## Universidade de São Paulo Escola Superior de Agricultura "Luiz de Queiroz"

Comparação do GGE biplot-ponderado e AMMI-ponderado com out	ros
modelos de interação genótipo × ambiente	

## **Kuang Hongyu**

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Estatística e Experimentação Agronômica

## Kuang Hongyu Bacharel em Matemática

Comparação do GGE biplot-ponderado e	AMMI-ponderado com outros modelos de
interação genót	tipo × ambiente
	Orientador: Prof. Dr. CARLOS TADEU DOS SANTOS DIAS

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Estatística e

Experimentação Agronômica

## Dados Internacionais de Catalogação na Publicação DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Hongyu, Kuang

Comparação do GGE biplot-ponderado e AMMI-ponderado com outros modelos de interação genótipo x ambiente / Kuang Hongyu. - - Piracicaba, 2015. 154 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Interação genótipos x ambientes 2. Análise de mega-ambiente 3. Modelo AMMI 4. Biplot 5. GGE Biplot 6. W-AMMI 7. W-GGE Biplot I. Título

CDD 630.2195 H773c

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

## **DEDICATÓRIA**

Aos meus pais,

Kuang Songzhi e Zhen Yaci, pelo exemplo de vida, por sempre acreditarem em mim, por todo apoio e confiança em mim depositada, pois sem a participação de vocês não teria chegado até aqui.

A minha esposa,

Fabiane de Lima Silva, por toda ajuda, incentivo, ensinamentos, conselhos, companheirismo, o amor sempre constante e por me fazer tão feliz... Te amo!.

A minha irmã,

Kuang Qixia, pelo amizade, incentivo, apoio e por fazer parte da minha vida.

#### **AGRADECIMENTOS**

A Deus por sempre iluminar o meu caminho, para que eu pudesse prosseguir os meus estudos com sabedoria e perseverança.

Ao meu orientador, Prof. Dr. Carlos Tadeu dos Santos dias pela orientação, oportunidades, compreensão e amizade em todos esses anos de convivência. Agradeço pela paciência, exemplo, todos os ensinamentos e principalmente pela confiança em mim depositada desde o ano de 2010. Muito obrigado.

Ao meu co-orientador tanto no mestrado quanto no meu doutorado, Prof. Dr. Lúcio Borge de Araújo, pela confiança, amizade, ensinamentos e apoio na elaboração da tese.

Aos professores da banca de qualificação, Prof. Dr. Lúcio Borge de Araújo, Prof. Dra. Sônia Maria de Stefano Piedade e Prof. Dr. Antônio Carlos Simões Pião pelas sugestões e contribuições no desenvolvimento do trabalho.

Aos amigos Neuber e Anderson pela contribuição e ajuda na tese.

Aos professores do Departamento de Matemática da Universidade Federal de Uberlândia por todos os ensinamentos e amizade.

Aos professores do curso de Pós-graduação em Estatística e Experimentação Agronômica pelos ensinamento que contribuíram para minha formação acadêmica.

Aos caros colegas do Departamento de Estatística da Universidade Federal de Mato Grosso (UFMT) pela força, compreensão e incentivos diários para a conclusão desta tese na escaldante e distante Cuiabá.

Aos amigos Ana Patrícia Pexoto, Anderson Castro Soares de Oliveira, Elizabeth Mie Hashimoto, Gilmar Jorge de Oliveira Junior, Gláucia Ferrari, José Nilton Cruz, Juliano Bortolini, Luiz André Ribeiro Zardo, Marcelino Pascoa, Marisol Peña, Renato Pereira, Neuber José Segri, Ricardo Olinda, Sérgio Alarcón e Tiago Almeida. Agradeço a vocês pela amizade fraterna, conversas, incentivo e pelos incontáveis momentos alegres compartilhados durante todos esses anos!

A todos os alunos do curso de Pós-Graduação em Estatística e Experimentação Agronômica da ESALQ/USP, com os quais tive o prazer de conviver.

À secretária do LCE, Luciane Brajão, pelo apoio, amizade, ajuda e confiança depositada em mim, que me ajudou muito durante esses dois anos de mestrado e três anos de doutorado.

