31^b
B	86.37^b	13.66^a	28.65^a	12.78^c	6.01^b	94138^{bc}	703.97^c	2.72^b	3.75^c	4.27^b
C	85.59^b	12.89^c	28.65^a	13.36^b	5.53^d	96115^b	646.77^c	2.73^b	4.04^b	4.68^a
D	101.46^a	12.47^d	28.55^a	13.67^a	5.82^c	98636^a	775.16^b	2.63^c	4.19^a	4.63^a
E	84.86^b	12.75^c	28.29^b	13.09^c	6.45^a	88994^d	777.97^b	3.01^a	3.55^d	4.13^c

*Values with common letters within columns are not significantly different (P<0.05)

5.4.2.3 Environment groups

The clustering of environments based on the first four PC scores of the G x E matrix across all series resulted in six environment-groups (dendrogram not shown). In general, ratoons of the same trial grouped together first, before grouping with other trials. There was no clear separation between the B and S sites in grouping, however, groups 1, 2, and 3 were dominated by B environments while groups 4, 5, and 6 were dominated by S environments. All environments from post-release trials conducted in frost pockets clustered together in group 2, adding confidence to the clustering strategy used.

Table 5.3 represents the mean values of environment covariates, as well as the average TCANE for each group. In general, it was found that there were no significant differences between groups in TT. Group 4 was the lowest yielding and also had the lowest WSI values. Examination of the groups revealed that environments in group 4 were typically plant crops that were usually established in November and harvested at approximately 20 months of age (average AGE was 20.6 months) around July (average TOH was 3.8) the following year. Group 4 environments were therefore subjected to one summer cycle only before being harvested in the second winter cycle. This may be the primary reason for the greater levels of stress (low average WSI) associated with group 4. Group 2, which consisted mainly of

environments in frost pockets, were characterized by low AGE, lower RAD, and similar EVP compared with the other groups. Despite trials being harvested much younger, environments in group 2 still produced above average yields. This may be attributed to the large proportion of high yielding, fast growing genotypes that are typically included in frost trials to maximize growth in relatively shorter growing periods. Group 2 also showed above average WSI. This may be due to the fact that most of the frost trials are conducted in valley bottom sites (where frost is most common), which are normally characterized by deeper soils that have higher water retention characteristics. Group 3 environments, harvested at above average AGE, showed the lowest EVP, highest RAIN and WSI, and produced the highest average yields compared with the other groups. Group 3 primarily comprised environments from the B sites, which were characterized by more favourable soils (Figure 5.3). This allowed the high rainfall to be effectively captured, resulting in lower moisture stress and higher yields. Groups 1 and 6 had similar EVP, RAD, and WSI, however, they differed slightly in RAIN and AGE. Group 5 environments were slightly warmer than the other groups, with high EVP and RAD, which translated into greater moisture stress (low WSI values).

Table 5.3 Average daily values for thermal time (TT), evaporation (EVP), radiation (RAD), rainfall (RAIN), water stress index (WSI), age at harvest (AGE), time of harvest (TOH) and cane yield (TCANE) of six environment groups identified through cluster analysis

Environment group	TT (°C days)	EVP (mm/day)	RAD (MJ/m^2/sec/day)	RAIN (mm/day)	WSI (0-1)	AGE (months)	TOH (1-9)	TCANE (t/ha)
1	8.03[a*]	3.95[a]	16.78[a]	2.05[a]	0.49[a]	22.41[a]	3.55[ab]	89[b]
2	8.00[a]	3.93[a]	15.80[b]	2.13[a]	0.49[a]	17.03[b]	3.67[a]	90[b]
3	8.07[a]	3.67[b]	16.09[b]	2.21[a]	0.51[a]	22.21[ab]	3.48[b]	115.1[a]
4	8.06[a]	3.77[ab]	16.20[ab]	2.15[a]	0.45[b]	20.6[b]	3.80[a]	76.9[b]
5	8.16[a]	3.96[a]	16.72[a]	2.17[a]	0.47[b]	21.81[ab]	3.82[a]	88.1[b]
6	8.02[a]	3.92[a]	16.69[a]	2.21[a]	0.49[a]	22.05[ab]	3.32[b]	82.3[b]

*Values with common letters within columns are not significantly different (P<0.05)

5.4.2.4 Trait relations within environment groups

In general, the trait relations among the genotype sets were fairly consistent across the six environment groups (Figure 5.4). For each environment-group, there was strong association between DIAM, CAN, and STKMS, which appeared to be strong characteristics of genotype sets A and E. However, these relationships were weaker in environment group 5 (Figure 5.4e). The above traits were generally negatively correlated with DM, POP and FIB, which

were associated with genotype sets D and C across all environment groups. The SKEW and LOD were always positively correlated to each other and both these traits were negatively associated with CAN and DIAM across all environment groups. The ERC was negatively correlated with TCANE in all groups except environment group 4 (Figure 5.4d), where it was correlated with TCANE, TERC, LOD, SKEW, and STKMS. Environment group 4 was the lowest yielding and most stressed (low TCANE and WSI) group, and under these conditions, the relationship between ERC and TERC was stronger compared with other groups, where the overriding effect of TCANE on TERC was more apparent. Genotype set B was characterized by high ERC, irrespective of the environment group. Genotype set D was characterized by high TCANE and TERC in environment groups 1 (Figure 5.4a), 2 (Figure 5.4b), 3 (Figure 5.4c), and 5 (Figure 5.4e), but this relationship was weaker for groups 4 (Figure 5.4d) and 6 (Figure 5.4f). The TCANE was always correlated with POP and STKMS, even though the latter two traits were always negatively correlated with each other. However, in environment group 6 (Figure 5.4f), STKMS showed a stronger correlation with TCANE than POP, suggesting that in such environments (early harvest, high rainfall, no moisture stress, high radiation, and moderate evaporation), TCANE is influenced more by STKMS than by POP. The negative correlation between ERC and FIB was consistent across all groups, while in group 6 ERC was not correlated with any other trait. The positive correlation between FIB and TCANE was not evident in environment group 4 (Figure 5.4d), suggesting that this relationship may not be strong in low yielding environments. The trait relations in environment group 2 (i.e. consisting of trials in frost pockets) did not differ markedly from the other groups, suggesting that there may be limited opportunity for differential trait selection strategies for such scenarios in the midlands region.
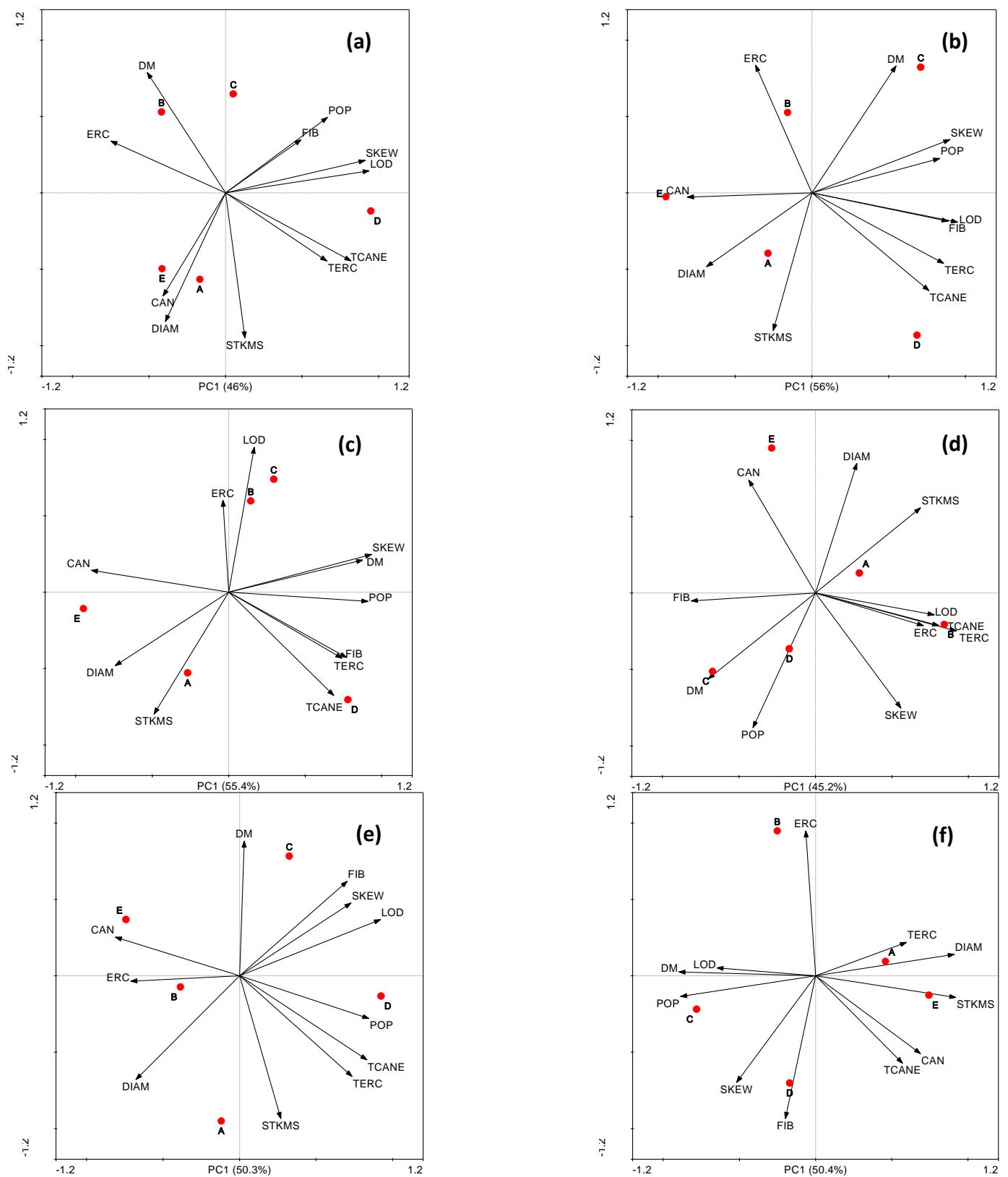
Figure 5.4 Trait relations among five genotype-sets across six environment-groups identified through cluster analysis. The genotype-sets are represented by red points on the biplots, while the traits are represented by arrows. The biplots represent environment-groups 1 (a), 2 (b), 3 (c), 4 (d), 5 (e), and 6 (f) that were identified from cluster analysis. TCANE=cane yield, ERC=estimated recoverable crystal percentage, TERC=ERC yield, DM=dry matter percentage, FIB=fibre percentage, DIAM=stalk diameter, POP=stalk population, STKMS=stalk mass, LOD=lodging, CAN=canopy formation, and SKEW=stalk skewness

5.5 Discussion

The results of the GGE biplot analysis showed that none of the selection sites consistently formed a single ME across all series. All sites were generally correlated with each other to varying degrees in each series, with occasional separation into two MEs. However, such separation of the sites into two MEs was also inconsistent in which sites fell into which ME. These responses suggest that the selection sites in the midlands region represent a single, complex ME with frequent crossover G x E interaction between sites. Similar conclusions were made by Yan and Rajcan (2002), who used GGE biplot analysis and concluded that there was not enough evidence to suggest that an Ontario soybean sub-region could be separated into different MEs. The overall lack of consistency in the ME constitution across the different series may be caused by changes in the genotypes tested in each series, variations in age, TOH, planting and ratooning dates across series, and changes in experimental fields used for trials in each series. Such factors may impact on the consistency of correlations between the trial sites, thereby impacting on ME identification. However, despite this, the analysis still revealed some consistent correlations, which allowed for definitive recommendations on future network design. This includes the redundancy of site S3, which did not produce a unique response in most series, and clustered frequently with sites S2 and B2 in the biplots. It is therefore recommended that site S3 be removed from the network in favour of a more diverse site. Such a site should preferably be within a different geographic zone e.g. the midlands south region of the industry. The practical and logistical implication of such a change will have to be considered by breeders. Additionally, the similarity of site B3 to both B1 and B2 suggests that the relevance of this site to the trial network be re-evaluated after a few more series (B3 was only evaluated across four series in this study).

The four MEs identified in the across-series analysis must be interpreted with caution, as the analysis only comprised six commercial genotypes. Nevertheless, the interpretations were in agreement with the within-series analysis e.g. the redundancy of site S3 and its similarity with site B2. The across-series analysis also suggested the existence of a ME made up of sites in frost pockets in the midlands. Frost-prone areas in the midlands currently represent approximately 15% of the area, and genotypes suited to this production scenario are usually identified through post-release evaluation trials. The scenario involves planting fast growing, high ERC genotypes that can accumulate acceptable TCANE in 12 months, despite cool temperatures. With many growers requesting such genotypes for the midlands in recent times,

it may be worthwhile considering a selection site dedicated to the shorter cutting cycle. Such an initiative may also be seen as a pro-active approach toward mitigating climate change, as temperatures are expected to increase in the midlands moving into the future. The removal of site S3 in favour of a site in a frost prone area is therefore recommended.

Site S2 generally showed greater discriminating ability than the other S sites, as it was located further away from the other S and B sites on the biplots. The ability of S2 to discriminate genotypic performance may be related to the poor soil characteristics associated with this site (Figure 5.3), suggesting that it is an appropriate low potential site that should remain in the network. Additionally, sites S3 and B2 did not differ significantly in ERD, Clay% and Ncat (Figure 5.3). This may be one of the reasons for the frequent similarity between these sites in genotypic performance observed in the GGE biplots. Furthermore, sites S3 and B2 are geographically close, suggesting that they would be exposed to similar climatic influences. Despite the high OM% associated with site B3, this site did not produce differential genotype performance, as B3 frequently grouped with B1 and B2 in the biplots. This shows that OM% may not be as influential as Clay% and ERD in defining the discriminating ability of the site. The above interpretations show that differential site groupings identified through an empirical G x E analysis could be explained to a certain extent by differences in trial site characteristics. Limited studies have attempted to link G x E patterns to actual site differences, and such an approach adds confidence to any recommendations for network redesign that may be required.

The genotype sets that were identified through cluster analysis showed that set D consisted primarily of genotypes where typical *Saccharum spontaneum* traits were more evident, while sets A and E consisted of genotypes where typical *Saccharum officinarum* traits were more evident. Sets B and C may have represented genotypes where traits of the two types were combined. The trait relations observed were generally in keeping with known sugarcane trait associations (Kang et al., 1989). These trait relations varied only slightly from one environment group to another, suggesting that the differences between the environment groups were not sufficient to have major influences on trait relations. The narrow trait variations could also reflect the stability of trait expression as well as the narrow trait variation in the population of varieties in the study. Indeed, the grouping of the test sites into a single ME in many of the biplots is testament to their similarity. Even within environment group 2 (which consisted of trials in frost pockets) there did not appear to be any major deviation from the trait relations, suggesting that differential trait selection strategies for

variable conditions in the midlands may not be very successful. However, the trait analysis did provide insights that will assist in future selection. For example, the strong, consistent correlations between SKEW and LOD, CAN and DIAM, and FIB and POP suggest that there may be duplication in efforts in measuring all these traits. Resources may be better allocated for other trait measurements that are not routinely included in the current protocols. Additionally, the consistent trait correlations mentioned above also suggest that there may be an opportunity for indirect selection using traits that are easier to measure.

The approach in this study was to use genotype sets and environment groups (both derived from pattern analysis) to aid in the interpretation of complex G x E interactions. The fact that the groups identified from G x E patterns showed differences in covariate parameters suggests that the G x E patterns may have been attributable to those covariate differences. However, such differences in the covariates were not explicit enough to allow for the formulation of recommendation domains for the midlands region, and in general, it was difficult to find commonalities between the two grouping approaches used (G x E grouping and covariate grouping). The lack of explicit differences in covariates between the environment groups may be because of the lack of diversity between the environments tested in the study or the limited diversity in genotypes in the study. In general, the midlands region is fairly uniform in soil and climatic factors compared with other regions (coast and hinterland) in the South African sugar industry. Therefore, the approach may be more useful when environments sampled from the complete target population of environments for the industry are analyzed. Additionally, the use of daily averages may not have been the most appropriate scale for the covariates, especially considering that the average ratoon length studied was approximately 24 months. The averaging of data across this two-year cycle would have effectively nullified any intermittent periods of extreme climatic variability that may have affected G x E patterns. Future studies should therefore summarize covariate values within growth phases to identify such effects.

5.6 Conclusions

This study was the first to investigate the similarities between selection sites in the midlands region in the way they discriminate between genotypic performance. In addition to identifying a redundant site (S3) that should be relocated, as well as flagging another site (B3) for possible redundancy, this study has shown that an analytical approach to the G x E study allowed for comprehensive interpretations of the G x E interactions. A simple

characterization of soil factors gave confidence to site similarities suggested by the traditional empirical G x E analysis. A further characterization of environments in relation to covariates helped explain specific similarities/differences between environments. The grouping approach used, although aimed at simplifying interpretations, still required some in-depth scrutiny of patterns. However, coupled with adequate local knowledge of the cropping systems in the region, this method allowed for sensible and realistic interpretations of the G x E patterns. Future studies will apply this approach to more diverse datasets across all regions of the sugar industry. Other studies of sugarcane G x E interactions can use the approaches and principles demonstrated here to enhance the value of METs for the purposes of understanding and exploiting G x E interactions.

5.7 References

Anon., 2000. Information Sheet 7.1: Recommendations for nitrogen. The South African Sugar Association Experiment Station, Mount Edgecombe, Durban, South Africa, pp. 2-3.

Anon., 2009. GenStat® Executable Release 12.1. Lawes Agricultural Trust, Rothamstead Experimental Station/ Clarendon Press, Harpenden, U.K/London, U.K.

Dehghani, H., Ebadi, A., Yousefi, A., 2006. Biplot analysis of genotype by environment interaction for Barley yield in Iran. *Agronomy Journal* 98, 388-393.

DeLacy, I.H., Basford, K.E., Cooper, M., Bull, J.K., McLaren, C.G., 1996. Analysis of multi-environment trials – An historical perspective. p 39-124. *In* M. Cooper and G.L. Hammer (Ed.) Plant Adaptation and crop improvement . CAB International, Wallingford, UK.

DeLacy, I.H., Kaul, S., Rana, B.S., Cooper, M., 2010. Genotypic variation for grain and stover yield of dryland (rabi) sorghum in India: 2: A characterization of genotype x environment interactions. *Field Crops Research* 118, 236-242.

De la Vega, A.J., Chapman, S.C., 2010. Mega-environment differences affecting genetic progress for yield and relative value of component traits. *Crop Science* 50, 574-583.

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

Gauch H.G., 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam.

Gilmour, A.R., Thompson, R., Cullis, B.R., 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51, 1440-1450.

Glaz, B., Kang, M.S., 2008. Location contributions determined via GGE Biplot analysis of multienvironment sugarcane genotype-performance trials. *Crop Science* 48, 941-950.

Gowda, C.L.L., Upadhyaya, H.D., Dronavalli, N., Singh, S., 2011. Identification of large-seeded high-yielding stable Kabuli Chickpea germplasm lines for use in crop improvement. *Crop Science* 51, 198-209.

Hamblin, J., Fischer, H.M., Ridings, H.I., 1980. The choice of locality for plnat breeding when selecting for high yield and general adaptation. *Euphytica* 29, 161-168.

Jackson, P., McRae, T., Hogarth, M., 1995. Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. *Field Crops Research* 43, 119-130.

Kang, M.S., Sosa, O., Miller, J.D., 1989. Path analysis for fiber, cane and sugar yield in sugarcane. *Crop Science* 29, 1481-1483.

Malla, S., Ibrahim, A.M.H., Little, R., Kalsbeck, S., Glover, K.D., Ren, C., 2010. Comparison of shifted multiplicative model, rank correlation, and biplot analysis for clustering winter wheat production environments. *Euphytica* 174, 357-370

Mirzawan, P.D.N., Cooper, M., DeLacy, I.H., Hogarth, D.M., 1994. Retrospective analysis of the relationships among test environments of the Southern Queensland sugarcane breeding programme. *Theoretical and Applied Genetics* 88, 707-716.

Nuss, K.J., 1998. Aspects considered in the search for new farms for the Experiment Station. *Proceedings of the South African Sugar Technologists Association* 72, 42-45.

Patterson, H.D., Thompson, R., 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545–554

Queme, J.L., Orozco, H., Melgar, M., 2010. GGE biplot analysis used to evaluate cane yield of sugarcane (Saccharum spp.) cultivars across sites and crop cycles. *Proceedings of the International Society of Sugar Cane Technologists* 27, 1-7.

Ramburan, S., Van den Berg, M., 2011. Review and analysis of post-release variety evaluation of sugarcane: A South African perspective. *International Sugar Journal* 113, 444-452.

Setimela, P.S., Crossa, J., Banziger, M., 2010. Targeting of early to intermediate maize hybrids for yield performance and stability using SREG model. *South African Journal of Plant and Soil* 27, 207-216.

Sharma, R.C., Morgounov, A.I., Braun, H.J., Akin, B., Keser, M., Bedoshvili, D., Bagei, A., Martius, C., Van Ginkel, M., 2010. Identifying high yielding stable winter wheat genotypes for irrigated environments in central and west Asia. *Euphytica* 171, 53-64.

Van Eeuwijk, F.A., Elgersma, A., 1993. Incorporating environmental information in an analysis of genotype by environment interaction for seed yield in perennial ryegrass. *Heredity* 70, 447-457.

Van Oosterom, E.J., Mahalakshmi, V., Bidinger, F.R., Rao, K.P., 1996. Effect of water availability and temperature on the genotype-by-environment interaction of pearl millet in semi-arid tropical environments. *Euphytica* 89, 175-183.

Voltas, J., Van Eeuwijk, F.A., Sombrero, A., Lafarga, A., Igartua, E., Romagosa, I., 1999. Integrating statistical and ecophysiological analyses of genotype by environment interaction for grain filling of barley I. Individual grain weight. *Field Crops Research* 62, 63-74.

Yan, W., 2001. GGEbiplot – A Windows application for graphical analysis of multienvironment trial data and other types of two-way data. *Agronomy Journal* 93, 1111-1118.

Yan, W., Hunt, L.A., Sheng, Q., Szlavnics, Z., 2000. Cultivar Evaluation and Mega-Environment investigation based on the GGE Biplot. *Crop Science* 40, 597-605.

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

Zhe, Y., Lauer, J.G., Borges, R., de Leon, N., 2010. Effects of genotype x environment interaction on agronomic traits in soybean. *Crop Science* 50, 696-702.

CHAPTER 6

A GROWTH PHASE APPROACH TO INTERPRET SUGARCANE GENOTYPE X ENVIRONMENT INTERACTIONS: INVESTIGATING TIME OF HARVEST EFFECTS[4,5]

6.1 Abstract

Environmental conditions during specific growth phases of sugarcane may be a determinant of complex G x E interactions. The objective of this study was to employ multivariate analysis techniques to evaluate the appropriateness of a growth phase approach to analyze sugarcane cultivar x time of harvest (TOH) interactions, and to assess the feasibility of these methods being applied to complex G x E datasets in future. Two field trials consisting of the same set of seven commercial cultivars were established on adjacent fields and harvested either early (April/May) or late (October/November) in the season for six ratoons. Climatic data were summarized within individual growth phases (establishment, elongation, and ripening) and used to create environmental covariates for thermal time (TT), rainfall (RAIN), radiation (RAD), and a water stress index (WSI). Additive main effects and multiplicative interaction (AMMI) analysis, principal components analysis (PCA), factorial regression, and covariate-effect biplot analysis were used to analyse and interpret the G x E interaction for TERC, TCANE, and ERC.

The G x E interactions were significant (p<0.001) for all three variables. The AMMI biplots showed distinct separations between environments, and characterized cultivars according to their adaptability to early (NCo376, N36, N35, N29) and late (N27, N19, N17) harvesting. The PCA and factorial regression identified RAD and TT as significant environmental covariates affecting TOH, while WSI and RAIN had insignificant effects. The effects on G x E interactions were greatest when these covariates varied within the stalk elongation phase. The covariate-effect biplot identified correlations between plant population and early

---

[4]Ramburan, S., 2011. Sugarcane cultivar x time of harvest interactions in South Africa. *South African Journal of Plant and Soil* 28, 75-84.

[5]Ramburan, S., 2011. Interpreting sugarcane varietal adaptability to time of harvest. *Proceedings of the South African Sugar Technologists Association* 84, 375-388.

harvests, while stalk mass was correlated to late harvests. These findings showed that a growth phase approach coupled with various multivariate methods is appropriate for comprehensive analyses of sugarcane G x E interactions. Future studies will apply these concepts to larger G x E datasets.

6.2 Introduction

The shortcomings of using environmental covariates averaged across a cropping season to characterize METs were highlighted in the previous two studies (Chapters 4 and 5). Although providing valuable information on the relative effects of covariates on G x E interactions, this method did not allow for more comprehensive interpretations that may have value beyond the realms of conventional G x E research. For example, a simple measurement of average rainfall at trial sites may help explain site differences and variable genotypic performance across sites. However, the value of this information to other disciplines such as crop physiology is limited if the distribution of rainfall and specific water stress events across the season cannot be captured from these observations. Most crop improvement programs (and other research disciplines) are under resource constraints, and the expansion of programs to include selection and evaluation under specific limiting biotic and abiotic factors is seldom feasible. Therefore, more effort must be invested in extracting greater value from METs, not just as tools for crop improvement, but also to inform other research disciplines. It is in this context that a growth phase approach to interpreting sugarcane G x E interactions is envisaged. This concept is investigated in this chapter using genotype x time of harvest (TOH) interactions observed in a cultivar evaluation trial.

In the South African sugar industry, cane is harvested and milled over a nine-month period between April and December. The TOH refers to the period within a milling season when the crop is harvested, and this can be categorized into early (April-June), mid (July-September), and late (October-December) season (Ramburan et al., 2010). Cultivar x TOH interactions are known to exist, however, this receives less attention during selection due to resource constraints. The effects of cultivar x TOH interactions on sugarcane productivity have been investigated in isolated studies. Gilbert et al. (2006) demonstrated highly significant differences in TOH and significant G x E interactions in three case studies with sugarcane in Florida. In Mauritius, Julien and Delaveau (1977) showed that dry matter partitioning in sugarcane was influenced to a greater extent by TOH than by location, while Nayamuth et al. (2005) proposed that genotypes could be classified into three distinct maturity groups (early,

mid, and late) based on their sucrose accumulation patterns.   In Australia, Di Bella et al. (2009) found significant cultivar x TOH interactions for cane and sugar yields in ratoon crops at multiple locations. Similarly, other examples of cultivar x TOH interactions in South Africa are available (Parfitt, 2000; Thomas & Parfitt, 2000; Donaldson et al., 2008; Ramburan et al., 2010). The previous studies in South Africa have focused on TOH effects in the irrigated northern regions, while no attempts have been made to evaluate the cultivar x TOH interactions in the southern rainfed regions, where the additional effects of water stress must also be considered.

Studies on cultivar x TOH interactions have focused on cultivar characterization and have also acted as decision aids for the design of sugarcane breeding programmes (Gilbert et al., 2006). Few dedicated attempts have been made to explain and interpret such interactions in relation to major environmental factors. Donaldson et al. (2008) found that early season crops intercepted lower fractions of incident radiation compared with late season crops in South Africa and also suggested that differences in biomass production between early and late harvests were due to lower radiation use efficiencies and premature ripening of late season crops. Singels et al. (2005b) reviewed biomass production of sugarcane in South Africa and also highlighted important temperature and radiation influences on TOH variability. These detailed studies explained TOH variability and interpreted cultivar adaptability to TOH by focusing largely on single factors and their influence on important crop development traits and biomass partitioning concepts used in crop modelling. The relative contributions of different environmental factors to the cultivar x TOH interaction, and their effects on commercial production have not been evaluated. Studies involving the interpretation of G x E interactions through use of environmental covariates (Vargas et al., 1999; Yan & Kang, 2003) could contribute to the understanding of TOH effects on commercial sugarcane production.

Currently there is uncertainty about the adaptability of new commercial cultivars to TOH, and the reasons for production differences between early and late season harvests in rainfed regions in South Africa remain unclear. Because of limited selection for TOH and confounding effects of site, crop age, and season during selection, there is consequently a need for the objective characterization of commercial cultivars post-release. Additionally, information on the environmental drivers characterizing early and late season harvests and cultivar responses to these drivers will assist in understanding their physiological/genetic make-up and help direct future breeding efforts for TOH. Furthermore, little is known about the individual contribution of yield components to the cultivar x TOH interactions.

Multivariate analysis techniques have been commonly employed to interpret cultivar growth phase sensitivities to environmental factors in crops such as ryegrass (Van Eeuwijk & Elgersma, 1993), wheat (Voltas et al., 2005), barley (Voltas, et al., 1999), and pearl millet (Van Oosterom et al., 1996). Analysis of such cultivar sensitivities may identify specific growth phases (and environmental factors) responsible for production variability between early and late harvests of sugarcane in South Africa. If successful, such multivariate methods could be extended to include more complex G x E datasets, where a large number of environmental covariates are summarized within growth phases to interpret G x E interactions. The main objective of this study was to evaluate the appropriateness of a growth phase approach to analyze sugarcane cultivar x TOH interactions and assess the feasibility of these methods being applied to complex G x E datasets in future. A secondary objective was to characterize new commercial cultivars in relation to TOH adaptability, and investigate the relative effects of yield components to the cultivar x TOH interaction through the application of multivariate analysis techniques.

6.3 Materials and methods

6.3.1 Field trials and weather variables

Two field trials were conducted on adjacent fields on a SASRI experimental farm (28°43'S, 31°53'E, 102 m.a.s.l.) on the north coast of Kwa-Zulu Natal province, South Africa. The trials were established under rainfed conditions in randomized complete-block designs with six replicates and seven cultivars in 57.6 m^2 plots (6 rows, 8 m long and 1.2 m row spacing). The treatments comprised new (N36, N35, N29, N27) and established (N19, N17, NCo376) cultivars, whose relative adaptabilities to different harvest times were unknown. One trial was established in November 2000 and harvested annually in the late season (October/November) for six successive ratoons (regrowth after harvest), while the other trial was established in March 2001 and harvested annually in the early season (April/May) for the same duration. Both trials were established on a deep Hutton soil, with 41% clay and a total available moisture of 48 mm. Nutrient and weed management proceeded as required, according to routine soil analyses and weed dispositions, respectively.

At each harvest, net cane plots of 38.4 m^2 were cut and weighed using a mechanical grab apparatus attached to an electronic scale to determine TCANE. Plot samples of 12 stalks were taken from the trials at each harvest to determine a range of quality traits, the most important of which was ERC. The TERC was calculated as the product of ERC and TCANE. Other

relevant cultivar traits determined at the plot level at each harvest included stalk population (POP) in stalks/ha, stalk mass (STKMS) in g/stalk, and stalk length (STKLTH) in cm. The POP was determined by counting the number of mature stalks per 8 m row and expressing that on a per hectare basis, STKMS was determined from the weight of a 12-stalk sample, while STKLTH was determined by measuring the length (from base to natural breaking point) of 20 stalks from each plot.

Daily weather data from an on-farm weather station were recorded during each crop cycle. The average duration of three growth phases, namely crop establishment (1), stalk elongation (2), and ripening (3) were estimated from previous observations (unpublished data[6]), and the weather data within each phase were then summarized as environmental covariates. These covariates included average daily rainfall (RAIN) in mm, average daily solar radiation (RAD) in MJ m^{-2} sec^{-1}, thermal time (TT) in heat units (base 10°C), and a water stress index (WSI). The WSI was calculated by considering the total available moisture from soil depth and clay content. This information was fed into the Canesim crop growth model (Singels & Donaldson, 2000), which utilized climatic data from the weather station and estimated the actual evapotranspiration (EVT$_{act}$) for each crop cycle. The model was also run for each crop cycle as if fully irrigated to estimate potential evapotranspiration (EVT$_{pot}$) had there been irrigation. The EVT$_{pot}$ represented unstressed crop growth and the WSI was calculated as the ratio of EVT$_{act}$: EVT$_{pot}$, where a high WSI represented no moisture stress, while values closer to zero represented higher levels of stress. Table 6.1 contains a summary of the daily averages for the environmental covariates for each crop cycle and illustrates how they originated. For example, RAIN1, RAIN2 and RAIN3 refer to average daily rainfall during establishment, elongation and ripening, respectively. Similar designations were used for the other covariates. Figure 6.1 contains the long-term mean temperatures, solar radiation, and rainfall for the trial site and illustrates the approximate harvest/ratoon times for the early and late trials.

---

[6] MA Smit, South African Sugarcane Research Institute, Mount Edgecombe, South Africa, 2010.

Table 6.1 Mean daily values of rainfall (mm) thermal time (Heat units), radiation (MJ/m^2/sec), and water stress index (0-1), during three growth phases (establishment, elongation, ripening) of early and late harvests

Covariate	Environments											
	Early harvests						Late harvests					
	E1	E2	E3	E4	E5	E6	L1	L2	L3	L4	L5	L6
Rainfall (mm)												
Establishment (RAIN1)	1.44	1.27	1.03	1.16	1.71	3.20	1.83	1.84	3.86	1.53	2.16	2.48
Elongation (RAIN2)	1.86	3.87	1.58	2.17	2.83	3.00	1.83	1.84	1.35	1.83	3.29	3.57
Ripening (RAIN3)	1.94	3.85	2.24	4.12	2.51	2.18	1.42	1.12	1.51	1.06	1.49	1.85
Thermal time (Heat units)												
Establishment (TT1)	10.03	10.08	9.61	10.85	9.99	9.49	14.48	14.46	14.90	15.33	14.10	14.71
Elongation (TT2)	14.13	14.37	15.91	14.41	14.94	14.04	12.07	12.13	11.98	13.04	12.45	12.58
Ripening (TT3)	14.02	13.49	14.05	13.91	14.19	12.80	10.40	10.33	9.53	10.76	10.20	9.73
Radiation (MJ m^{-2} sec^{-1})												
Establishment (RAD1)	13.13	12.74	14.36	12.74	12.97	12.97	19.59	19.58	20.35	21.11	18.92	21.19
Elongation (RAD2)	19.27	17.96	21.66	18.90	21.73	20.32	14.29	14.36	14.77	15.50	15.12	15.05
Ripening (RAD3)	16.67	14.89	16.52	16.13	16.73	17.58	14.34	13.34	14.93	12.37	13.19	13.77
Water stress index (0-1)												
Establishment (WSI1)	0.45	0.39	0.61	0.02	0.70	0.80	0.44	0.37	0.58	0.44	0.53	0.53
Elongation (WSI2)	0.09	0.36	0.24	0.002	0.31	0.42	0.04	0.08	0.26	0.28	0.59	0.45
Ripening (WSI3)	0.06	0.82	0.37	0.002	0.40	0.55	0.21	0.14	0.30	0.18	0.31	0.45
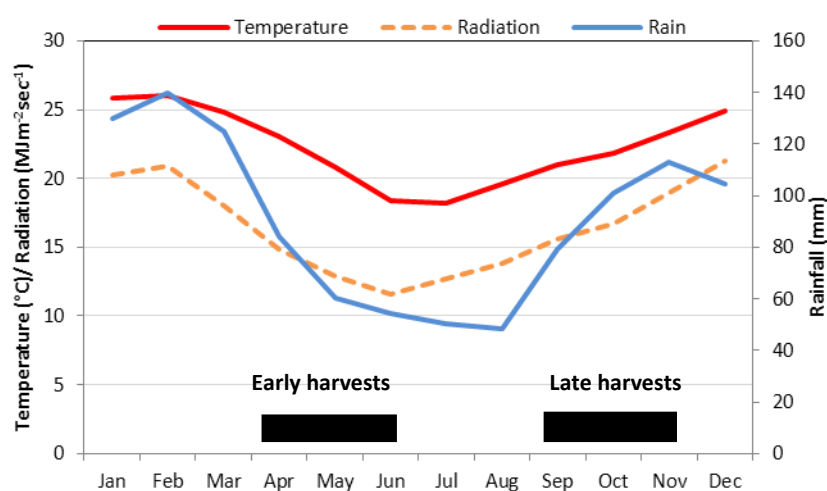
Figure 6.1 Mean temperatures, solar radiation and rainfall for the trial site, with the approximate harvest windows for the early and late season trials

6.3.2 AMMI analyses

A trial and ratoon combination was considered an environment, which resulted in a total of 12 environments, designated E1 to E6 (early season) and L1 to L6 (late season). The G x E interaction for TCANE, ERC and TERC was analysed using the AMMI model, as described in previous chapters. For each of the relevant variables, AMMI2 biplots of interaction principal component axes (IPCA)1 and two scores for genotypes and environments were constructed to help visualize the TOH adaptability of cultivars and to determine if these responses were consistent across seasons. The AMMI2 biplot is commonly used in the analysis of METs and is derived from the singular value decomposition of the matrix of interaction residuals (Gauch, 1992). In these biplots, cultivars and environments are depicted as points in a two dimensional space and their coordinates are defined by their IPCA1 and IPCA2 scores. Here, cultivars and environments are represented by arrows and points, respectively. The distance from the origin of the biplot represents the amount of interaction that is exhibited by the respective cultivar or environment. The angles between two cultivar arrows correspond to their correlation. In general, acute angles represent positive correlations, right angles represent no correlation, and obtuse angles represent negative correlations (Voltas et al., 1999). Similar interpretations hold true for environments. The magnitude of interactions between cultivars and environments can be interpreted by their respective vector direction, where similar directions indicate positive interactions and vice versa (Gauch, 1992).

6.3.3 Interpretation of G x E interactions using environmental covariates

In order to identify the most relevant environmental covariates and reveal possible co-linearity, a preliminary PCA was conducted on the environment x covariate two-way table. This analysis helped visualize the relations and possible redundancies in the covariate dataset. The results of the PCA were further supplemented by the correlation of the AMMI IPCA1 and IPCA2 environment scores with the environmental covariates (Van Eeuwijk and Elgersma, 1993). These techniques aided to limit the number of environmental covariates to the most relevant and provided candidates for inclusion in factorial regression analysis.

The general form of a factorial regression model based on a two-way G x E table with concomitant environmental variables was described by model 7 in Chapter 2. Factorial regression can be used to describe the G x E interaction as differential sensitivity of genotypes to measured environmental variables (Voltas et al., 1999). After fitting the main effects of genotype and environment, associated variables on the levels of the environmental

factor can then be introduced into the model in a multiple regression approach. One of the advantages of factorial regression is that hypotheses about the influence of associated environmental variables on the G x E interaction can be statistically tested and viewed within an ANOVA table.

6.3.4 Relationships among yield components

In an attempt to explain the G x E interaction of TCANE relative to cultivar traits, a covariate-effect biplot was produced (Yan and Tinker, 2005). This biplot is produced through the singular value decomposition of a trait x environment two-way table of correlation coefficients between TCANE and the relevant trait in the relevant environment. A comprehensive description of this type of biplot is given by Yan and Tinker (2005). With this technique, the correlation coefficients were used as a measure of the effects of the explanatory traits (POP, STKMS, STKLTH) on TCANE, and this helped determine the extent to which TCANE was influenced by different traits at different harvest times. All statistical analyses were conducted using Genstat® Version 12.1 statistical software (Anonymous, 2009) and biplots were produced using Canoco® for Windows Version 4.51 (Ter Braak and Smilauer, 2003).

6.4 Results

6.4.1 AMMI analysis

The AMMI2 analysis of variance indicated that environments (E), genotypes (G), and the G x E interactions were highly significant (P<0.001) for all three variables (Table 6.2). The E accounted for the largest amount of variation, comprising 71, 61 and 64% of the total sums of squares (SS) for TCANE, ERC, and TERC, respectively. The SS for G was greater than the G x E SS for TCANE and ERC, while these two components had similar contributions to variation in TERC. Both the IPCA1 and IPCA2 axes were also significant for the three variables, and although explaining only a small proportion of the total SS, they did explain the majority of variation of the G x E interactions. For TCANE, IPCA1 and IPCA2 explained 56 and 23% of the G x E interaction, respectively. The relative proportions of the G x E interactions explained by IPCA1 (32%) and IPCA2 (28%) for ERC were more balanced, while 53 and 23% of the G x E interaction for TERC was explained by IPCA1 and IPCA2, respectively.

Table 6.2 AMMI2 analysis of variance for cane yield (TCANE), estimated recoverable crystal percentage (ERC) and tons ERC (TERC), including the first two interaction principal component analyses (IPCA) axes

Source	df	SS	MS	%SS[a]	SS	MS	%SS	SS	MS	%SS
			TCANE			ERC			TERC	
Total	503	218310	434		2086.8	4.15		3611	7.18	
Genotype (G)	6	11674	1946**	5.3	256.8	42.8**	12.3	218	36.35**	6.0
Environment (E)	11	154293	14027**	70.7	1281	116.46**	61.4	2307	209.74**	63.9
G X E	66	10466	159**	4.8	118.7	1.8**	5.7	224	3.39**	6.2
IPCA 1	16	5853	366**	2.7	38.4	2.4**	1.8	118	7.38**	3.3
IPCA 2	14	2427	173**	1.1	33.6	2.4**	1.6	50	3.57*	1.4
Residuals	36	2186	61		46.7	1.3		56	1.55	
Error	360	29534	82		378.6	1.05		609	1.69	

*$P \leq 0.05$, **$P \leq 0.01$

[a] % of total sums of squares for each term or interaction

The AMMI means of each cultivar in each environment revealed that the average TCANE of the early harvests were generally higher than the later harvests; however, the early TOH had greater variability as a whole (not shown). Cultivars that produced high TCANE in the early and late season were N36, N27 and N17. The late season harvest produced considerably higher ERC levels than the early season, with cultivars N29, N35, and N36 performing the best at both harvest times. Although TCANE was greater for early harvests, the higher ERC of the late harvests had an overriding effect on the final TERC, which was greater at the latter harvest time.