À secretária da Pós-Graduação em Estatística e Experimentação Agronômica, Solange Paes de Assis Sabadin, pelo apoio, amizade, ajuda sempre que preciso, que me ajudou muito durante esses dois anos de mestrado e três anos de doutorado.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP Eduardo Bonilha, Jorge Alexandre Wiendl, Rosni Honofre Aparecido Pinto e Mayara Segatto pela amizade, ajuda e apoio técnico.

Às bibliotecárias da ESALQ/USP, Eliana Maria Garcia e Maria da Glória Eloi da Silva pela ajuda e revisão da tese.

A CAPES e CNPq pela concessão de bolsa de estudos.

As pessoas que contribuíram de alguma forma ou com um simples gesto para a realização do presente trabalho, pois na vida não conquistamos nada sozinhos e sempre precisamos do auxílio de outras pessoas para alcançar os nossos objetivos.

Em especial, à minha esposa Fabiane. Agradeço pelo incentivo, apoio nos momentos em que mais precisei e por ser minha inspiração de cada dia. Obrigado por fazer parte da minha vida e estar do meu lado na realização desse sonho... Te amo!

"Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá.

Ayrton Senna

"Se você acreditar que uma coisa é impossível, você a tornará impossível. Lembre-se, o sucesso é uma jornada, não um destino. Tenha fé em sua capacidade.

Bruce Lee

"Não deixe o barulho da opinião dos outros abafar sua voz interior.

E mais importante, tenha a coragem de seguir seu coração e sua intuição.

Eles de alguma forma já sabem o que você realmente quer se tornar.

Tudo o mais é secundário."

Steve Jobs

## SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	19
$1 INTRODUÇ\~AO \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	23
1.1 Objetivos	25
2 REVISÃO BIBLIOGRÁFICA	27
2.1 Interação genótipo $\times$ ambiente	27
2.1.1 Classificação da GEI	29
2.2 Análise de variância	30
2.3 Os modelos estatísticos com base em regressão e decomposição por valores singulares	33
2.4 Decomposição do valores singulares	34
2.5 Análise de Componentes Principais	38
2.6 Histórico dos modelos AMMI	42
2.7 Histórico Biplot	43
2.8 Tipos de Biplot	46
2.9 Teoria e conceito do Biplot AMMI	49
3 MATERIAL E MÉTODOS	51
3.1 Características dos dados	51
3.2 Análise de variância conjunta	51
3.3 Modelo AMMI	53
3.4 Análise AMMI ponderado (W-AMMI)	58
3.5 Modelo SREG	59
3.6 GGE biplot	60
3.7 DVS em GGE biplot	63
3.8 W-GGE Biplot (GGE biplot ponderado)	64
4 RESULTADOS E DISCUSSÃO	67
$4.1$ Análise AMMI e Biplot para $1^o$ conjunto dados multi-ambientes (MET)	67
4.1.1 Análise de variância conjunta para dois fatores	67
$4.2\mathrm{GGE}$ biplot para análise de $1^o$ conjunto de dados MET	74
4.2.1 Análise Mega-ambiente (1º conjunto de dados)	77
4.2.2 Avaliação de genótipos baseada em GGE biplot (1º conjunto de dados)	79