The AMMI2 biplot for TCANE, which captured 79% of the total G x E interaction, revealed two distinct non-overlapping clusters for the early and late season harvests (Figure 6.2a). These groupings were spread more explicitly across IPCA1. Cultivars N17, N19, and N36 demonstrated clear adaptation to the late season harvest, while cultivars N27, N29, N35, and NCo376 interacted positively with the early season harvest. The ERC biplot (Figure 6.2b) also showed TOH separation along the IPCA1 axis, however, significant spread was also observed along the second axis with no clear distinction between early and late harvests. Here, the varieties N29, N35, and N36, which are all characterized as high ERC varieties in the industry, were positively correlated with each other. For the TERC biplot (Figure 6.2c), the first two axes accounted for 75% of the G x E interaction and also showed distinct clustering of the early and late harvests along the IPCA1 axis. Cultivars N17, N19, and N27

showed similar responses to late season harvesting, while varieties NCo376, N29, N35, and N36 interacted positively with the early season.
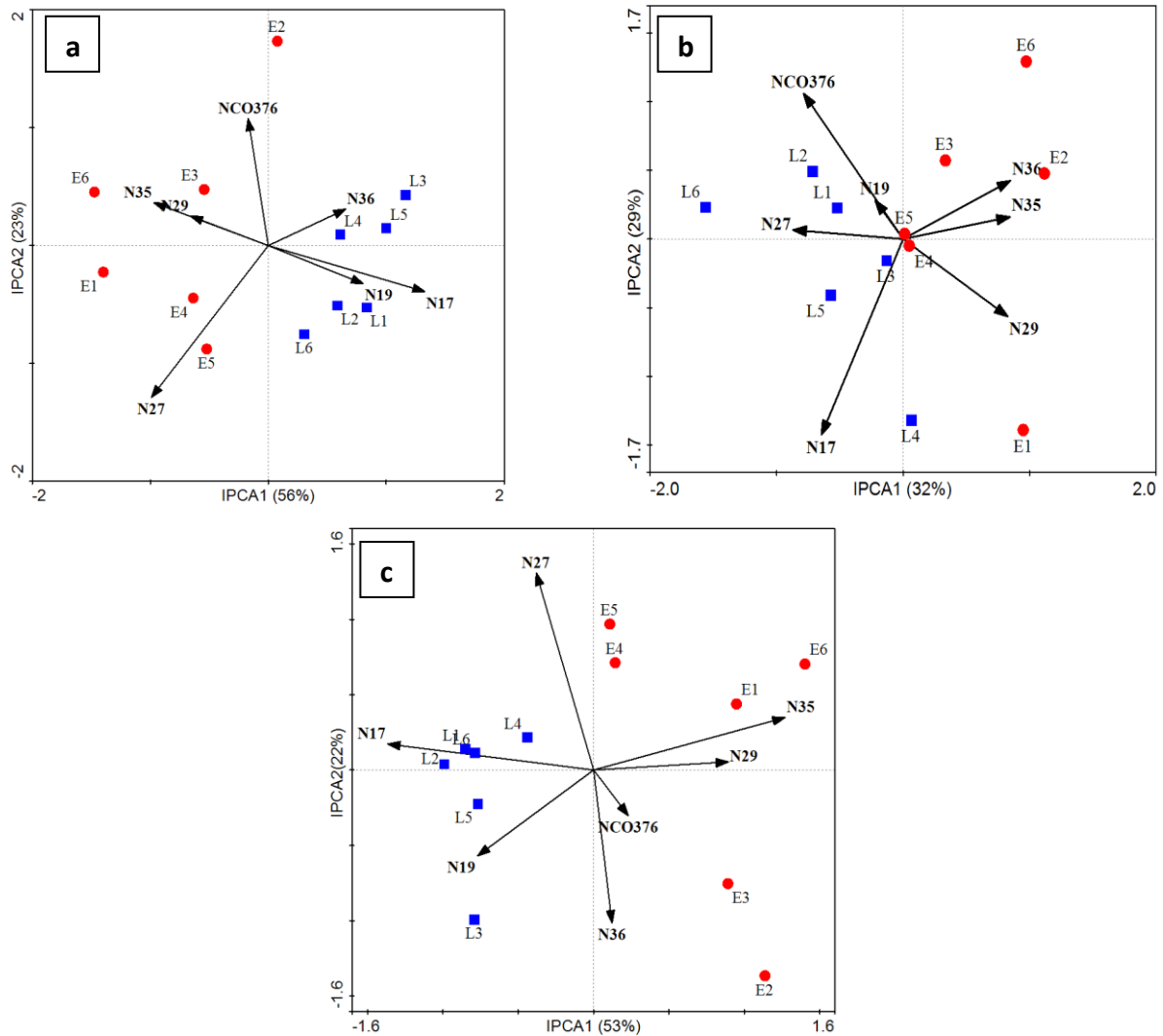
Figure 6.2 AMMI2 biplots for TCANE (a), ERC (b), and TERC (c). Varieties are represented by arrows while environments are represented by points. Early season harvests are represented by red circles and late season harvests are represented by blue squares. (The percentage of the interaction explained by the IPCA axes are indicated in parentheses in the axis titles)

6.4.2 Identification and interpretation of environmental covariates

6.4.2.1 PCA analysis

The PCA of the environment x covariate two-way table helped identify covariates that had the greatest influence on TOH variability (Figure 6.3). The first two PCA axes for the environmental covariates accounted for 93% of the total variation, with most of the variation

accounted for by the first axis (88%). This suggests that the wide spread of covariates along the PC2 axis in Figure 6.3 is rather misleading when considering the total variation in the environment x covariate table. It also suggests that covariates spread along the PC1 axis have a greater effect on variability among the harvest times. The TT1 and TT3 were negatively correlated (as indicated by the opposite directions of their vectors). This was expected, as high temperatures during establishment (TT1) of a late season crop is often accompanied by lower temperatures during ripening (TT3) of the crop the following winter. The RAIN3 and WSI3, together with RAIN1 and WSI1 were highly correlated, as high rainfall often implies less water stress (hence high WSI values). The RAD2, RAD3, TT2, and TT3 were all correlated, with RAD2 accounting for the greatest amount of variation (longest vector) thereby suggesting that it could be a representative of that group of covariates for subsequent analyses. Similarly, the PCA suggested that RAD1, RAIN3, and WSI1 could be used as the other candidate covariates for factorial regression.
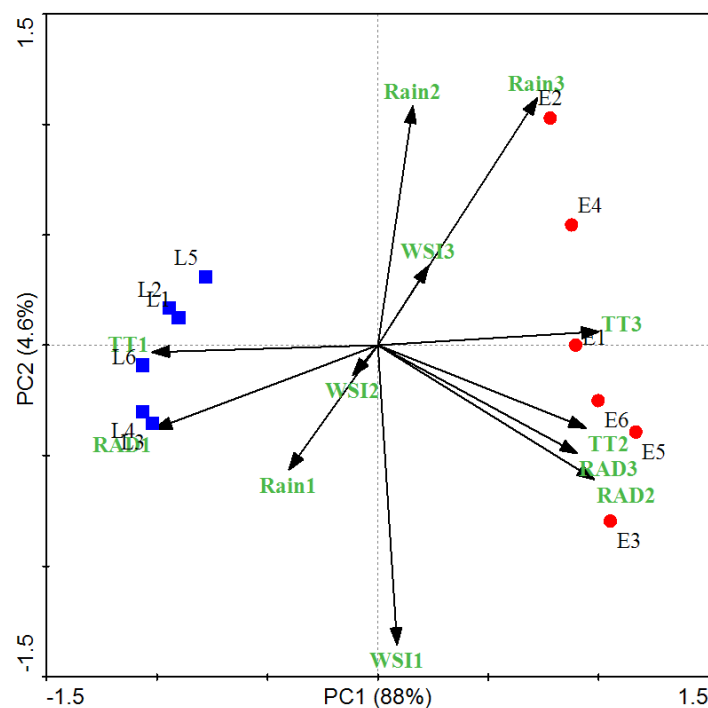
Figure 6.3 Principal component biplot of the environment x covariate two-way table. Environmental covariates are in green and depicted by arrows. Early and late harvests are depicted as red circles and blue squares, respectively. (The percentage of variation explained by the PC axes are indicated in parentheses in the axis titles)

115

6.4.2.2 Correlations with IPCA scores

The covariates analysed with PCA were also evaluated for significance through correlation with the environment IPCA1 and IPCA2 scores obtained from AMMI analysis (Van Eeuwijk & Elgersma, 1993). Table 6.3 contains the correlations between IPCA scores and environmental covariates for TCANE, ERC, and TERC. Based on the clustering of early and late season environments according to IPCA scores in Figure 6.2a, and the correlations between the covariates and IPCA scores in Table 6.3, a number of biological interpretations can be made. For TCANE, the significant positive correlations with IPCA1 of RAD1 and TT1 suggest that early season harvests (low IPCA scores in Figure 6.2a) are characterized by lower average daily radiation and average daily heat units during the establishment phase. In contrast, late season harvests (high IPCA scores in Figure 6.2a) were characterized by higher levels of RAD1 and TT1. The significant negative correlations with IPCA1 of RAD2, RAD3, TT2 and TT3 can be interpreted in a similar way. Late season harvests (high IPCA scores) were characterized by lower average daily radiation and average daily heat units during the elongation and ripening phases, while early season harvests (low IPCA scores) were characterized by higher levels of these covariates. These interpretations can be confirmed through examination of Figure 6.1, where it is shown that late season harvests encountered summer conditions during the early growth phases, while early season harvests experienced summer conditions during later phases of growth.

On the basis of correlation analysis, it can be deduced that early season harvests generally show greater TCANE than late harvests because of more favourable radiation and temperature conditions during the elongation and ripening stages. In sugarcane, most yield accumulation occurs during the stalk elongation phase (Robertson et al., 1996); therefore, favourable conditions during this phase will result in greater exploitation and net gain in yield compared with favourable conditions during any other growth phase. It follows that late season harvests produce lower TCANE because of poor radiation and temperature conditions during elongation.

The separation of environments along IPCA1 was not distinct for ERC (Figure 6.2b), with occasional overlap between early and late harvests being observed. In general, however, early season harvests had higher IPCA1 scores compared with late season harvests. An analysis of correlations (Table 6.3) suggested that late season harvests were characterized by lower RAD2, RAD3, TT2, and TT3 (significant positive correlations with IPCA1) compared with early harvests. The poor radiation and temperature conditions experienced during winter are

known to improve the ERC% of sugarcane in South Africa, and these results confirm these observations. Conversely, the early harvests were characterized by higher RAD2, RAD3, TT2, and TT3, which stimulated vegetative growth during summer, thereby utilizing stored sucrose and reducing ERC%.

Correlations for TERC were similar to ERC, with significant negative correlations of IPCA1 with RAD1 and TT1 and significant positive correlations with RAD2, RAD3, TT2, and TT3 (Table 6.3). These interpretations correspond to explanations derived from the analyses of TCANE. Early season harvests were characterized by higher RAD2, RAD3, TT2, and TT3 compared with late harvests. However, the TERC of early harvests were generally lower than the late harvests (not shown), suggesting that ERC had a more dominant influence on TERC than TCANE. In addition, a significant positive correlation with RAIN3 was observed, suggesting that late season harvests (lower IPCA1 scores) were characterized by lower rainfall during ripening compared with early harvests (this can be confirmed from Figure 6.3). Low moisture during ripening of late season harvests created water-deficit stresses, thereby stimulating the accumulation of sucrose and higher ERC% compared with early harvests.

The spread of early season harvests along IPCA2 for TCANE and TERC suggests greater variability in water stress during ripening compared with the late season harvests, which tended to cluster more closely along IPCA2 (Figures 6.2a,b). For example, a comparison of environments E2 and E5, which were widely spaced along IPCA2 in Figure 6.2a, indicates that E2 had a WSI3 value of 0.82 (low water stress) compared with a value of 0.4 for environment E5. A similar interpretation can be made for environments E1 (WSI3 = 0.06) and E6 (WSI3 = 0.55) in Figure 6.2b. In general, these interpretations suggest that IPCA2 captured the effects of water stress during ripening.

6.4.2.3 Factorial regression
For brevity, only the factorial regression analysis of TCANE is reported here. Based on the significance of correlations with IPCA scores, the magnitude of interaction accounted for from the PCA, and the levels of colinearity observed in the PCA, only a few relevant covariates were chosen for factorial regression analyses. Various combinations of these selected covariates were attempted in the factorial regression model to maximize the proportion of G x E explained. The model that accounted for the greatest amount of variation, with an acceptable number of covariates included the terms RAD2, TT1 and RAIN3 (Table

6.4). Still, of these covariates, only RAD2 was significant (P<0.001). Genotypic sensitivity analyses showed that cultivars N17, N19, and N36 showed positive responses to RAD2 (radiation during stalk elongation), while cultivars N27, N29, N35, and NCo376 responded negatively (Table 6.5). A similar grouping of the cultivars can be observed in Figure 6.2a, suggesting that the differential responses of varieties to radiation during stalk elongation may be one of key factors influencing variability in time of harvest.

Table 6.3 Correlation coefficients between radiation (RAD), rainfall (RAIN), thermal time (TT) and water stress index (WSI) during three growth phases (establishment (1), elongation (2), and ripening (3)) and IPCA1 and IPCA2 scores from AMMI analysis for cane yield (TCANE), estimated recoverable crystal percentage (ERC), and tons ERC (TERC)

Covariate	IPCA1			IPCA2		
	TCANE	ERC	TERC	TCANE	ERC	TERC
RAD1	0.83^{**}	-073^{**}	-0.87^{**}	-0.27^{ns}	-0.32^{ns}	-0.19^{ns}
RAD2	-0.84^{**}	0.61^{*}	0.82^{**}	0.12^{ns}	0.27^{ns}	0.26^{ns}
RAD3	-0.84^{**}	0.57^{*}	0.77^{**}	0.07^{ns}	0.39^{ns}	0.27^{ns}
RAIN1	0.25^{ns}	-0.21^{ns}	-0.26^{ns}	0.02^{ns}	0.19^{ns}	-0.07^{ns}
RAIN2	-0.09^{ns}	-0.03^{ns}	0.21^{ns}	0.23^{ns}	0.28^{ns}	-0.03^{ns}
RAIN3	-0.46^{ns}	0.39^{ns}	0.57^{*}	0.32^{ns}	0.32^{ns}	-0.04^{ns}
TT1	0.87^{**}	-0.71^{**}	-0.92^{**}	-0.29^{ns}	-0.38^{ns}	-0.15^{ns}
TT2	-0.75^{**}	0.59^{*}	0.81^{**}	0.23^{ns}	0.22^{ns}	0.09^{ns}
TT3	-0.82^{**}	0.70^{*}	0.83^{**}	0.17^{ns}	0.15^{ns}	0.22^{ns}
WSI1	-0.14^{ns}	0.10^{ns}	0.21^{ns}	0.10^{ns}	0.27^{ns}	-0.01^{ns}
WSI2	0.15^{ns}	-0.09^{ns}	0.04^{ns}	0.31^{ns}	0.11^{ns}	-0.21^{ns}
WSI3	-0.04^{ns}	0.19^{ns}	0.41^{ns}	0.64^{*}	0.58^{*}	-0.47^{ns}

*P<0.05, ** P<0.01, ns Not significant

Table 6.4 Factorial regression model for partitioning of G x E interaction for cane yield (TCANE). The percentage of sums of squares (%SS) of the G x E interaction is indicated

Source	df	SS	%SS (G x E)	MS	P-value
Genotype (G)	6	11674.34		1945.72	<.001
Environment (E)	11	154293.28		14026.66	<.001
G X E	66	10466.29		158.58	<.001
G X RAD2	6	4376.07	41.8	729.35	<.001
G X TT1	6	468.58	4.5	78.1	0.456
G X RAIN3	6	368.11	3.5	61.35	0.61
Residual	415	33941.63		82	

Table 6.5 Estimates of cultivar sensitivities to radiation during stalk elongation (RAD2) according to the factorial regression model

Cultivar	Sensitivity[α]
N17	1.73
N19	1.09
N27	-0.97
N29	-0.85
N35	-1.13
N36	0.71
NCO376	-0.59

[α]Estimates the expected change in TCANE of cultivars exposed to a one-unit change in RAD2

6.4.3 Relationships among yield components

The covariate effect biplot based on a trait x environment table of correlation coefficients explained 90% of the total variation in the table, suggesting that it was a good approximation thereof (Figure 6.4). In this biplot, the length of a trait vector represents the magnitude of its effects on TCANE. Interestingly, POP and STKLTH had slightly greater effects on TCANE than STKMS. The near opposite directions of the POP and STKMS vectors indicate that these two cultivar traits had contrasting effects on TCANE. Similarly, STKLTH showed no strong correlation to either STKMS or POP (almost right angles), suggesting that its effects on TCANE was not similar to the other traits. The early and late harvests demonstrated a moderate degree of separation on the biplot, which were more distinct along IPCA2. The early harvests E1, E3, E4, E5 and E6 were positively correlated to POP, suggesting that TCANE is influenced to a greater extent by POP in the early season than the late season. In contrast, the late harvests L2, L3, L4, L5, and L6 were positively correlated with STKMS, suggesting that TCANE is affected more by STKMS in the late season compared with the early season. The STKLTH did not appear to have distinctive correlations to either harvest time. The covariate effect biplots for ERC and TERC (data not shown) indicated similar trait relations as those observed with TCANE, however, early and late harvests showed considerable overlap suggesting no distinct clustering for these traits.
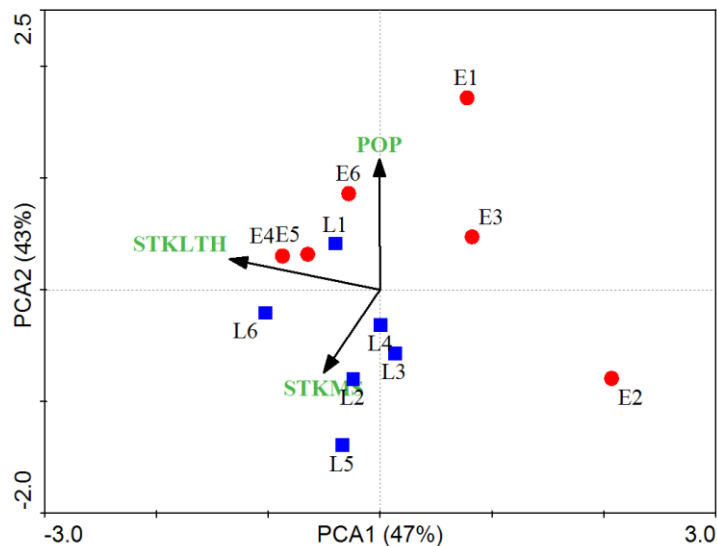
Figure 6.4 Trait x environment biplot based on a trait x environment two-way table of correlation coefficients between traits and TCANE in each environment. Stalk population (POP), stalk mass (STKMS) and stalk length (STKLTH) are represented by arrows. Early season harvests are represented by red circles and late season harvests are represented by blue squares. (The percentage of the interaction explained by the IPCA axes are indicated in parentheses in the axis titles)

6.5 Discussion and conclusions

The separation of cultivars and environments in the AMMI2 biplots of all three variables demonstrated distinct cultivar adaptation to TOH in the sugar industry. This has direct implications on the design of coastal selection programs, where currently, the majority of advanced selection trials are harvested late, and evaluation of adaptability to early season harvesting primarily occurs post-release. These results suggest that selection for early season harvesting along the coast may result in more significant gains. However, the practical implications of splitting advanced selection trials equally into early and late harvests must be considered.

In this study, the higher TCANE of the early harvests was attributed to greater exploitation of higher temperatures and radiation during stalk elongation and ripening, while the phases of stalk elongation and ripening in late harvests corresponded to periods of low temperatures and radiation. This synchronization of phenology and environmental potential is proposed as the main reason for differences in TCANE between early and late harvests. Donaldson et al. (2008) indicated that lower biomass production associated with late harvests were caused by

120

premature ripening of stalks in response to winter temperatures experienced the following year. They proposed that the accumulation of sucrose in immature stalks triggers a feedback mechanism which inhibits further structural growth, resulting in lower biomass production. Indeed, a feedback mechanism may very well be responsible at the physiological level, and the outcomes of this study may be explained using this concept. Furthermore, the differential responses of cultivars to early and late harvests demonstrated differences in their ability to exploit these growing conditions at specific growth phases. Singels et al. (2005b) reached similar conclusions about differential cultivar responses to TOH and explained these differences within the context of biomass partitioning and radiation use.

Many of the responses observed in this study may be explained using concepts identified in other physiological studies (Singels et al., 2005a; Singels et al., 2005b; Donaldson et al., 2008), and they correspond to what is generally known about sugarcane growth and development in the industry. This suggests that the multivariate techniques employed here (although only allowing for basic physiological interpretations from the available data) do have potential in future studies involving interpretations of G x E interactions. The approach may seem rather laborious for a study involving only 12 environments; however, when applied to more complex datasets comprising detailed crop and environmental parameters, this approach should yield more valuable interpretations that can inform selection and/or further physiological studies. Additionally, cultivars chosen for detailed physiological studies are conventionally selected based on expert opinion, subjective information and anecdotal evidence. This type of study should be viewed as an objective way to characterize cultivars to inform further physiological/crop modelling studies, and is conventionally regarded as 'hypothesis generating', rather than 'hypothesis testing' research.

The analysis of environment x trait relations in the covariate-effect biplot revealed positive correlations between stalk population and TCANE in four of the six early season harvests (Figure 6.4). This implies that early harvests were more dependent on the yield-population relationship, and selection for higher yield in the early season may be accelerated through selection of high stalk population cultivars. This response may be linked to the effects of temperature on the duration of tillering in grasses. Lower temperatures increase the duration of tillering in other crops such as wheat, thereby resulting in greater tiller numbers at harvest (Ramburan and Greenfield, 2007). Hence, tillering and establishment of early harvests through winter may promote higher stalk populations in a similar manner. This may in turn be followed by favourable temperature and moisture conditions during spring, thereby

limiting tiller mortality until harvest. Conversely, cane yields of late season harvests were more dependent on individual stalk mass, which implies that gains from selection for cane yield in the late season could be accelerated by selecting for higher stalk mass. The biplot also demonstrated the well-known negative correlation between POP and STKMS in sugarcane. These results indicate possible benefits of employing differential trait selection strategies for different harvest times in the South African industry. However, a more comprehensive study involving a greater number of relevant cultivar traits is needed.

Rainfall and the water stress index were not correlated to IPCA1 or IPCA2 to the extent that temperature and radiation was, suggesting that significant variability between early and late harvests were not due to these former covariates. This may be due to the fact that the trials were located adjacent to each other and conducted simultaneously, thereby creating confounding effects of environmental factors. For example, RAIN1 (average daily rainfall during establishment) for early trials would have been similar to RAIN3 (average daily rainfall during ripening) for late season trials. This overlap of seasons and environmental factors between the two harvests may have effectively nullified any real differences in rainfall. This highlights the need for diverse environments when studying G x E interactions. Nevertheless, the significant correlation between the IPCA2 score and the WSI3 suggests that the WSI covariate may have potential in further studies involving cultivar sensitivities to drought stress.

In this study, the duration of growth phases was estimated from stalk population and stalk height measurements done in previous studies, and were not based on actual measurements of phenology and development within the trials. It is acknowledged that this crude estimation is not ideal, especially when considering the effects of different seasons and responses of different cultivars in development. Most G x E interpretive studies of this nature have been conducted on determinate grain crops such as wheat and maize, where growth phase changes can be identified from visual observation (e.g. flag leaf appearance in wheat, or silking in maize). In sugarcane, such distinct growth phase switches are difficult to define, and may require detailed monitoring of different crop parameters for more accurate estimations. However, even when a crude estimation was used in this study, the biological interpretations from the analyses corresponded to known knowledge of sugarcane growth. Nevertheless, future studies of this nature should include better estimates of growth phases, either through detailed measurements or through the applications of crop models.

Much difficulty was experienced in fitting the factorial regression model. Although the number of environmental covariates for further analysis was culled through the use of PCA and correlations with IPCA scores from AMMI, the high degree of co-linearity between covariates prevented the fitting of a more accurate model. The sequence in which terms were added to the model made a large difference to the proportion of G x E accounted for, suggesting that the results from the factorial regression should be interpreted with caution. Future attempts to fit factorial regression models should involve more contrasting environmental conditions.

It is acknowledged that the dataset used in this study was not ideal for interpretation of G x E, as the experiments did not constitute a MET (Gauch, 1992) *per se*. In fact, the experiments analysed in this study were designed to provide practical industry recommendations, yet the application of multivariate analysis techniques demonstrated how such experiments can be utilized to investigate and interpret growth and development in sugarcane. Despite the fact that the trials were adjacent to each other and experienced overlapping climatic influences, the relative effects of temperature and radiation at different growth phases were clearly evident. This implies that application of this methodology to METs using sites that vary substantially in their characteristics in the SA sugar industry may reveal trends and mechanisms defining the adaptability of cultivars to different conditions. The following chapter will build on the methodologies outlined here to provide comprehensive interpretations of G x E interactions in the diverse coastal region of the industry.

6.6 References

Anonymous, 2009. GenStat® Executable Release 12.1. Lawes Agricultural Trust, Rothamstead Experimetnal Station, Harpenden, U.K. Clarendon Press, London, U.K.

Di Bella, L.P., Stringer, J.K., Wood, A.W., Royle, A.R., Holzberger, G.P., 2009. What impact does time of harvest have on sugarcane crops in the Herbert River district (in Australia). *Sugar Cane International* 27, 143-148.

Donaldson, R.A., Redshaw, K.A., Rhodes, R., Van Antwerpen, R., 2008. Season effects on productivity of some commercial South African cultivars, I: Biomass and radiation use efficiency. *Proceedings of the South African Sugar Technologists Association* 81, 517-527.

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

Gilbert, R.A., Shine, J.M., Miller, J.D., Rice, R.W., Rainbolt, C.R., 2006. The effect of genotype, environment and time of harvest on sugarcane yields in Florida, USA. *Field Crops Research* 95, 156-170.

Julien, M.H.R., Delaveau, P., 1977. The effects of time of harvest on the partitioning of dry matter in three sugarcane varieties grown in contrasting environments. *Proceedings of the International Society of Sugar Cane Technologists* 16, 1755-1769.

Nayamuth, A.R., Mangar, M., Ramdoyal, K., Badaloo, M.G.H., 2005. Early sucrose accumulation, a promising characteristic to use in sugarcane improvement programs. *Proceedings of the International Society of Sugar Cane Technologists* 25, 421-429.

Parfitt, R.C., 2000. Genotype by environment interaction among secondary variety trials in the Northern region of the South African sugar industry. *Proceedings of the South African Sugar Technologists Association* 74, 245-248.

Ramburan, S., Greenfield, P.L., 2007. The effects of chlormequat chloride and ethephon on agronomic and quality characteristics of South African irrigated wheat. *South African Journal of Plant and Soil* 24, 106-113.

Ramburan, S., Paraskevoupolos, A., Saville, G., Jones, M., 2010. A decision support system for sugarcane variety selection in South Africa based on genotype-by-environment analyses. *Experimental Agriculture* 46, 243-257.

Ramburan, S., Van Den Berg, M., 2011. Review and analysis of post-release variety evaluation of sugarcane: A South African perspective. *International Sugar Journal* 113, 444-452.

Robertson, M.J., Wood, A.W., Muchow, R.C., 1996. Growth of sugarcane under high input conditions in tropical Australia. I. Radiation use, biomass accumulation and partitioning. *Field Crops Research* 48, 11-25.

Singels, A., Donaldson, R.A., 2000. A simple model of unstressed sugarcane canopy development. *Proceedings of the South African Sugar Technologists Association* 74. 151-154.

Singels, A., Donaldson, R.A., Smit, M.A., 2005a. Improving biomass production and partitioning in sugarcane: theory and practice. *Field Crops Research* 92, 291-303.

Singels, A., Smit, M.A., Redshaw, K.A., Donaldson, R.A., 2005b. The effect of crop start date, crop class and cultivar on sugarcane canopy development and radiation interception. *Field Crops Research* 92, 249-260.

Ter Braak, C.J.F., Smilauer, P. 2003. Canoco® for Windows Version 4.51. Biomteris Plant Research International, Wageningen, The Netherlands.

Thomas, D.W., Parfitt, R.C., 2000. The effect of early and late season harvesting on the ranking of 150 sugarcane clones from one cross. *Proceedings of the South African Sugar Technologists Association* 74, 249-251.

Van Eeuwijk, F.A., Elgersma, A., 1993. Incorporating environmental information in an analysis of genotype by environment interaction for seed yield in perennial ryegrass. *Heredity* 70, 447-457.

Van Oosterom, E.J., Mahalakshmi, V., Bidinger, F.R., RAO, K.P., 1996. Effect of water availability and temperature on the genotype-by-environment interaction of pearl millet in semi-arid tropical environments. *Euphytica* 89, 175-183.

Vargas, M., Crossa, J., Van Eeuwijk, F.A., Ramirez, E., Sayre, K., 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype x environment interaction. *Crop Science* 39, 955-967.

Voltas, J., Lopez-Corcoles, H., Borras, G., 2005. Use of biplot analysis and factorial regression for the investigation of superior genotypes in multi-environment trials. *European Journal of Agronomy* 22, 309-324.

Voltas, J., Van Eeuwijk, F.A., Sombrero, A., Lafarga, A., Igartua, E., Romagosa, I., 1999. Integrating statistical and ecophysiological analyses of genotype by environment interaction for grain fillin of barley I. Individual grain weight. *Field Crops Research* 62, 63-74.

Yan, W., Kang, M.S., 2003. GGE Biplot Analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL.

Yan, W., Tinker, N.A., 2005. An integrated biplot analysis system for displaying, interpreting, and exploring genotype x environment interaction. *Crop Science* 45, 1004-1016.

CHAPTER 7

INTEGRATING EMPIRICAL AND ANALYTICAL APPROACHES TO INVESTIGATE SUGARCANE G X E INTERACTIONS IN THE COASTAL AND HINTERLAND REGIONS OF SOUTH AFRICA[7]

7.1 Abstract

The interpretation of sugarcane G x E interactions through simultaneous consideration of environmental covariates and genotypic traits across many trial series has not been attempted. The objectives of this study were (1) to investigate the magnitude of G x E interactions and test site similarity in the long and short cycle selection programs along the coast, (2) to identify factors responsible for G x E interactions by integrating genotypic trait and climatic data into routine G x E analyses, and (3) to evaluate and illustrate the integrated use of crop models, climatic and soil data for the growth-phased interpretation of sugarcane G x E interactions. The ERC yield data from eight series of METs were analyzed using variance components, GGE biplot analysis and AMMI. Environmental covariates were computed for each environment (trial x ratoon combination). The time to canopy formation was estimated from the Canesim crop model and environmental covariates were summarized before and after canopy formation. Environmental covariates and genotypic traits were correlated to AMMI IPCA scores and superimposed on the biplots.

The G x E interaction accounted for more variation than the main effect of genotype across all series for the long cycle, while the opposite was true for the short cycle. The repeatable component of G x E interaction (genotype x site) was more dominant in the long cycle than in the short cycle, where the non-repeatable component (genotype x ratoon and genotype x ratoon x site) dominated. The GGE biplots identified sites K1 and G2 as redundant within the long cycle, while site T2 was recommended for removal from the short cycle program. The G sites were characterized by lower yield potential conditions compared with the K sites, which generally experienced higher rainfall and lower water stress. The joint examination of AMMI biplots with covariates and genotypic traits superimposed allowed for identification of environmental factors that affected the different genotypic traits. The biplots permitted a range of biological interpretations and trait relations to be studied. The splitting of

---

environmental covariates into pre-and post-canopy values did not add much value to the interpretations. The approaches and methods illustrated in this study may be utilized in other sugarcane G x E studies to improve the utility and value of MET data.

7.2 Introduction

The G x E interactions in plant improvement programs have historically been analyzed using empirical approaches that have a direct impact on selection decisions. Such empirical analyses are also aimed at quantifying the size and repeatability of the interactions, and identifying the similarity of test sites for selection. Although breeding advance has been achieved through this approach, it is acknowledged that applying integrated approaches to these analyses through the use of associated soil and climatic data, together with derived information from crop models may allow for more comprehensive interpretations of MET data. Integrated studies of this nature are very rare, because of the lack of soil and climatic data associated with METs. Where associated soil and climatic data are available, a range of statistical techniques have been applied to interpret G x E interactions.

The most common method of interpreting G x E interactions involves the correlation of climatic data to PC scores derived from PCA-based methods such as AMMI. This approach has been successfully used to understand causes of G x E interactions in crops such as barley (Voltas et al., 1999), pearl millet (Van Oosterom et al., 1996), ryegrass (Van Eeuwijk and Elgersma, 1993), wheat (De Vita et al., 2010) and others. Yan and Hunt (2001) correlated PC scores derived from GGE biplot analysis to agronomic traits, disease incidence ratings, and climatic covariates in a study of winter wheat METs in Ontario. Most studies of this nature have focused on correlating IPCA scores to environmental covariates only (as demonstrated in Chapters 4 and 6). Comparatively fewer studies have integrated interpretations of environmental and genotypic covariates. Such an integrated analysis may allow for the understanding of the genotypic characteristics that contribute to a superior genotype, and the environmental factors that can be manipulated to facilitate selection for such genotypes (Yan and Hunt, 2001). Furthermore, a joint visual examination of biplots containing both environmental covariates and genotypic traits may facilitate easier interpretations compared with a straightforward correlation analysis.

Chapter 5 illustrated an approach for the combined use of empirical and analytical methods to better understand G x E interactions and trait relations. The grouping approach used showed

that G x E patterns could be explained by environmental differences between sites and provided insights into how different groups of genotypes responded to different types of environments. Although very useful, the approach requires extensive statistical analyses and data processing. The correlation of AMMI IPCA scores to covariates and the subsequent enrichment of the biplots with the covariates (as described above and illustrated in Chapters 4 and 6) is much simpler and can provide comparable interpretations. However, this approach has not been tested across many trial series to evaluate repeatability of the interpretations (Chapters 4 and 6, and other studies applied this method to single datasets only).

Also, in Chapter 6, environmental covariates were summarized within growth phases that were crudely estimated. It is envisaged that a more scientific approach to estimating changes in crop development, through the use of crop models, may be more accurate. This approach, combined with an interpretation of genotypic trait and environmental covariate effects on G x E interactions may allow for more comprehensive insights into factors causing differential genotype performance in different environments. If the approach is successful (i.e. shows consistent interpretations across trial series), it is envisaged that the technique may be applied with other covariates that include pest and disease ratings or factors relating to crop nutrition. The outcome would be an integrated analysis technique that allows for the testing of trial site similarity within a MET network, the identification of environmental factors characterizing trial sites, investigating genotypic trait relations, characterizing genotypes according to trait parameters and responses to environmental stresses (for recommendation purposes), and the identification of major causes of G x E interactions to inform future trial site selection. This chapter therefore focuses on the above issues within the diverse coastal and hinterland regions of the industry.

In addition to understanding the causes of G x E interactions from an environmental and genotypic perspective, an integrated analysis that incorporates all growth influencing factors may also provide knowledge of crop growth that goes beyond the realms of traditional G x E research. For example, integrated G x E analysis techniques have been used to determine and confirm the sensitivities of crops to environmental factors at different growth phases (Voltas et al., 1999). For sugarcane, knowledge of crop responses to environmental factors at different growth phases is traditionally gathered from detailed field trials that are not routinely repeated in space and time. Consequently, these trials do not accommodate G x E influences on different crop parameters. An integrated analysis method that accounts for G x E interactions and provides information that contributes to understanding of crop growth and

development is therefore needed. Such analyses may be regarded as preliminary investigations that will inform further detailed physiological/crop modelling research. An initial step in developing such methodology would be to evaluate the extent of interpretations that are possible from a conventional MET dataset comprising diverse genotypes and environments. Advanced plant breeding selection trials in the coastal and hinterland regions of the industry are ideally suited for such an evaluation.

The coastal and hinterland regions represent the most heterogeneous population of environments in the industry, characterized by large temporal and spatial fluctuations in production conditions. Commercial genotypes are selected for this region from two separate selection programs that are defined by the average age at which sugarcane is harvested in the region. The designated short and long cycle programs are designed to identify promising genotypes suited to harvesting at 12 to 15 and 15 to 18 months of age, respectively. The Gingindlovu (G) and Kearsney (K) selection schemes run concurrently to identify genotypes for production on a long cutting cycle at low and higher altitude (hinterland), respectively. This scheme is characterized by permanent selection sites located either on SASRI owned research farms (G1 and K1), or on grower co-operator farms in the region (K2, G2, and G3). The so-called "off-station" sites K2, G2, and G3 represent production conditions that are slightly different from the research farms. The Gingindlovu research farm also supports selection for the short cycle program. In this instance, the Gingindlovu (U) and Empangeni (T) selection schemes run concurrently to identify genotypes for production on a short cutting cycle, under low and high yield potential conditions, respectively. Again, this scheme is characterized by permanent selection sites on SASRI owned research farms (U1 and T1), or on grower co-operator farms in the region (U2 and T2), under different conditions. Site T3 represents a late season harvest on the SASRI owned T1 site. Since inception of the long and short cycle programs, no studies have actively investigated the nature of G x E interactions, the contribution of different components of variation to the G x E interaction, and the similarities between the test sites used for selection. Such information is essential to optimize the design of the selection network. As shown in Chapter 5, variance components analysis and GGE biplot analysis are techniques that are suited to investigating the above issues.

Given the above considerations, the objectives of this study were (1) to investigate the magnitude of G x E interactions and test site similarity in the selection programs using variance components and GGE biplot analysis, (2) to identify factors responsible for G x E interactions by integrating trait and climatic data into routine AMMI analyses, and (3) to

evaluate and illustrate the integrated use of crop models, climatic and soil data as tools for the growth-phased interpretation of sugarcane G x E interactions.

7.3 Materials and methods

7.3.1 The MET dataset

The dataset used for this study comprised 77 advanced plant breeding selection trials established between 2000 and 2008 at 10 selection sites in the coastal and hinterland regions of the industry. Within each program, a set of 20 to 27 promising sugarcane genotypes that had progressed through the initial selection stages were annually included in advanced METs, together with two to eight commercial checks. The genotype sets differed completely between the long and short cycle programs. Consequently, the dataset was split into the long (41 trials) and short (36 trials) cycle programs and analyzed separately. Trials that were established in each planting year (series) were harvested as plant, first and second ratoon crops. However, some trials were not harvested across all ratoons. In each program, the genotype x trial datasets were balanced within series but unbalanced across series (except for commercial checks). All trials were planted in randomized complete blocks with three to four replicates. Experimental units comprised five sugarcane rows that were 10 m long, and spaced 1.2 m apart. At each harvest event, three net rows were hand-harvested and weighed using a digital scale mounted on a tractor-operated hydraulic boom, to determine TCANE. Twelve-stalk samples were taken from each plot to determine individual stalk mass (STKMS) and a range of milling characteristics, including estimated recoverable crystal (ERC), and fibre (FIB) percentages. The TERC was calculated as a product of TCANE and ERC. Other important traits measured at harvest included stalk population (POP) in stalks/ha and stalk diameter (DIAM) in mm. Visual ratings (1 to 9) of lodging (LOD), canopy formation (CAN) and stalk bending (SKEW) were also done in each trial at appropriate growth stages.