4.2.3 Avaliação do ambiente de teste a teste-ambiente (1º conjunto de dados) 81
$4.3$ Análise AMMI-ponderado e Biplot para $1^o$ conjunto dados multi-ambientes (MET) $83$
$4.4\mathrm{GGE}$ biplot ponderado (W-GGE biplot) para análise de $1^o$ conjunto dados MET . $~89$
4.4.1 Análise Mega-ambiente com W-GGE biplot (1° conjunto de dados) 91
4.4.2 Avaliação de genótipos baseado em W-GGE biplot (1º conjunto de dados) $$ 93
4.4.3 Avaliação do Ambiente de Teste baseado no W-GGE biplot (1º conjunto de dados) $95$
4.5 Comparação das análises (AMMI vs AMMI-ponderado vs GGE biplot vs W-GGE
biplot) para 1º conjunto de dados MET
4.5.1 Modelo AMMI versus GGEbiplot
4.5.2 Comparação dos modelos para o 1º conjunto de dados MET
$4.6$ Análise AMMI e Biplot para $2^o$ conjunto de dados MET $\ \ldots \ \ldots \ \ldots \ \ldots \ 101$
4.6.1 Análise de variância conjunta para dois fatores
$4.7\mathrm{GGE}$ biplot para análise de $2^o$ conjunto dados MET
4.7.1 Análise Mega-ambiente (2º conjunto de dados)
4.7.2 Avaliação de genótipos baseada em GGE biplot (2º conjunto de dados) 111
4.7.3 Avaliação do Ambiente de Teste (2º conjunto de dados)
$4.8$ Análise AMMI-ponderado e Biplot para $2^o$ conjunto dados multi-ambientes (MET) $116$
$4.9\mathrm{GGE}$ biplot ponderado (W-GGE biplot) para análise de $2^o$ conjunto dados MET . $120$
4.9.1 Análise Mega-ambiente com W-GGE biplot (2º conjunto de dados) 122
4.9.2 Avaliação de genótipos baseado em W-GGE biplot (2º conjunto de dados) $$ 124
$4.9.3$ Avaliação do Ambiente de Teste baseado no W-GGE biplot (2º conjunto de dados) 126 $^{\circ}$
$4.10\mathrm{Compara}$ ção das análises (AMMI v s AMMI-ponderado v s GGE biplot v s W-GGE
biplot) para 2º conjunto de dados MET
5 CONCLUSÕES
5.1 Pesquisas futuras
REFERÊNCIAS
APÊNDICES

#### **RESUMO**

# Comparação do GGE biplot-ponderado e AMMI-ponderado com outros modelos de interação genótipo $\times$ ambiente

Interação genótipo × ambiente (GEI) é uma questão extremamente importante no melhoramento genético de plantas e produção. A seleção e recomendação de genótipos superiores são dificultadas devido à ocorrência de GEI e representa um grande desafio para os pesquisadores. Nesse contexto, as análises biplot têm sido cada vez mais utilizadas na análise de dados agronômicos, em que os dados são representados por uma tabela de dupla entradas de médias de GEI. Entretanto, as particularidades existentes no gráfico biplot dificultam sua interpretação, podendo induzir o pesquisador a erros. Existem vários modelos na literatura para análise de DGE (dados de GEI), entre eles, os mais utilizados são os modelos AMMI (Additive Main effects and Multiplicative Interaction) e GGE biplot (Genotype main effects + Genotype environment interaction). O modelo AMMI é um método estatístico para compreender a estrutura de interações entre genótipos e ambientes, que combina a análise de variância e a análise de componentes principais, para ajustar, respectivamente, os efeitos principais (G e E) e os efeitos da GEI. O GGE Biplot agrupa o efeito aditivo de genótipo com o efeito multiplicativo da GEI, e submete estes à análise de componentes principais. Existem dois problemas na utilização destes modelos: i) só pode ser utilizado para analisar dados MET (multi-ambientes), que tenha uma única característica e ii) cujos ambientes são heterogêneos. O presente trabalho tem como objetivos propor novos modelos W-GGE biplot (Weighted Genotype main effects + Genotype environment interaction) e AMMI-ponderado para análise de dados multi-ambientes, além de fazer uma comparação entre os modelos existentes como AMMI e GGE biplot; análise de mega-ambiente; avaliação de genótipos, ambiente de teste dentro de cada mega-ambiente e compreender as causas da GEI.

Palavras-chave: Interação genótipos × ambientes; análise de mega-ambiente; Modelo AMMI; Biplot; GGE Biplot; W-AMMI; W-GGE Biplot

#### ABSTRACT

# Comparison of weighted-GGE biplot and weighted-AMMI with other models of interaction genotype $\times$ environment

Genotype × environment interaction (GEI) is an extremely important issue in plant breeding and production. The selection and recommendation of superior genotypes are hampered due to the occurrence of GEI and represents a major challenge for researchers. In this context, biplot analyzes have been increasingly used in analyzing agronomic data, in which data are represented by a table of two entries of means of GEI. However, the particularities in the biplot graphic hamper its interpretation, and could lead the researcher to errors. There are several models in the literature for DGE analysis (GEI data), among them, the most used are the AMMI model (Additive Main effects and Multiplicative Interaction Models) and GGE biplot (Genotype main effects + Genotype environment interaction). The AMMI model is a statistical method to understand the structure of interactions between genotypes and environments, combining the analysis of variance and principal component analysis, to adjust, respectively, the main effects (G and E) and the effects of GEI. The GGE Biplot groups genotype of additive effect with multiplicative effect of GEI, and submit these to the principal component analysis. There are two problems in using these models: i) can only be used to analyze MET data (multi-environments), which has a unique feature and ii) whose environments are heterogeneous. This paper aims to propose new W-GGE biplot models (Weighted Genotype main effects + Genotype environment interaction) and AMMI-weighted multi-environments for data analysis, and make a comparison between the existing models as AMMI and GGE biplot; mega-environment analysis; genotype evaluation, test environment within each mega-environment and understand the causes of GEI.