7.3.2 Seasonal covariates

For each trial in the dataset, covariates that varied seasonally with each ratoon (hence termed *seasonal covariates*) were determined. The descriptions of the covariates and how they were determined are given in Chapter 4. Briefly, these included average daily rainfall in mm (RAIN), average daily thermal time (TT) in heat units (base 10°C), average daily solar radiation (RAD) in $MJ/m^2/sec/day$, average daily A-pan evaporation (EVP) in mm, and a water stress index (WSI). The WSI in this study was calculated as a ratio of simulated rainfed evapotranspiration to simulated irrigated evapotranspiration from the Canesim crop

model as described in Chapter 4. The age at harvest (AGE) and time of harvest (TOH) were also recorded as covariates. The Canesim model, which was run for each environment using the relevant soil and weather parameters, simulated canopy formation and stalk elongation. This allowed for an estimation of the approximate number of days to full canopy formation for each environment. Using this information, the average daily values of seasonal covariates were then calculated "before" and "after" canopy formation. Figure 7.1 is an example of an output from a model simulation, depicting how the approximate time of canopy was determined for environments tested in this study. Each of the five seasonal covariates described above were then calculated for the pre-canopy (1) as well as the post-canopy (2) phases. For example, RAIN1 and RAIN2 referred to average daily rainfall before and after canopy formation, respectively. The other covariates were coded similarly (except for AGE and TOH).
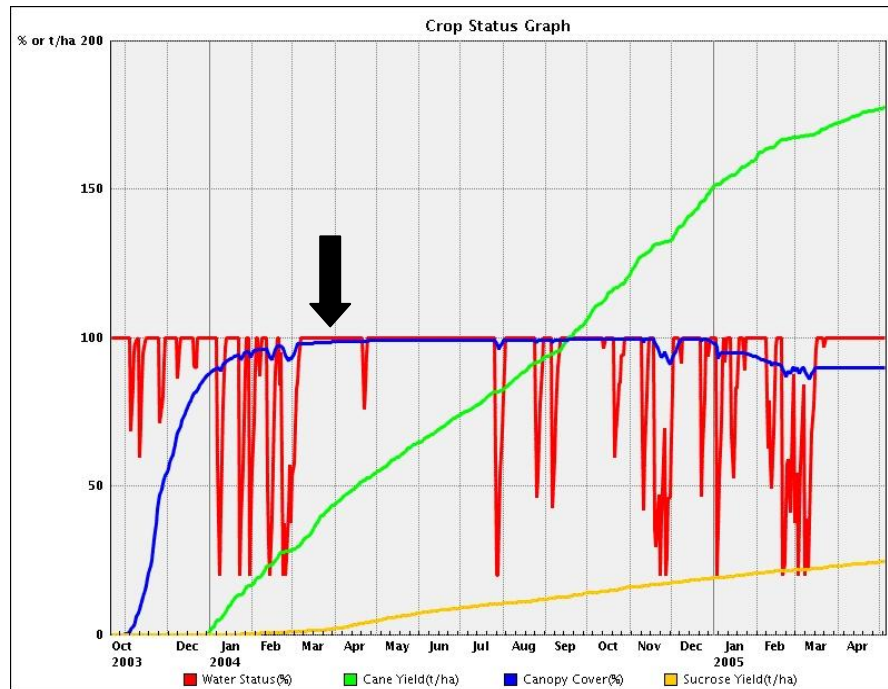
Figure 7.1 An example of an output of a Canesim simulation of a sugarcane crop. The black arrow shows the point of 100% canopy cover, which was used to split the crop cycle and calculate covariate values before and after canopy formation

7.3.3 Variance components analysis

The following random model was fitted to the TERC data for each trial series (2000-2009)

$$Y_{ijkl} = \mu + T_j + B(T)_{l(j)} + G_i + GT_{ij} + GB(T)_{il(j)} + R_k + RT_{jk} + RB(T)_{kl(j)} + RG_{ik} + RGT_{ijk} + e$$

131

where Y_{ijkl} is the observation l of genotype i in ratoon k of trial j, μ is the grand mean, T_j is the trial main effect, $B(T)_{l(j)}$ is the effect of block within trial, G_i is the genotype main effect, GT_{ij} is the effect of the genotype x trial interaction, $GB(T)_{il(j)}$ is the effect of genotype x block (within trial), R_k is the main effect of ratoon, RT_{jk} is the effect of the ratoon x trial interaction, $RB(T)_{kl(j)}$ is the effect of the ratoon x block (within trial) interaction, RG_{ik} is the effect of the ratoon x genotype interaction, RGT_{ijk} is the effect of the genotype x trial x ratoon interaction, and e is the error. Restricted maximum likelihood (Patterson and Thompson 1971) using the sparse average information algorithm (Gilmour et al. 1995) was used, as implemented in GenStat 12.0 (Anonymous, 2009).

7.3.4 GGE biplot analysis

Within each series, the two-way matrix of G x E TERC means was analyzed using GGE biplot analysis (Yan et al., 2000). A detailed description of this analysis and software is given in Chapter 3. In this study, the polygon viewing option was implemented as it is the most appropriate to study site similarities and MEs. A description of this viewing option was given in previous chapters.

7.3.5 AMMI analysis and G x E interpretation

In order to identify the causes of G x E interactions, AMMI analysis was conducted on a balanced subset of genotypes within each series (genotypes that were not included in all trials were removed). Details of the AMMI model and associated biplots were given in previous chapters. The environment IPCA scores that were derived from AMMI analysis were regressed against the seasonal covariates. The derived regression coefficients from the above regression then defined the positions of the covariate on the AMMI biplot (Van Eeuwijk and Elgersma, 1993). For example, the regression coefficient derived from the regression of the covariate and IPCA1 scores was used as the x-coordinate in the biplot, while the regression coefficient derived from the covariate and IPCA2 regression was used as the y-coordinate in the biplot. In this way, the AMMI biplots were enriched with the seasonal covariates. When the covariates were superimposed onto the biplot, as described above, the adaptability of genotypes to the different environments are interpreted as they are in a conventional biplot. However, the difference now is that the environmental characteristics (covariates) of the different environments could be simultaneously viewed. This allowed for an interpretation of the environmental factors characterizing the different test sites, and helped identify genotypes that showed responses to those factors.

Similarly, the genotype IPCA scores derived from AMMI analysis were regressed against genotype traits, and the derived regression coefficients then defined the positions of traits on the biplots. The traits were subsequently superimposed onto the biplot, as described above. The resultant biplots (one containing covariates and the other containing genotype traits) were then examined together to allow for the simultaneous interpretation of genotype adaptabilities to environments, the environmental covariates driving those trends, and the genotype traits that were primarily responsible for differential responses. This analysis was conducted for each series with the objective of identifying trends that were repeatable, thereby adding confidence to interpretations and recommendations.

7.4 Results

7.4.1 Variance components

The variances associated with the most relevant terms of the random model are expressed in Table 7.1 as a percentage of total variation within each series. For the long cycle, site differences accounted for the largest proportion of total variation in four (2000, 2001, 2003, and 2005) of the eight series. In the other four series the ratoon (series 2006), and site x ratoon interaction (2002, 2004, and 2007) accounted for the largest proportion of variation. The main effect of ratoon (confounded with season) varied dramatically, accounting for 0% of the variance in series 2001, 2002, and 2003 (this suggests that the variation accounted for by the main effect of ratoon was less than the variation accounted for by the individual ratoon interaction components). This variance accounted for by ratoon ranged up to 44.6% of total variance in series 2006. The low variance associated with the site x ratoon interaction in series 2000 is attributable to the fact that only two crops were harvested from each trial within that series. The G x E interaction (expressed as a sum of the genotype x site, genotype x ratoon, and genotype x site x ratoon interactions) accounted for considerably more variation than the main effect of genotype across all series. The genotype x site term accounted for more variation than the combined effects of genotype x ratoon and genotype x site x ratoon in five out of eight series. This implies that the repeatable component of G x E (i.e. genotype x site) is exploitable and homogenous regions can be identified. The site x ratoon interaction accounted for a higher proportion of variation than the genotype x ratoon interaction across all series (except for series 2000). This suggests that ratooning ability is influenced to a larger extent by site differences than by genotype differences.

The percentage of total variation accounted for by site differences was considerably lower for the short cycle (Table 7.1). Only in the 2004 series did the site component account for the largest proportion of variation. In general, the ratoon and/or the site x ratoon interaction accounted for most of the variation across the series, except for series 2004 and 2006 where large proportions of variation were attributed to error. The genotype term accounted for more variation than the G x E interaction (expressed as a sum of the genotype component terms) in series 2001, 2003, and 2004. This is in contrast to the long cycle where the G x E interaction always accounted for more variation than genotype. Also, in contrast to the long cycle, the genotype x site term accounted for less variation than the combined effects of genotype x ratoon and genotype x site x ratoon in seven out of eight series. This implies that for the short cycle, the non-repeatable component of G x E (i.e. genotype x ratoon and genotype x site x ratoon) is more dominant than the repeatable component. Therefore, the identification of homogenous regions for selection and recommendations may be more difficult in the short cycle.

Table 7.1 The percantage of total phenotypic variance associated with relevant random terms from variance components analysis, over eight trial series on the long and short cycles

Random term	Series (Long Cycle)								
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Site	64.1	45.4	16.4	47.8	6.3	32.8	13.9	20.2	-
Genotype	0.5	0.9	0.2	2.3	4.3	4.4	2.9	1.9	-
Genotype x Site	10.7	10.2	14.8	3.0	2.9	2.7	2.2	8.2	-
Ratoon	10.4	0	0	0	7.6	21.7	44.6	21.1	-
Site x Ratoon	0.2	27.1	37.6	28.6	35	12.9	16.1	24.9	-
Genotype x Ratoon	1.3	0.1	1.6	0.8	0.5	1.7	1.5	1.1	-
Genotype x Site x Ratoon	0.8	1.3	3.8	2.1	4.3	3.3	1.5	1.3	-
Residual	10.4	8.4	23.8	15.7	21.4	11	12.5	12.5	-
	Series (Short Cycle)								
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Site	-	13.6	6.3	0.3	13.6	57.7	16.2	7.8	17.8
Genotype	-	3.1	7.1	8.1	12.2	3.7	8.3	7.6	3.2
Genotype x Site	-	0.4	2.9	1.9	3.4	2.7	9.3	5.2	1.1
Ratoon	-	37.5	34.1	0	7.9	1.3	6.2	0	27.9
Site x Ratoon	-	19.5	10.9	62.1	13	8.6	7.8	39.1	27.5
Genotype x Ratoon	-	1.8	1.5	1.6	4.8	3.5	2.1	1.3	0.7
Genotype x Site x Ratoon	-	0.8	6.5	1.6	2.9	1.9	3.3	5	3.5
Residual	-	8	18.7	20.1	31.1	9.9	22.8	16.7	11.1

\-   No data available for analysis in that series

7.4.2 GGE Biplot analysis

7.4.2.1 Long cycle

Depending on the series, the biplots for the long cycle accounted for 47.8 to 67% of genotype plus G x E interaction, and therefore represented reasonable approximations of the genotype and G x E effects (Figure 7.2). In most long cycle biplots, ratoons of the same trial clustered together tightly before clustering with other trials. The G and K trial sites frequently formed separate mega-environments (series 2000, 2001, 2002, and 2003) or formed separate clusters within a ME that was shared (series 2005). Ratoons from sites K1 and K2 occasionally fell within a unique ME sector (series 2001, 2002, and 2003) and they also tended to cluster together when sharing a ME with the G sites (series 2005). Of the two K sites, K2 was more discriminative, as it was located in unique ME sectors in five out of eight biplots (series 2000, 2001, 2002, 2003, and 2004) and was located further away from the G sites in the other biplots. Site K1, on the other hand, either grouped together with K2 in a ME sector, or clustered with the G sites. This suggests that K1 may in fact be a redundant site as it did not uniquely discriminate genotype responses in any series.

As a group, the three G sites generally clustered together in the biplots, forming groups that were clearly different from the K sites. However, none of the G sites consistently formed a distinct ME across all series. The ratoons from site G1 either fell into the same ME sector as the corresponding ratoons from site G2 (series 2001, 2003, 2004, 2005, and 2006), or they were located in very close proximity to each other at the diagonal defining the MEs (series 2000, 2002, and 2007). This suggests that G1 and G2 may be similar in the way they discriminate genotype performance. Ratoons from site G3 occasionally showed a tendency to detach from the larger G cluster (series 2001, 2004, 2005, 2006, and 2007), suggesting that it may discriminate genotype performance differently to the other G sites. However, in general, the G sites did not cluster according to any observable pattern. In series 2006, the clustering of the G sites seemed to be dependent on the ratoon numbers rather than the actual sites, suggesting that ratoon (season) may have had a dominating effect in that series. This is reflected in the variance components analysis, where ratoon accounted for a large proportion (44.6%) of the total variation.

7.4.2.2 Short cycle

For the short cycle, the ratoons of the same trial generally clustered together on the biplots (Figure 7.3). Therefore, for easier interpretation, the genotype values were averaged across

ratoons of the same trial. Consequently, the biplots for the short cycle accounted for 62.1 to 83.4% of the genotype plus G x E interaction. In general, the separation of the high potential (T) and low potential (U) sites into different MEs was only observed in the 2005 and 2006 series. In the other series, all sites either fell into the same ME (series 2002 and 2004), or a single site formed a separate ME (series 2001, 2003, 2007 and 2008). In the latter group of series, site U2 formed a unique ME in two of the series (2003 and 2007), while site T1 and U1 formed a ME in series 2001 and 2008, respectively. Sites U1 and U2 distinctly fell into the same ME in four out of the eight series. Of these two sites, U2 appeared to be more discriminative, as it was frequently located further away from the T sites compared with U1. Conversely, site U1 did not appear to be very discriminative, as it formed a unique ME in one series only (2008), and it clustered frequently with either T1 or T2 in the other series. Despite this, however, there was not enough evidence to suggest that U1 was a redundant site. Sites T1 and T2 fell into the same ME sector in six out of eight series, suggesting that one of these sites may be redundant. Of these two sites, T1 appeared to be more discriminative (was located further away from the U sites) in most series. This suggests that T1 should remain in the trial network while T2 could be dropped, as it provided no new information on genotype performance. Additionally, site T3 was only evaluated across four series but always fell into the same ME as site T1. Once again, this suggests that T3 provides no new information on genotype performance and could therefore be considered for removal from the trial network.

Figure 7.2 GGE biplots for eight series (2000 to 2007) of multi-environment trials conducted on the long cutting cycle program along the coast. The Gingindlovu (G) and Kearsney (K) environments are shown in red, and are represented by the site and ratoon numbers, respectively. Genotypes are represented by "+"
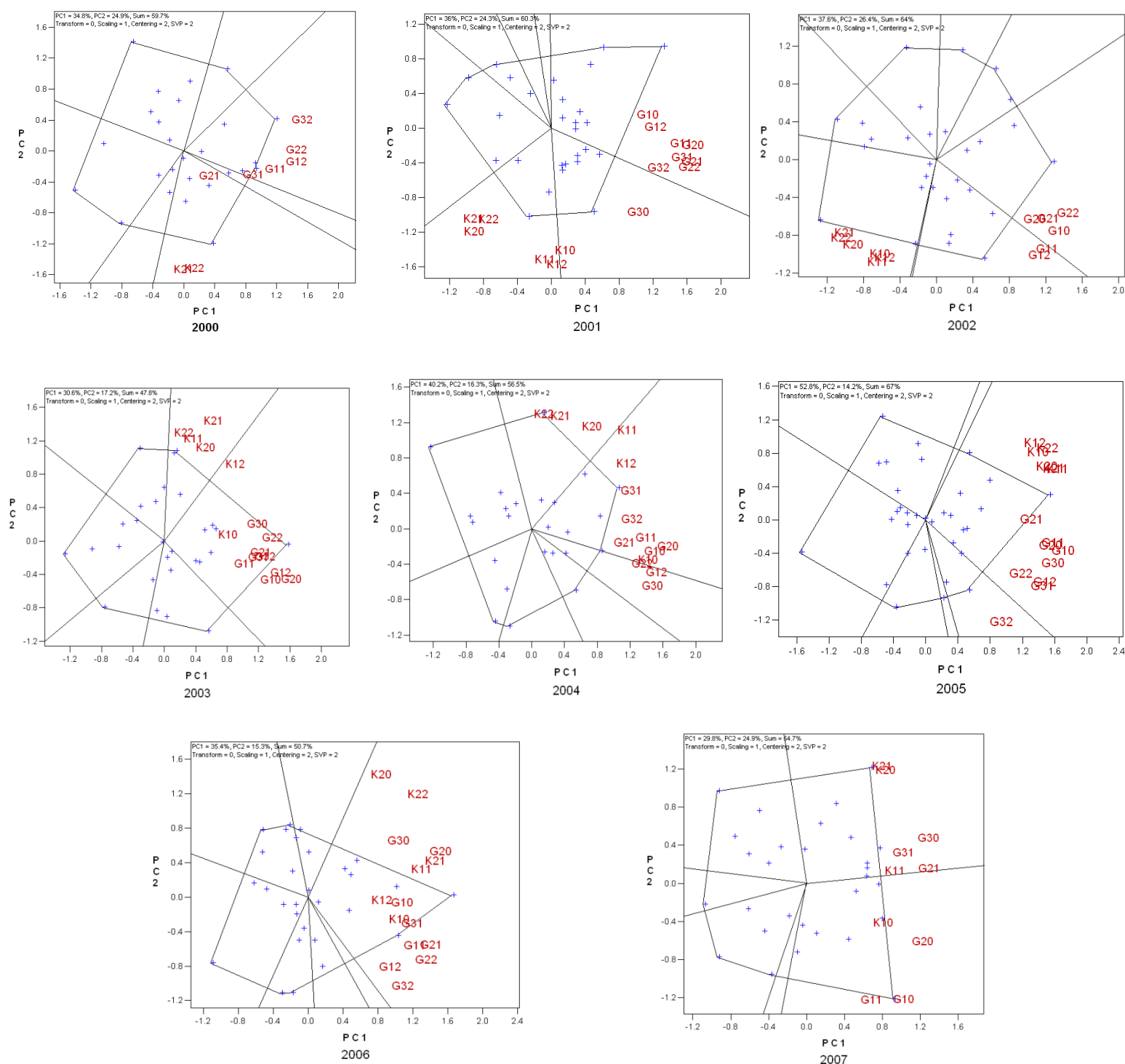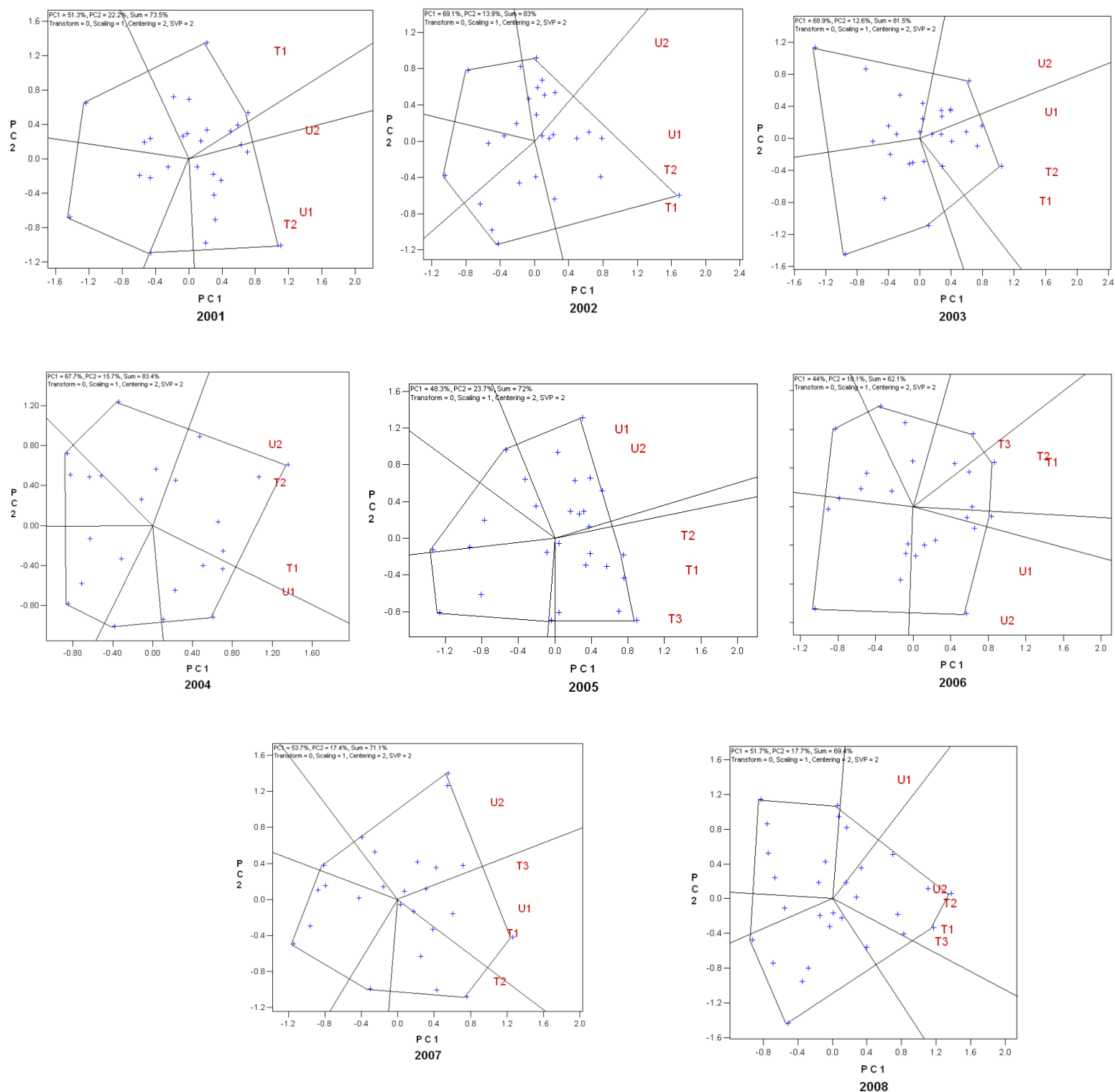
Figure 7.3 GGE biplots for eight series (2001 to 2008) of multi-environment trials conducted on the short cutting cycle program along the coast. The Gingindlovu (U1 and U2) and Empangeni (T1, T2 and T3) sites (averaged across ratoons) are shown in red. Genotypes are represented by "+"

7.4.3 Interpretation of G x E interactions

7.4.3.1 Long cycle

The AMMI biplots with environmental covariates and genotypic traits superimposed are shown for all series from the long cycle in Figures 7.4, 7.5 and 7.6. The AMMI biplot for each series was produced twice; once with the environmental covariates superimposed, and once with the genotypic traits superimposed. The resulting two biplots for each series are displayed adjacent to each other and may be interpreted together. In these biplots, the length and direction of the covariates or traits are indications of their influence on the separation of environments or genotypes on the biplots. For example, in series 2000 (Figure 7.4) the separation of the K environments from the G environments may be attributed to the covariates RAIN1, RAIN2, EVP1, WSI2, and TT2 because these covariates were spread explicitly along the IPCA1 axis together with the environments. Similarly, the separation of sites G21 and G32 along the IPCA2 axis may be attributed to differences in TOH, EVP2, RAD1 etc. A similar interpretation will apply for the genotypic traits. Here, the length and direction of the trait vectors are indications of their influence on the separation of the genotypes on the biplots. For example, genotypes that were located on the right hand side of the biplot were characterized by high POP and TCANE, while those on the left hand side were characterized by high SKEW. These traits were therefore the most influential in causing separation of the genotypes along the IPCA1 axis in the 2000 series.

In all series the K sites showed considerable separation from the G sites on the AMMI biplots. However, environment K10 separated from the cluster of K environments in the 2003, 2004 (Figure 7.5), and 2007 (Figure 7.6) series. In general, the K and G environments separated along the IPCA1 axis in most series. Thereafter, further separation within the K and G sites occurred along the IPCA2 axis. In series 2000, 2001, 2002 (Figure 7.4), and 2007 (Figure 7.6), the K environments were explicitly associated with higher WSI2, RAIN2 and EVP2. In the remaining series, the K environments (especially those from site K2) were always associated with WSI2 (low levels of water stress after canopy formation). The G environments, however, were always associated with higher temperatures, radiation, and age irrespective of the growth stage. Across all biplots, the age at harvest was always associated with the G environments, while the time of harvest did not show any consistent associations.

Although sites K1 and K2 often separated along the IPCA2 axis, this separation could not be attributed to any specific group of covariates, as the covariates were not concomitantly spread

along this axis with the two K sites. Regarding the covariates causing the separation of the G environments on the biplots, it was evident in the 2000 series that the 2nd ratoons at all three sites were associated with RAD1, RAD2, AGE, WSI1, and TT1, while the 1st ratoons were associated with TOH and EVP2. However, this distinct pattern was not repeated in the other series. The covariates causing the separation of the G environments along IPCA2 were not constant. However, RAIN2 did produce high absolute IPCA2 scores in series 2003, 2004 and 2005 (Figure 7.5) when ratoons of site G2 showed concomitant deviations along this axis. This implies that site G2 may be characterized by higher rainfall after canopy formation compared with the other two G sites.

The biplots with genotypic traits superimposed did not show any distinct, consistent clustering of genotypes into groups across all series (because different genotypes were used in each series). Instead, the genotypes were widely spread in each series, showing their diversity in genetic make-up and the level of diversity in their responses to the environments. The lack of the same genotype set in each series prevented the identification of traits that consistently caused separation of the genotypes on the biplots. The only genotype common to all series was NCo376, which showed consistent associations with the K environments and was always characterized by high TCANE. Nevertheless, some other trait relations and their associations with environments and covariates could be identified. For example, in seven out of eight series, FIB and POP were correlated to each other. As a group, FIB and POP were negatively correlated to SKEW in six out of eight series. This shows that high FIB and POP genotypes are characterized by straight stalks. As expected, SKEW and LOD were always positively correlated, as these are traits that are related to each other. The STKMS and DIAM were associated with each other in seven out of eight series, showing that thicker stalks are generally heavier. The TCANE and ERC were always negatively associated.

The joint examination of biplots for each series allowed for a number of biological interpretations. For example, in the 2000 series (Figure 7.4), the high RAIN1 associated with the K2 environments (left biplot) may have contributed to improved POP (right biplot). The high RAIN2 and WSI2 in the K environments thereafter ensured that STKMS was also improved. The improvement in these yield components (POP and STKMS) resulted in an overall improvement in TCANE, which determined the affinity of some genotypes to the K environments. In contrast, the G environments from the 2000 series were characterized by higher temperatures and radiation (left biplot), which lead to quicker canopy formation (right biplot). Furthermore, the low rainfall associated with the G environments resulted in higher

ERC. The above interpretations were also relevant to the 2001 and 2002 series (except that POP was associated more with the G environments in 2001 and 2002). In some series such as 2004 and 2005 (Figure 7.5), the K environments were not explicitly associated with higher rainfall but were still associated with high WSI2. This may be because of the deeper so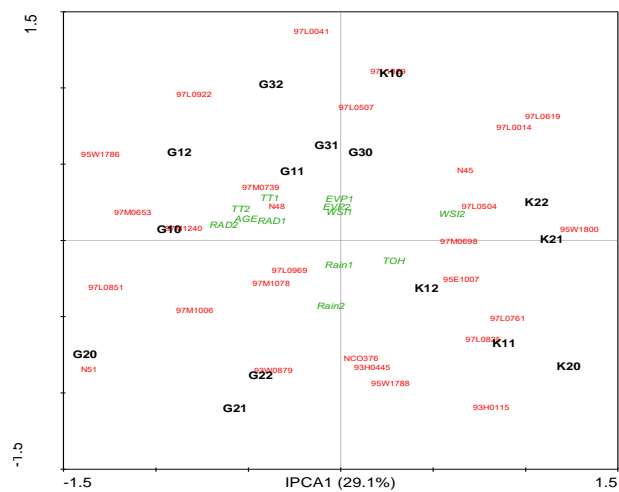ils and higher total available moisture (not shown) associated with the K sites in general, which most often produced higher TCANE.

The joint examination of the biplots for each series also allowed for some relationships between covariates and traits to be studied. For example, environments from site K2 were always associated with WSI2 on the covariate biplots and STKMS on the trait biplots. This is clear evidence of the effects that good moisture conditions after canopy formation may have on stalk mass. In contrast, environments from the G sites were most often associated with TT1 on the covariate biplots and CAN on the trait biplots. This represents the effects of high temperatures on the speed of canopy formation, which is a well-known relationship. Additionally, in all series, environments associated with low rainfall on the covariate biplots were also characterized by high ERC on the trait biplots. This represents the effects of moisture on the stimulation of vegetative growth (using up stored sucrose), resulting in a reduction in ERC.

For each covariate studied, the pre-and post-canopy splits were generally positively associated with each other on the biplots. An exception to this was series 2002 (Figure 7.4) where the pre-canopy evaporation and WSI was negatively associated to the corresponding post-canopy covariates. In this respect, the post-canopy evaporation and WSI was associated with ratoons of site K2, while the pre-canopy evaporation and water stress was associated with all the other environments. A similar response was observed in 2007 (Figure 7.6), where the pre-canopy WSI and rainfall was negatively associated with the corresponding post-canopy covariates. Once again, ratoons of site K2 were associated with higher post-canopy values.

Figure 7.4 AMMI2 biplots based on ERC yield of genotypes (red font) and environments (black font) evaluated in the long cycle of series 2000, 2001, and 2002. Environmental covariates (green font) are superimposed on the biplots on the left while genotypic traits (blue font) are superimposed on biplots on the right. (The percentage variance accounted for by each IPCA axis are indicted in parentheses)
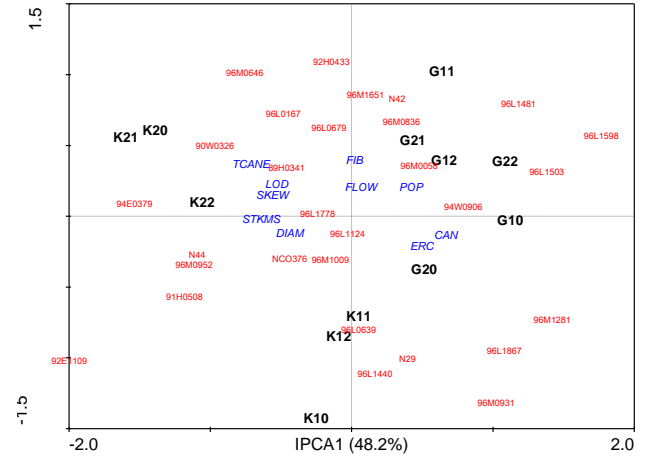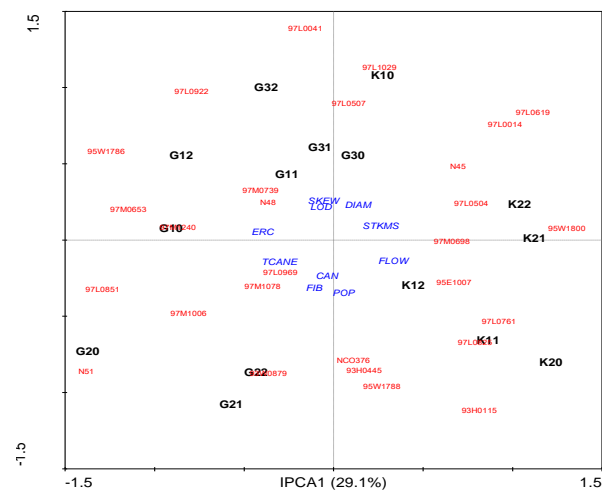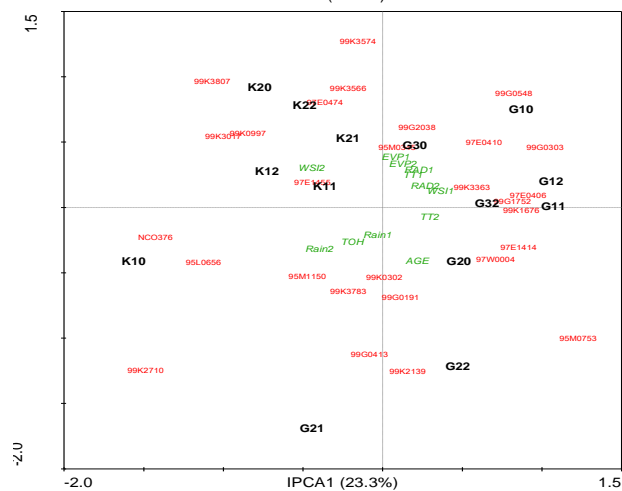
Figure 7.5 AMMI2 biplots based on ERC yield of genotypes (red font) and environments (black font) evaluated in the long cycle of series 2003, 2004, and 2005. Environmental covariates (green font) are superimposed on the biplots on the left while genotypic traits (blue font) are superimposed on biplots on the right. (The percentage variance accounted for by each IPCA axis are indicted in parentheses)
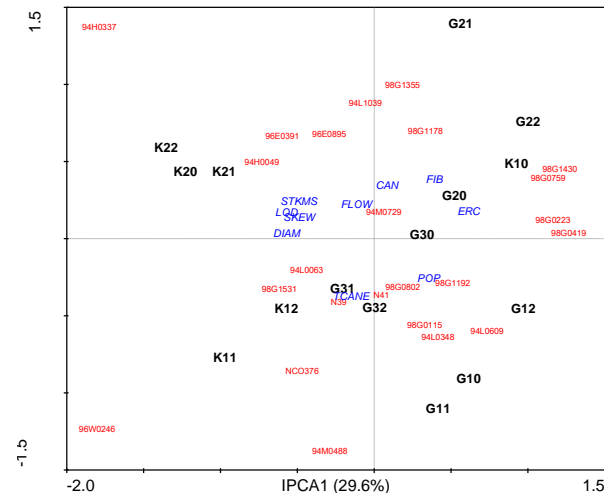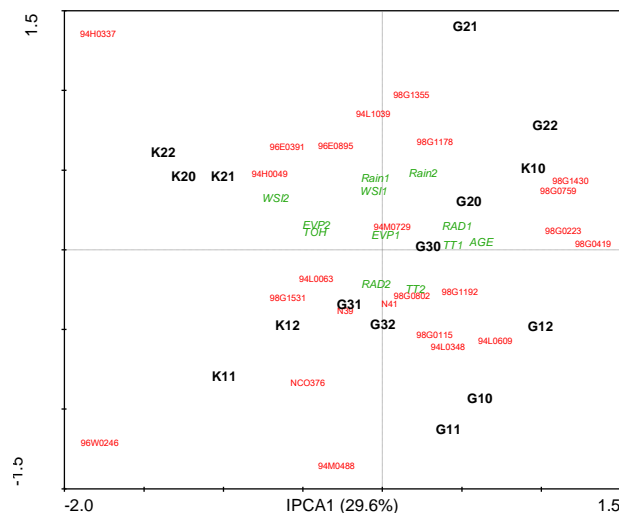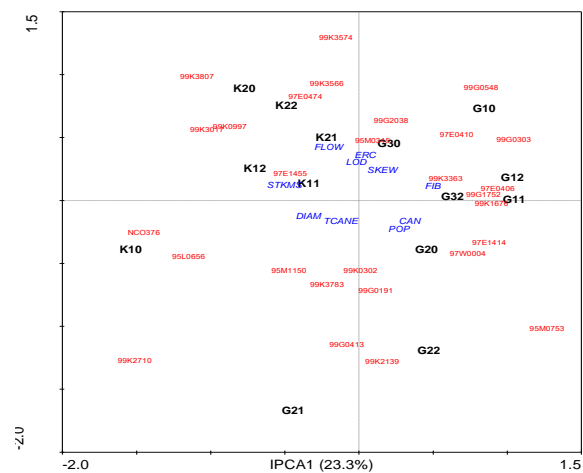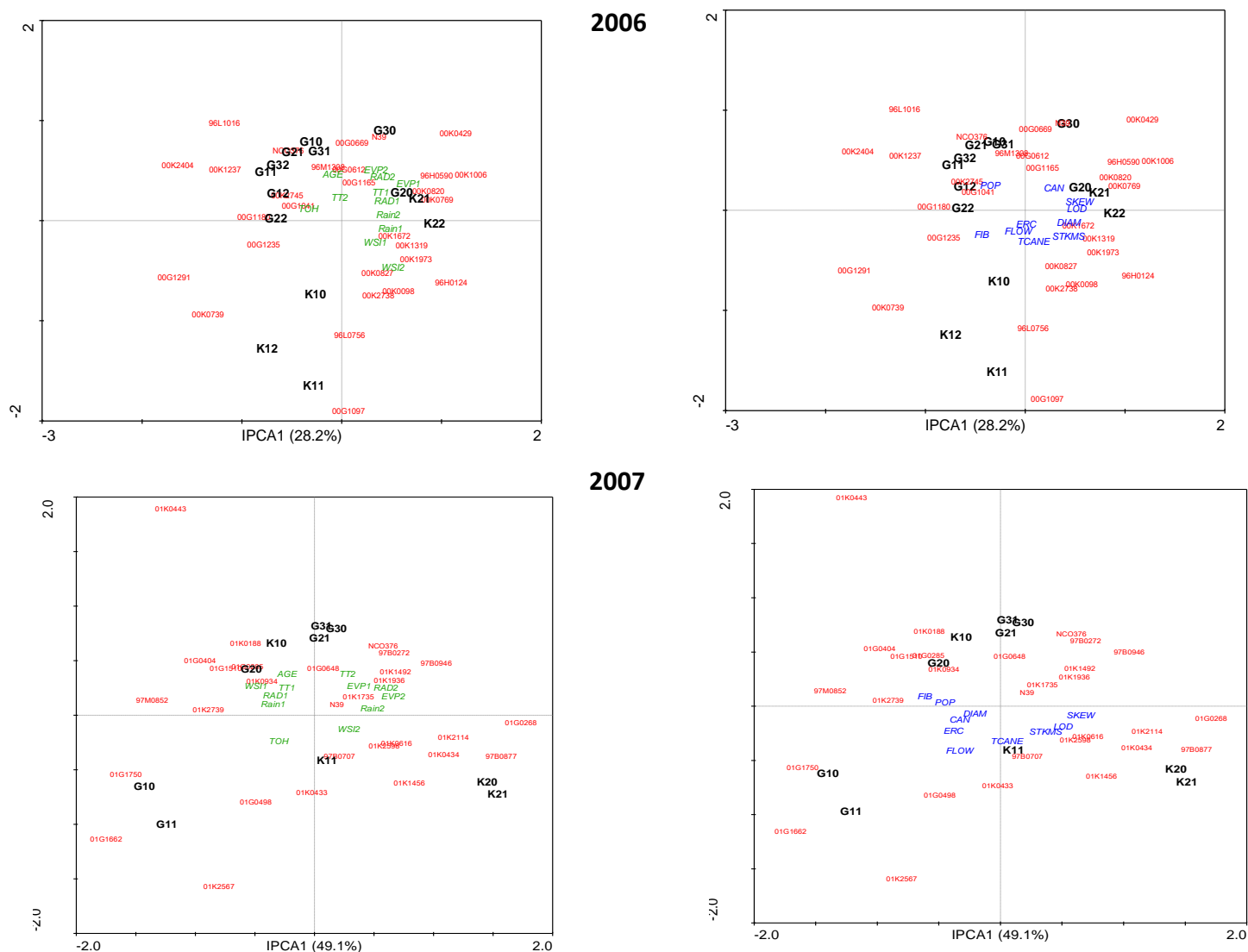
Figure 7.6 AMMI2 biplots based on ERC yield of genotypes (red font) and environments (black font) evaluated in the long cycle of series 2006 and 2007. Environmental covariates (green font) are superimposed on the biplots on the left while genotypic traits (blue font) are superimposed on biplots on the right. (The percentage variance accounted for by each IPCA axis are indicted in parentheses)

7.4.3.2 Short cycle

In keeping with the GGE biplot analysis, the short cycle G x E interactions were interpreted by averaging the genotype performance, environmental covariates, and genotypic trait values across ratoons for each trial. The AMMI2 biplots for the short cycle did not reveal any distinct and consistent clustering of the trial sites (Figures 7.7, 7.8, and 7.9). In 2001, 2002, and 2003, sites U1 and U2 were separated from sites T1 and T2 along the diagonals of the biplots, while in 2005 the T and U sites were separated along the vertical axis. In the other series, no consistent patterns of separation between the T and U sites were evident. In general, sites T1 and T2 were located in close proximity to each other on the biplots in five

144

(2002, 2003, 2006, 2007, and 2008) out of the eight series. In the four series where site T3 was evaluated, it separated from the other T sites on the biplots. Aside from the 2001 series, sites U1 and U2 were generally separated from each other, often located in different quadrants of the biplots.

In the first three series, the environmental covariates formed a single cluster (2001) or separated (2002 and 2003) into different clusters on the biplots according to the location of site U1 (Figure 7.7). In 2004, no distinct covariate clustering could be identified. However, in 2005, 2006, 2007 and 2008, the covariates formed two separate clusters; one of which was always correlated to site T3. The covariate cluster comprising TOH, TT1, RAIN1, RAD1, EVP1 and WSI1 were correlated to site T3 in 2005, 2006, 2007 and 2008. This was attributed to site T3 being dedicated to late season harvesting, which meant that crops were exposed to summer conditions (high temperatures, rainfall, radiation, evaporation etc.) before canopy formation. The same crops were subsequently exposed to winter conditions after canopy formation; hence the negative correlations between site T3 and TT2, RAD2, EVP2, and RAIN2. Interestingly, WSI2 was also correlated with site T3, suggesting that late season crops did not experience moisture stress after canopy formation. This may be primarily attributed to the low evaporative demand after canopy formation, as depicted by the negative correlation between site T3 and EVP2. Site U1 was separated from the other sites along the IPCA1 axis in most series. In all cases, U1 was also positively correlated with RAIN2 suggesting that this site generally received high rainfall after canopy formation. Aside from 2002 and 2003, site U2 did not show consistent associations with any covariates on the biplots.