Keywords: Genotypes × environments; mega-environment analysis; AMMI model; biplot; GGE Biplot; W-AMMI; W-GGE Biplot

## LISTA DE FIGURAS

Figura 1 -	Interações Genotipos × Ambientes (GEI) (SQUILASSI, 2003)	29
Figura 2 -	Biplot AMMI1 (Médias vs PC1) para os dados de produtividade de milho	
	$(kg\ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)	72
Figura 3 -	Biplot AMMI2 (PC1 vs PC2) para os dados de produtividade de milho	
	$(kg\ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)	73
Figura 4 -	O GGE Biplot para os dados de produtividade de milho $(kg\ ha^{-1})$ , com	
	nove genótipos (G) e vinte ambientes (E)	76
Figura 5 -	O GGE Biplot ("Which-won-where") mostra que para os dados de pro-	
	dutividade de milho $(kg\ ha^{-1})$ , quais dos genótipos apresentaram melhor	
	desempenho e em qual ambientes	78
Figura 6 -	O GGE Biplot ("Média versus Estabilidade") com eixo do ambiente-média	
	(EAM) para mostrar o desempenho médio e estabilidade dos genótipos	
	para os dados de produtividade de milho $(kg\ ha^{-1})$	79
Figura 7 -	O GGE Biplot com eixo do ambiente-média (EAM) para classificar	
	os genótipos em relação ao genótipo ideal (no centro dos círculos	
	concêntricos)	80
Figura 8 -	O GGE biplot "discriminação e representatividade" para mostrar a ca-	
	pacidade discriminação e de representatividade dos ambientes de teste $$ . $$ .	82
Figura 9 -	O GGE biplot discriminação e representatividade para classificar ambi-	
	entes de teste em relação a um ambiente de teste ideal (representado por	
	centro dos círculos concêntricos)	83
Figura 10 -	-Biplot AMMI1-ponderado (Médias vs PC1) para os dados de produtivi-	
	dade de milho $(kg\ ha^{-1}),$ com 15 genótipos (G) e quatro ambientes (A) .	88
Figura 11 -	-Biplot AMMI2-ponderado (PC1 vs PC2) para os dados de produtividade	
	de milho ( $kg\ ha^{-1}$ ), com 15 genótipos (G) e quatro ambientes (A)	88
Figura 12 -	-O W-GGE Biplot para os dados de produtividade de milho $(kg\ ha^{-1})$ ,	
	com nove genótipos (G) e vinte ambientes (E) $\dots \dots \dots \dots$	90
Figura 13 -	-O W-GGE Biplot ("Which-won-where") mostra que para os dados de pro-	
	dutividade de milho $(kg\ ha^{-1})$ , quais dos genótipos apresentaram melhor	
	desempenho em que ambientes	92