Once again, the biplots with genotypic traits superimposed could not reveal consistent clustering of genotypes, as different genotypes were used in each series. The genotype NCo376 was common to all series, and again showed positive correlations with TCANE in six (2001, 2002, 2004, 2005, 2006, and 2007) out of eight series. In general, the trait relations identified in the short cycle biplots were almost identical to those observed in the long cycle. These included the positive correlations between FIB and POP; this was observed in all of the series. The SKEW and LOD were once again positively correlated with each other across all series. Similarly, STKMS and DIAM consistently showed strong correlations with each other. The TCANE and ERC were usually negatively correlated, however, in series 2001, 2003, and 2005, these two traits showed some correlation. This may be because of the use of

genotypes that were 'well-balanced' in yielding ability and quality characteristics in those series.

The joint examination of the covariate and trait biplots for each series also allowed for some relationships between covariates and traits to be studied. For example, in 2002 and 2003, site U1 was positively correlated to WSI1, RAIN1, RAIN2, EVP2, TT2, WSI2 and RAD2. Genotypes that performed well at site U1 in these series were subsequently characterized by high TCANE, which is a trait known to be influenced by favourable conditions brought about by the above covariates. In contrast, sites T1, T2, and U2 were not explicitly associated with any of the above covariates in the 2002 and 2003 series, and these sites were subsequently negatively associated with TCANE. In 2005, 2007, and 2008, site T3 (late season harvest) was associated with TT1, RAD1, EVP1 and RAIN1. Favourable conditions before canopy formation may have stimulated the tillering phase, which normally results in lower tiller numbers (due to shorter phase duration). Consequently, genotypes that were adapted to site T3 were generally those that were negatively correlated to final stalk population (POP) in series 2005, 2007 and 2008. It was interesting to note that site T3 was always negatively correlated to FIB, suggesting that late season harvesting may not be favourable when fibre is a desired product.
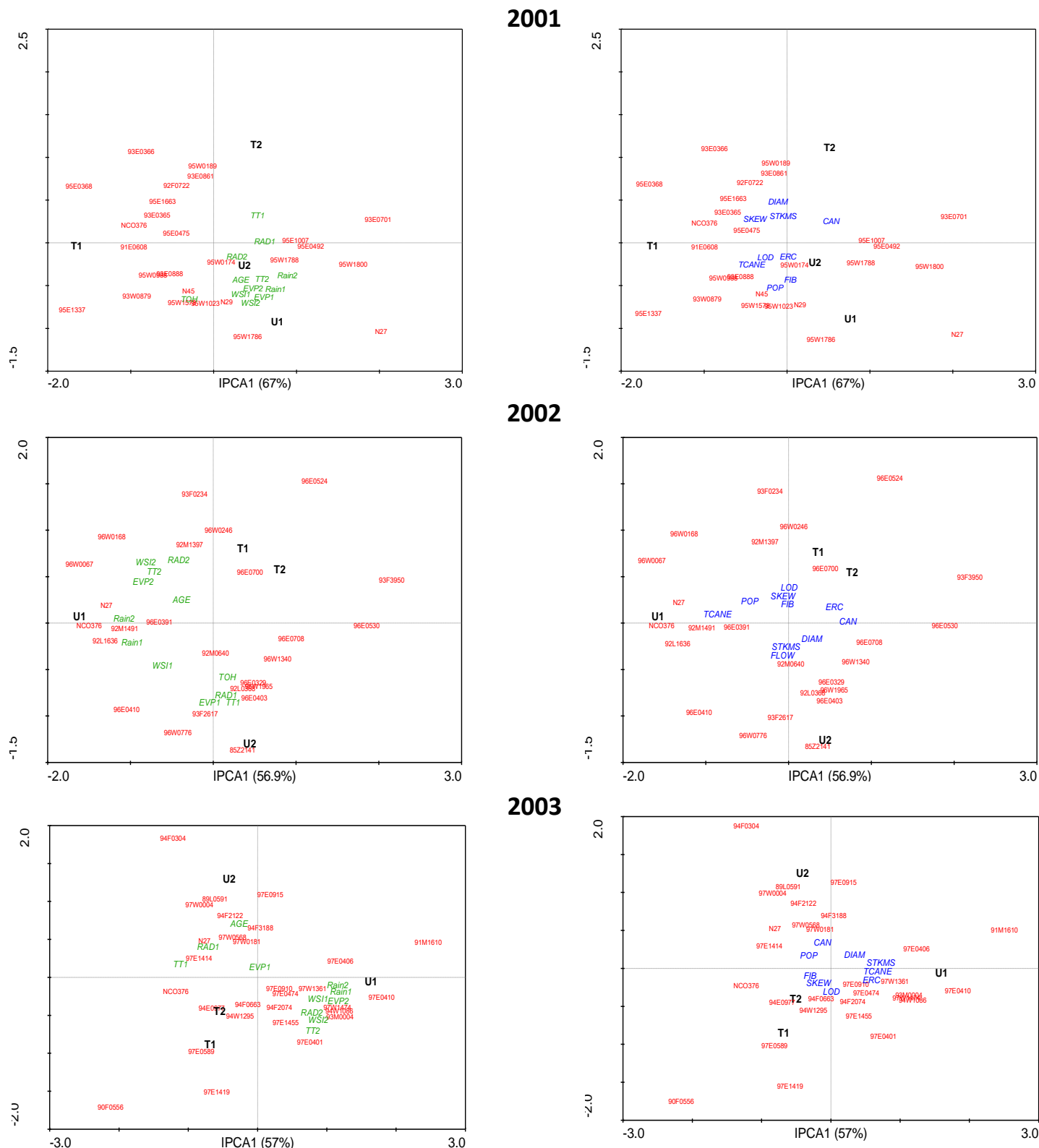
Figure 7.7 AMMI2 biplots based on ERC yield of genotypes (red font) and environments (black font) evaluated in the short cycle of series 2001, 2002, and 2003. Environmental covariates (green font) are superimposed on the biplots on the left, while genotypic traits (blue font) are superimposed on biplots on the right. (The percentage variance accounted for by each IPCA axis are indicted in parentheses)

Figure 7.8 AMMI2 biplots based on ERC yield of genotypes (red font) and environments (black font) evaluated in the short cycle of series 2004, 2005, and 2006. Environmental covariates (green font) are superimposed on the biplots on the left, while genotypic traits (blue font) are superimposed on biplots on the right. (The percentage variance accounted for by each IPCA axis are indicted in parentheses)
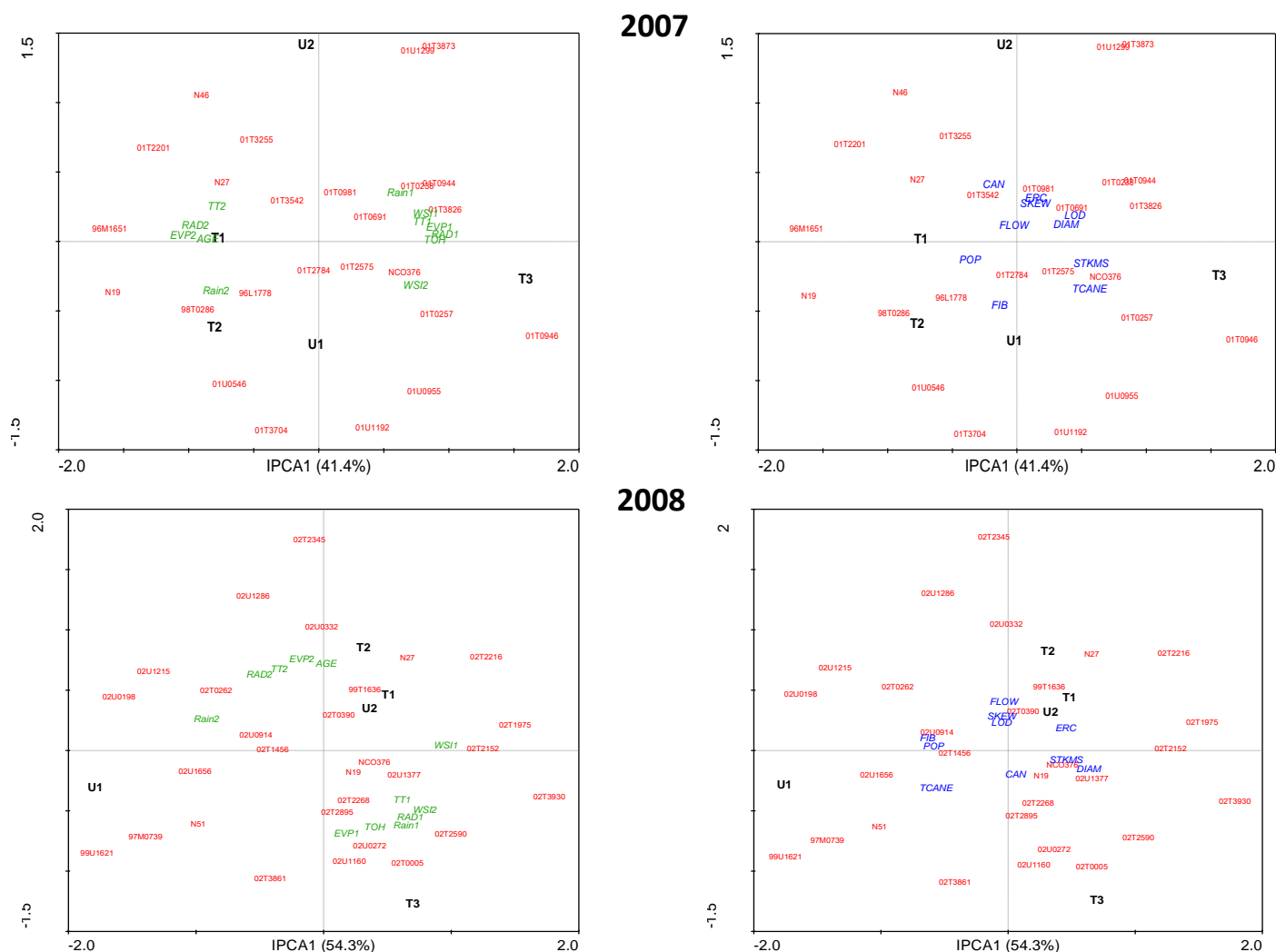
Figure 7.9 AMMI2 biplots based on ERC yield of genotypes (red font) and environments (black font) evaluated in the short cycle of series 2007 and 2008. Environmental covariates (green font) are superimposed on the biplots on the left, while genotypic traits (blue font) are superimposed on biplots on the right. (The percentage variance accounted for by each IPCA axis are indicted in parentheses)

7.5 Discussion

The variance components analysis showed that the G x E interaction accounted for more variation than the main effects of genotype across all series on the long cycle. In contrast, some series on the short cycle showed the opposite trend i.e. genotype accounted for more variation than the G x E interaction. This implies that selection of genotypes for superior TERC may be more feasible on the short compared with the long cutting cycle along the

coast. Also, testing across sites was less important on the short cycle compared with the long cycle, where complex G x E interactions were more evident. Zhou et al. (2011) conducted an extensive study of variance components associated with selection programs along the coast and reached similar conclusions. Furthermore, on the long cycle the repeatable component of G x E (i.e. genotype x site) accounted for more variation than the combined effects of genotype x ratoon and genotype x site x ratoon. This implies that this component is exploitable and homogenous regions can be identified. Indeed, this was shown clearly by the GGE biplot analysis, where the G and K sites frequently formed separate clusters. This study, which has complemented that of Zhou et al. (2011) by further identifying mega-environments, has shown that the long and short cycles must be treated differently when considering any network re-structuring.

The GGE biplot analysis has proven to be an excellent tool for the investigation of test site similarity. It was apparent from the long cycle biplots that the K and G sites were different from each other and should therefore remain separate. However, within the K and G groups, there was clear evidence of site redundancy. Site K1 was less discriminative than K2, was always located within the same ME as K2, and occasionally clustered with G sites. This suggests that K1 is redundant and should be considered for removal from the trial network. However, K1 is a SASRI-owned selection farm, and the practicalities of staying with such a site may possibly outweigh its redundancy within the network. Plant breeders may have to conduct further evaluations in this regard to arrive at a final recommendation. All of the G sites were positively correlated to each other in all series and in general, site G1 and G2 were the most similar. As G1 is also a SASRI-owned selection site, it is therefore recommended that site G2 be dropped from the trial network instead in favour of an alternative site.

The investigation of site redundancy was less straightforward for the short cycle. Site U1 did not appear to be discriminative, however, it did constitute a ME in one series. It is therefore recommended that further studies be initiated before a decision is taken on the retention of site U1 within this program. Site T2, however, was found to be very similar to site T1 across all series and could therefore be removed from the trial network. Additionally, T3 also showed similarities to site T1. Site T3 represents a different time of harvest (relative to T1) on the SASRI-owned selection farm. The GGE biplot analysis in this study has shown that T1 (early harvest) and T3 (late harvest) may be similar. However, the AMMI biplots showed that T1 and T3 were separated on the biplots. The discrepancy lies in the fact that GGE biplot performs singular value decomposition on the G + G x E matrix after removal of the

150

environment means only (Yan et al., 2000) while AMMI removes the genotype and environment means (Gauch, 1992). This implies that the similarities between sites T1 and T3 on the GGE biplots are also influenced by genotypic effects. Given that Chapter 6 showed clear differential responses between early and late harvests at the same site, any recommendations about the similarity of T1 and T3 should only be made after further investigation. In general, however, GGE biplot and AMMI produced comparable results, and similar conclusions could be made with regard to site similarity using both methods.

The AMMI biplots with covariates and traits superimposed suggested that the G sites were characterized by lower yield potential conditions compared with the K sites. This was supported by the response of genotype N33 in the 2000 series. This genotype was released to the industry specifically for low potential conditions, and its affinity for the G environments in the 2000 series (Figure 7.4) confirms this. Also, the lack of a consistent genotype set across series prevented an accurate estimate of which traits were most influential on the separation of genotypes on the biplots. In general, the trait relations were fairly consistent from one series to the next, and any deviation in trait relations could be attributed to variations in the trait profiles of the specific genotype sets. Despite the use of different genotype sets, traits such as TCANE, FIB, SKEW, and ERC, were characterized by larger IPCA scores than the other traits on average, and also produced the most number of significant ($P<0.05$) correlations with the IPCA scores (not shown). This is very important, as one of the requirements of superimposing covariates onto AMMI biplots is that the correlation between the IPCA scores and the covariates are strong (Voltas et al., 2005). Additionally, this study has shown that standard agronomic traits that are routinely gathered from sugarcane METs can be superimposed onto an AMMI biplot for visualization of relative trait importance. This suggests that there may be potential to include even more detailed physiological and genetic trait measurements into such biplots in future.

The successful growth-phased approach to interpret G x E interactions in Chapter 6 prompted the expansion of this approach to a larger dataset in this study. However, the previous chapter revealed that the period of ripening in sugarcane was difficult to isolate, as it is a continuous process that occurs within the stalk and it is therefore very difficult to establish when ripening commences. It was therefore more logical to split the growth phases into two, using the crop model to estimate time to canopy for different crops. This approach has not been attempted in sugarcane previously. One of the limitations of this approach is that different genotypes exhibit different rates of canopy formation and using an average time to canopy formation for

a trial may be misleading. One way of overcoming this is to have an estimate of time to canopy formation for each genotype. In this study the CAN, trait was a general rating of canopy density at a specific time (usually done once in the growing season). Future studies should investigate the possibility of converting this relatively easy rating into an estimate of time to canopy formation by taking two or three CAN ratings at different times during the establishment phase.

In general, it was found that for each covariate, the pre- and post-canopy splits were associated with each other. Aside from two exceptions (series 2002 and 2007), the general correlation of the pre-and post-canopy covariates across all series show that either one of the splits, or even a seasonal average for the covariates would have been sufficient to interpret the responses. This is in contrast to the previous chapter, where it was shown that the covariates summarized within three growth phases gave better interpretations of the G x E interactions. The lack of separation of pre- and post-canopy covariates in this study (compared to the previous chapter) may be because of the use of more diverse environments that are characterized by large differences in climate. In other words, the differences in climatic conditions between sites had an overriding effect relative to the differences between pre- and post-canopy growth phases. This was indeed the case with the long cutting cycle where the G sites were generally characterized by high temperatures, low rainfall, high radiation, and high water stress, irrespective of the growth phase. In the short cycle, however, the separation of the pre- and post-canopy covariates on the 2005, 2006, 2007, and 2008 biplots were explicit, suggesting once again that the splitting of the covariates may lead to useful interpretations when differences between environments are subtle.

Past studies that have used the approach employed here to interpret G x E interactions mainly used single data sets, where specific sites were tested once only (Van Eeuwijk and Elgersma, 1993; Jackson et al., 1995; Vargas, et al., 1999; Voltas et al., 1999). This study is the first to attempt this approach across a range of trial series where the same trial sites are used across many seasons. This study has consequently evaluated the repeatability of this method of G x E interpretation. The consistent association of environmental covariates with certain sites (K sites and WSI2, G sites and TT1 etc.) across all series does indeed show that this approach can produce repeatable results. Additionally, other methods, such as partial least squares regression (PLS), have been used in past studies to incorporate environmental covariates and genotypic traits into G x E biplots (Vargas et al., 1999; Joshi et al., 2010). However, such a technique requires extensive understanding of complex matrix computations which are not

familiar to most crop scientists. Furthermore, when large numbers of environments, genotypes, environmental covariates, and genotypic traits are available, the PLS biplot may become difficult to interpret. Hence, the separation of the environmental covariates and genotypic traits into separate biplots in this study was seen as a more practical approach.

7.6 Conclusions

The variance components analysis conducted in this study has provided valuable information on the relative contributions of different components to the overall variation within the long and short cycle networks. Greater selection gains are more likely to be observed in the short cycle compared with the long cycle, because of the greater influence of G x E interactions in the latter. The G x site component for the long cycle accounts for large variation, suggesting that testing across sites on this cycle is sufficient. However, across-site testing on the short cycle did not account for a large proportion of variation, suggesting that more effort should be placed on identifying more diverse test sites.

The GGE biplot analysis identified sites K1 and G2 from the long cycle for possible removal from the trial network in favour of more diverse test sites. Similarly, site T2 from the short cycle did not provide any new information of genotype performance across eight series and can therefore be removed from the network. It is further recommended that follow-up studies investigating similarities between sites T1, T3, and U1 be initiated once data from a few more series become available. The factors responsible for differential genotype performance at sites on the long cycle include temperature, evaporation, rainfall, radiation, and water stress caused by differences in soil properties. The G sites may be characterized as low yield potential relative to the K sites on the long cycle. The computation of covariates based on pre- and post-canopy growth did not add much value to the interpretations in this dataset. However, further refinement of this technique may provide more comprehensive interpretations.

This was the first study to attempt an integration of a sugarcane crop model within a conventional G x E analysis framework. The use of the model to compute the water stress index was successful and will be applied to further sugarcane studies. There are many external factors and traits that can affect G x E interactions. However, this study has shown that a relatively simple environmental characterization using soil and weather data, coupled with the use of a crop model and basic agronomic traits, may help identify key aspects of

sugarcane growth causing superior performance of genotypes in different environments. The use of empirical and analytical methods that complement each other synergistically has been illustrated, and other sugarcane studies may benefit from the principles and approaches used here.

7.7 References

Anonymous., 2009. GenStat® Executable Release 12.1. Lawes Agricultural Trust, Rothamstead Experimental Station/ Clarendon Press, Harpenden, U.K/London, U.K.

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

Gauch H.G., 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam.

Gilmour, A.R., Thompson, R., Cullis, B.R., 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51, 1440-1450.

Jackson, P., McRae, T., Hogarth, M., 1995. Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. *Field Crops Research* 43, 119-130.

Joshi, A.K., Crossa, J., Arun, B., Chand, R., Trethowan, R., Vargas, M., Ortiz-Monasterio, I., 2010. Genotype x environment interaction for zinc and iron concentration of wheat grain in eastern Gangetic plains of India. *Field Crops Research* 116, 268-277.

Patterson, H.D., Thompson, R., 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545–554

Van Eeuwijk F.A. and Elgersma A., 1993. Incorporating environmental information in an analysis of genotype by environment interaction for seed yield in perennial ryegrass. *Heredity* 70, 447-457.

Van Oosterom E.J., Mahalakshmi V., Bidinger F.R. and Rao K.P., 1996. Effect of water availability and temperature on the genotype-by-environment interaction of pearl millet in semi-arid tropical environments. *Euphytica* 89, 175-183.

Vargas, M., Crossa, J., Van Eeuwijk, F.A., Ramirez, E., Sayre, K., 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype x environment interaction. *Crop Science* 39, 955-967.

Voltas J., Van Eeuwijk F.A., Sombrero A., Lafarga A., Igartua E. and Romagosa I., 1999. Integrating statistical and ecophysiological analyses of genotype by environment interaction for grain filling of barley I. Individual grain weight. *Field Crops Research* 62, 63-74.

Voltas, J., Lopez-Corcoles, H., Borras, G., 2005. Use of biplot analysis and factorial regression for the investigation of superior genotypes in multi-environment trials. *European Journal of Agronomy* 22, 309-324.

Yan, W., Hunt, L.A., 2001. Interpretation of genotype x environment interaction for winter wheat in Ontario. *Crop Science* 41, 19-2.

Yan, W., Hunt, L.A., Sheng, Q., Szlavnics, Z., 2000. Cultivar Evaluation and Mega-Environment investigation based on the GGE Biplot. *Crop Science* 40, 597-605.

Zhou, M., Joshi, S., Maritz, T., Koberstein, H., 2011. Components of genotype by environment interaction among SASRI regional breeding and selection programmes and their implications. *Proceedings of the South African Sugar Technologists Association* 84, 363-374.

CHAPTER 8

GENERAL DISCUSSION AND CONCLUSIONS

Prior to this study, information on the nature and causes of G x E interactions in the rainfed parts of the sugar industry was severely lacking. Given the diverse conditions, this study was initiated to conduct a systematic analysis of the G x E interactions through application of novel approaches that may add value to future sugarcane METs. Chapters 3 and 4 of this study were therefore seen as broad investigations of the G x E interactions across the rainfed regions and the purpose of such studies were to direct or inform the approach that would be followed for the detailed G x E studies within the different regions. To this end, these chapters illustrated the complex nature of the G x E interactions in the rainfed parts and showed how post-release recommendation trials can be utilized to ultimately inform pre-release selection programs. Chapters 3 and 4 identified that the midlands region was unique, while the coastal and hinterland regions showed some similarity and could subsequently be analysed together. Following this combined analysis in Chapter 7, it was concluded that the C and H programs were in fact different. The reason for the difference in results between Chapter 3 and Chapter 7 lies in the difference in data sets used. In Chapter 3, released cultivars with stable performance were evaluated whereas in Chapter 7, unreleased, diverse genotypes were tested.

Chapter 3 also showed how ratoons from the same trials and trials from the same location showed variable genotypic responses, thereby highlighting the need to characterize the METs in relation to environmental factors. Furthermore, Chapter 3 also highlighted the possibilities of site similarities (e.g. humic vs. sandy sites in the midlands) as well as the uniqueness of sites in frost pockets. These observations were subsequently confirmed in Chapter 5 when the midlands region was studied in detail. This therefore illustrates how important broad studies are, and how valuable they can be at flagging issues that need further attention when conducting a systematic analysis of G x E interactions at an industry/regional level. Chapters 3 and 4 therefore provided a base from which more detailed studies and some novel approaches to interpret G x E interactions could be conducted. This was despite the use of highly selected, relatively stable commercial cultivars in the G x E studies in Chapters 3 and 4.

The statistical methods used in this study were based on multivariate techniques that produced visual outputs (biplots) from which G x E interactions could be investigated. Given the large quantities of data originating from numerous selection programs in the industry, the multivariate methods were chosen based on the ease of interpretation. More statistically robust methodologies may have been more accurate; however, it is believed that many of the patterns and trends that were observed in these studies may have gone unnoticed. Of particular value was GGE biplot analysis, which was ideal for the evaluation of test site similarity and identification of homogenous MEs (particularly in Chapters 5 and 7). The inability to evaluate statistical significance of genotype and environment co-ordinates on the biplot is one of the limitations of GGE biplot (Yang et al., 2009). Therefore, despite the evaluation of test site similarity across numerous trial series in Chapters 5 and 7, any decisions about the removal of certain sites from the trial networks will only be made after further statistical evaluations and practical considerations. Still, GGE biplot analysis allowed for quick visual examinations of site similarities in this study and its use in other sugarcane G x E studies is recommended. Furthermore, its use as a tool to identify genotypic traits responsible for yield variations in different environments was clearly illustrated in Chapter 6, where a covariate effect biplot (Yan and Tinker, 2005) was produced.

Despite the demonstrated applications of GGE biplot to identify causes of G x E interactions (Yan and Hunt, 2001), the preferred method of interpreting interactions in this study was AMMI analysis instead. This was mainly because of the numerous demonstrated successful uses of AMMI to interpret G x E interactions in a range of crops (Van Eeuwijk and Elgersma, 1993; Van Oosterom et al., 1996; Romay et al., 2010). The methodology involves correlating AMMI IPCA scores to environmental covariates and genotypic traits, and further facilitating interpretation by superimposing these covariates and traits onto the AMMI biplots. The technique proved successful in identifying key covariates affecting G x E interactions across regions in Chapter 4 and when identifying factors causing genotype x time of harvest interactions (Chapter 6). Correlations between covariates and IPCA scores were generally weaker in Chapter 7. Nevertheless, the enrichment of the AMMI biplots with the covariates and genotypic traits in Chapter 7 allowed for a comprehensive interpretation that may not have been possible using other methods. Another key advantage of using this approach is that it is relatively simple to understand and implement as opposed to other more complicated analyses like partial least squares regression (Vargas et al., 1998). Furthermore, the use of

AMMI served to confirm many of the G x E patterns that were initially observed using GGE biplot analysis, thereby adding confidence to the recommendations.

The pattern analysis approach to interpret G x E interactions proposed in Chapter 5 is a novel way to understand general sugarcane growth responses to varying types of environments. The approach may be used in preliminary studies, where information is required on the trait profiles of germplasm collections and the types of environments characterising an industry. Once identified, representative genotypes in each set may be used in detailed growth experiments. Such experiments should be conducted in environmental conditions typical of the environment groups identified. This will allow for the growth of different genotype sets to be modelled as opposed to modelling the growth of specific genotypes, which is difficult in a genetically complex crop like sugarcane. The pattern analysis and grouping strategies proposed in this study therefore have potential beyond conventional G x E research, as it can assist in selecting representative genotypes (and environments) for detailed physiological and crop modelling research.

The use of variance components analyses in this study was primarily aimed at evaluating the relative contributions of different sources to yield variation. In Chapter 3, the variance components analysis provided insights into the variation accounted for by different components across regions. That study was the first to highlight that the genotype x ratoon term accounted for less variation than the site x ratoon interaction in the industry. This is especially relevant in an industry where there is a general perception that variation in ratoon performance is more dependent on genotype than site conditions. This point was highlighted once again by the variance components analysis in Chapter 7, which also indicated that selection gains are more likely on the short cutting cycle compared with the long cutting cycle. This has indeed been the case in the industry during the last few years as fewer commercial genotypes have originated from long cycle compared with the short cycle. The use of variance components to determine optimal numbers of trials, replicates etc. was out of the scope of this study and may be addressed in follow-up studies to Zhou et al. (2011) in order to optimize the efficiencies of the trial networks.

In general, the combination of statistical methods adopted in this study highlighted the different approaches that could be used for the analysis and interpretation of sugarcane METs. In most instances, the nature of the datasets determined the most appropriate analysis to employ. The main factors responsible for G x E interactions varied slightly from one

chapter to the next. This may have been because of the differences in methodologies. However, it is more likely attributable to differences in production conditions from one region to the next. An ongoing objective throughout the chapters of this study was to evaluate the appropriateness of multivariate methods for the interpretation of G x E interactions. In retrospect, it is apparent that no single methodology was appropriate across all datasets. Instead a combination of approaches, depending on the nature of the data, seemed more appropriate.

One of the novel features of this study was the use of weather data to supplement conventional sugarcane METs. Most sugarcane G x E studies have been largely empirical in nature (Queme et al., 2007; Jackson et al., 2007), and aside from a single analytical study (Jackson et al., 1995), no other attempts had been made to interpret sugarcane G x E interactions on a large scale. In many instances, the lack of appropriate weather stations at trial sites is a limitation to such interpretive studies. In fact, the weather data used in this study was also not ideal. At times, data had to be used from weather stations that were located at a distance from the actual trial sites, which may have compromised the accuracy of some of the analyses. In some cases data from the same weather stations were used for two different sites (with trials on the same cutting cycles). As a result, any grouping of sites based on weather data may have been attributable to the origin of the data rather than the true similarity of the sites being tested. Consequently, any future attempts to adopt the methodologies used in this study should ensure that weather data are sourced from an appropriately positioned weather station. For the selection programs investigated in this study, it is therefore recommended that an automatic weather station be established at each testing site.

In addition to the use of basic weather data, one of the novel approaches that featured prominently in this study was the use of crop growth models to characterize sugarcane METs. Such an approach has not been used previously in sugarcane and it has provided valuable insights and ideas that future studies can build on. Collectively, the studies reported in the individual chapters represent the largest ever simulation of sugarcane METs. These simulations were done using a single approach throughout the study. The accuracy of the simulations might have been improved had different approaches been tested. However, "optimizing the use of crop models to characterize METs" was out of the scope of this particular study. Chapter 5 did touch on this issue by testing different water stress indices derived from crop models, but in general, the different possibilities of characterizing trials

using models were not evaluated here. Aspects that should be tested in future include a comparison of model simulations using visual estimates of effective rooting depth (done throughout this study) vs. actual soil textural measurements conducted to depth to determine rooting depth. The latter method may provide a better estimate of the total available moisture which may result in more accurate simulations. Another application of crop models to G x E research would be to treat the simulated yields as the yield of a hypothetical genotype in each trial, and subject that dataset to conventional G x E analyses. This will investigate the bias of crop models to different environments, as well as the performance of other genotypes relative to the hypothetical genotype. Furthermore, if the growth parameters associated with the simulations are known (as they are with all crop models), then it may be possible to use it as a probe genotype to study the characteristics of other genotypes in the dataset. Another aspect needing attention is to investigate the effects of increasing distances from weather stations, and how this affects the correlations between actual and simulated yields. These are just some of the ideas that have evolved from this study which will be used to optimize the use of crop models in sugarcane G x E analyses in the future.

Chapters 4 and 5 suggested that the computation of environmental covariates be aligned with specific growth phases, as seasonal averages may have been misleading. In Chapter 6 this was tested (using a small dataset), and showed that relevant biological interpretations were possible when environmental factors were summarized within three growth phases. However, in Chapter 7 (larger dataset), the splitting of the growth phases did not appear to add much value to the interpretations on the long cycle. This was attributed to the more diverse set of environments that were studied in Chapter 7, where the general climatic differences between sites had an overriding effect on the subtle differences that may have existed between growth phases. Conversely, when the short cycle was analysed in Chapter 7, the pre- and post-canopy splits of the environmental covariates were explicitly associated with the late and early harvests, respectively. Therefore, it would appear that the splitting of the sugarcane growth cycle does have value when climatic differences between environments are subtle (as was the case in the short cycle dataset in Chapter 7). Further testing on a large scale with diverse environments and more appropriate weather data is needed to evaluate whether a general splitting of the growth phases will be valuable in future G x E studies. The strategy of using the crop model to estimate the growth phases has not been attempted in other studies previously, and future sugarcane studies should consider refining this approach.

In this study, the effects of environmental covariates on the G x E interactions were studied within selection programs that were already well established (Chapters 5 and 7). This implies a certain degree of homogeneity within the programs in soil and climate, which would have been the criteria for site selection upon establishment. It is hypothesised that the effects of environmental covariates would have been more pronounced had evaluation occurred across selection programs rather than within. Evidence of this is shown in Chapter 4, where post-release trials from different regions were tested. In that study, the environmental covariates produced large significant correlations with AMMI IPCA scores. This is in contrast to the size and significance of correlations observed between covariates and IPCA scores in the within-program analyses in Chapter 7. One of the limitations of conducting such analyses across diverse regions is the lack of genotype-balanced datasets across years and sites. Future sugarcane studies aimed at understanding the causes of G x E interactions should therefore consider applying these methodologies across selection programs rather than within programs. In the South African industry, it is recommended that a genotype-balanced MET be established across the different regions to follow up on some of the concepts outlined in this study.

An aspect that featured prominently in the chapters of this study was that of sugarcane trait relations. In Chapter 5, it was shown that the trait relations between genotype sets did not vary much from one environmental group to the next. This suggests that the differences in environmental conditions were not large enough to alter trait relations, or there was low variability in trait expression in the population of genotypes studied. In general, the chapters in this study illustrated and confirmed some valuable trait relations that may be utilized in selection programs. For example, the strong correlation between CAN, ERC, and DIAM, which were generally negatively correlated to traits like FIB and POP. Some of the trait relations were consistently strong, while others varied in strength from one dataset to the next. It was interesting (Chapter 6) to note that POP had a large effect on TCANE when harvesting early in the season, while STKMS had a larger effect on TCANE when harvesting late in the season. In Chapter 7, the short trials were mostly harvested early in the season and the biplots revealed that POP and TCANE were always positively correlated. This shows that even when using two different types of analyses and two different datasets, the same conclusions could be reached about how the trait relations varied, and this adds to the validity of the strategies employed. The long cycle trials in Chapter 7 were harvested at 18 months of age, implying that some crops were harvested early while others were harvested late in the

season. This confounding effect of age and time of harvest should be investigated further. Given the complex combinations of age and time of harvest possible in the South African sugar industry, it is recommended that "cropping scenarios" be defined for future studies.

In Chapter 7, the trait relations were also fairly consistent within selection programs, however, when comparing the short and long cycle programs there were deviations in trait relations. For example, on the long cycle, the relationship between TCANE and POP was generally weak across series, while TCANE and STKMS seemed to show a stronger relationship. On the short cycle, TCANE and POP were associated across most series; while TCANE and STKMS showed negative associations in five series. This illustrates the relative influence of different yield components on TCANE when different cutting cycles are implemented and shows how trait selection strategies have to be adapted when age at harvest is increased in sugarcane. The differences in trait relations described above may be the result of the greater incidences of Eldana (*Eldana saccharina spp.*) damage associated with the long cutting cycle. Eldana damage generally increases in older sugarcane and stalk boring results in the reduction in STKMS. It is hypothesised that the strong dependency of TCANE on STKMS on the long cycle is because of this effect of Eldana. In fact, the effects of Eldana and other pests and diseases were not taken into consideration throughout this study. This shortcoming is mainly attributable to the lack of pest and disease ratings in selection trials. Such biotic factors can have massive influence on the G x E interactions (Yan and Hunt, 2001) and any further attempts at understanding causes of G x E interaction in sugarcane should account for the effects of pests and diseases as well.

In general, the different methods of studying trait relations in this study proved to be accurate and comprehensive. However, future studies should consider using these techniques to analyse relations between traits defining certain lower level plant processes typically used in crop modelling. Most physiological and crop modelling research is done in isolation, somewhat detached from G x E research typically relevant to agronomists and plant breeders. In order to integrate these disciplines, the principles and approaches illustrated in this study are relevant, and may be downscaled to lower level plant processes to facilitate further understanding of sugarcane growth and development. To accomplish this, sugarcane METs must be comprehensively characterized in relation to phenotype, genotype, and environment. This should not be seen as the role of the plant breeder exclusively, but rather a joint effort including agronomists and physiologists as well. In this respect, more integrated studies of

this nature are necessary, especially in crops like sugarcane where G x E studies in general are uncommon.

The way forward following this study will be to immediately address the site similarity issues that were identified in the different selection programs. This may require further practical considerations before final decisions are made. The massive trial datasets that were generated and characterized in this study will be expanded through inclusion of data from current trial series. These datasets will then be explored further to investigate breeding progress and inform selection strategies. Further studies will also be conducted to try and optimize the use of crop models and weather data for the characterization of sugarcane METs. And finally, attempts will be made for the routine inclusion of more detailed crop growth measurements in selection trials to enhance their value to other research disciplines and contribute to the understanding of sugarcane G x E interactions.

8.1 References

Jackson, P., Chapman, S., Rattey, A., Wei, X., Cox, M., 2007. Genotype x region interactions, and implications for sugarcane breeding programs. *Sugar Cane International* 25, 3-7.

Jackson, P., McRae, T., Hogarth, M., 1995. Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. *Field Crops Research* 43, 119-130.

Queme, J.L., Crossa, J., Orozco, H., Melgar, M., 2007. Analysis of genotype-by-environment interaction for sugarcane using the sites regression model (SREG). *Proceedings of the International Society of Sugar Cane Technologists* 26, 764-769.

Romay, C.N., Malvar, R.A., Campo, L., Alvarez, A., Moreno-Gonzalez, J., Ordas, A., Revilla, P., 2010. Climatic and genotypic effects for grain yield in maize under stress conditions. *Crop Science* 50, 51-58.

Van Eeuwijk, F.A., Elgersma, A., 1993. Incorporating environmental information in an analysis of genotype by environment interaction for seed yield in perennial ryegrass. *Heredity* 70, 447-457.

Van Oosterom, E.J., Bidinger, F.R., Mahalakshmi, V., Rao, K.P., 1996. Effect of water availability pattern on yield of pearl millet in semi-arid tropical environments. *Euphytica* 89, 165-173.

Yang, R., Crossa, J., Cornelius, P.L., Burgueno, J., 2009. Biplot analysis of genotype x environment interaction: Proceed with caution. *Crop Science* 49, 1564-1576.

Yan, W., Hunt, L.A., 2001. Interpretation of genotype x environment interaction for winter wheat in Ontario. *Crop Science* 41, 19-25.

Yan, W., Tinker, N.A., 2005. An integrated bipot analysis stem for displaying, interpreting, and exploring genotype x environment interaction. *Crop Science* 45, 1004-1016.

Vargas, M., Crossa, J., Sayre, K., Reynolds, M., Ramirez, M.E., Talbot, M., 1998. Interpreting genotype x environment interaction in wheat by partial least squares regression. *Crop Science* 38, 679-689.

Zhou, M., Joshi, S., Maritz, T., Koberstein, H., 2011. Components of genotype by environment interaction among SASRI regional breeding and selection programmes and their implications. *Proceedings of the South African Sugar Technologists Association* 84, 363-374.

SUMMARY

Information on the nature and causes of G x E interactions in the rainfed parts of the South African sugar industry were lacking. The aim of this study was to systematically analyse, identify causes, and explore more comprehensive methods of analysing and understanding the G x E interactions of sugarcane, in order to optimize future MET networks. Data from plant breeding selection trials and post-release evaluation trials were systematically analysed using various statistical approaches combined with the use of soil and climatic data and crop models. Statistical methods based on multivariate methodologies such as GGE biplot, AMMI, and pattern analysis, were used to explore the effects of different environmental factors on sugarcane performance and agronomic traits.

The age at harvest was the main factor causing different genotypic responses in the midlands region, which was unique in comparison with the coastal and hinterland regions that shared similar characteristics. In the midlands region, two testing sites were identified as being redundant and were recommended for removal from the trial network in favour of a testing site in a frost pocket. In the coastal/hinterland region, three sites were identified as being redundant. Along the coast, time of harvest influenced G x E interactions, with yields showing stronger correlations to stalk population in the early season and stalk weight in the late season. In all regions, site x ratoon interactions accounted for more variation in yield than genotype x ratoon interactions, suggesting that variation in ratoon performance is influenced more by site differences than genotype differences. The repeatable component of G x E interaction (genotype x site) was larger than the non-repeatable components (genotype x ratoon and genotype x site x ratoon) on the long cycle program on the coast, however, this was not the case on the short cycle. This suggests that more effort be placed on identifying more diverse test sites on the short cycle.

In addition to providing direct recommendations for the industry selection programs, this study also illustrated novel methods of understanding sugarcane growth in different environments. The benefits of using a crop growth model to characterize sugarcane METs for water stress were illustrated throughout the study. The further use of the crop model to establish sugarcane growth phases also proved useful, and is likely to be more valuable when diverse datasets are analysed. Trials were characterized in terms of basic climatic and soil variables, which proved to be invaluable in understanding the causes of G x E interactions.

The characterisation of the current sites ensures that future site selection will be more rigorous, as plant breeders will be more aware of the conditions to select for or against.

The study showed that the integration of empirical and analytical statistical approaches was more valuable than using either approach in isolation, as is conventionally done in sugarcane. Additionally, these techniques were applied across many trial series and shown to produce repeatable results. The different strategies used to investigate sugarcane trait relations in this study have not been reported elsewhere, and future sugarcane studies dealing with similar traits (or other traits associated with lower level plant processes) may benefit from applying these methodologies. Furthermore, the integration of these multivariate methods with the largest ever simulation of sugarcane METs has opened new doors for the combined use of crop and statistical models in sugarcane research – an area not previously explored for this crop.   The study illustrated novel methods of identifying factors responsible for sugarcane G x E interactions and introduced new ways of characterizing sugarcane METs through the use of crop growth models and supplementary environmental data.