Figura 14 -	O W-GGE Biplot ("Média versus Estabilidade") com eixo do ambiente-
1	média (EAM) para mostrar o desempenho médio e estabilidade dos
	genótipos para os dados de produtividade de milho $(kg\ ha^{-1})$ 93
Figura 15 -	O W-GGE Biplot com eixo do ambiente-média (EAM) para classi-
i	ficar os genótipos em relação ao genótipo ideal (no centro dos círculos
•	concêntricos)
Figura 16 -	O W-GGE biplot "discriminação e representatividade" para mostrar a
•	capacidade discriminação e de representatividade dos ambientes de teste . 96
Figura 17 -	O W-GGE biplot discriminação e representatividade para classificar am-
1	bientes de teste em relação a um ambiente de teste ideal (representado por
•	centro dos círculos concêntricos)
Figura 18 -	Biplot AMMI1 (Médias vs PC1) para os dados de produtividade de milho
	$(kg\ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)
Figura 19 -	Biplot AMMI2 (PC1 vs PC2) para os dados de produtividade de milho
	$(kg\ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)
Figura 20 -	O GGE Biplot para os dados de produtividade de milho $(kg \ ha^{-1})$ , com
I	nove genótipos (G) e vinte ambientes (E)
Figura 21 -	O GGE Biplot ("Which-won-where") mostra que para os dados de pro-
	dutividade de milho $(kg\ ha^{-1})$ , quais dos genótipos apresentaram melhor
•	desempenho em que ambientes
Figura 22 -	O GGE Biplot ("Média versus Estabilidade") com eixo do ambiente-média
	(EAM) para mostrar o desempenho médio e estabilidade dos genótipos
]	para os dados de produtividade de milho $(kg\ ha^{-1})$
Figura 23 -	O GGE Biplot com eixo do ambiente-média (EAM) para classificar
•	os genótipos em relação ao genótipo ideal (no centro dos círculos
•	concêntricos)
Figura 24 -	O GGE biplot "discriminação e representatividade" para mostrar a capaci-
•	dade discriminação e de representatividade dos ambientes de teste 114
Figura 25 -	O GGE biplot discriminação e representatividade para classificar ambi-
•	entes de teste em relação a um ambiente de teste ideal (representado por
•	centro dos círculos concêntricos)
Figura 26 -	Biplot AMMI1-ponderado (Médias vs PC1) para os dados de produtivi-
	dade de milho $(kq \ ha^{-1})$ , com 15 genótipos (G) e quatro ambientes (A) . 119

Figura 27 -	-Biplot AMMI2-ponderado (PC1 vs PC2) para os dados de produtividade	
	de milho $(kg\ ha^{-1})$ , com 15 genótipos (G) e quatro ambientes (A)	120
Figura 28 -	-O W-GGE Biplot para os dados de produtividade de milho $(kg\ ha^{-1})$ ,	
	com nove genótipos (G) e vinte ambientes (E)	122
Figura 29 -	-O W-GGE Biplot ("Which-won-where") mostra que para os dados de pro-	
	dutividade de milho $(kg\ ha^{-1})$ , quais dos genótipos apresentaram melhor	
	desempenho em que ambientes	123
Figura 30 -	-O W-GGE Biplot ("Média versus Estabilidade") com eixo do ambiente-	
	média (EAM) para mostrar o desempenho médio e estabilidade dos	
	genótipos para os dados de produtividade de milho $(kg\ ha^{-1})$	124
Figura 31 -	-O W-GGE Biplot com eixo do ambiente-média (EAM) para classi-	
	ficar os genótipos em relação ao genótipo ideal (no centro dos círculos	
	concêntricos)	125
Figura 32 -	-O W-GGE biplot "discriminação e representatividade" para mostrar a ca-	
	pacidade discriminação e de representatividade dos ambientes de teste	126
Figura 33 -	-O W-GGE biplot discriminação e representatividade para classificar am-	
	bientes de teste em relação a um ambiente de teste ideal (representado por	
	centro dos círculos concêntricos)	127

## LISTA DE TABELAS

Tabela 1 -	Esquema da análise de variância para um experimento inteiramente	
	aleatorizado	32
Tabela 2 -	Esquema da análise de variância para experimentos de um mesmo grupo	
	de genótipos avaliado em $e$ locais com $r$ blocos	52
Tabela 3 -	Esquema da análise de variância para experimentos de um mesmo grupo	
	de genótipos avaliado em $e$ locais com $r$ repetições	53
Tabela 4 -	Análise de variância conjunta para um conjunto de dados com 15 genótipos	
	avaliados em 4 ambientes com 2 blocos	67
Tabela 5 -	Médias dos genótipos, ambientes e posição das médias em relação a pro-	
	dutividade	68
Tabela 6 -	Valores estimados da interação dupla de 15 genótipos e 4 ambientes para	
	a produção em $kg\ ha^{-1}$	69
Tabela 7 -	Porcentagem da soma de quadrados da GEI captada por componente prin-	
	cipal (PC)	71
Tabela 8 -	Coordenadas dos eixos das abscissas $(x)$ e de coordenadas $(y)$ , para os dois	
	gráficos Biplot	71
Tabela 9 -	Médias dos 15 genótipos (G1 a G15) testadas em quatro ambientes em	
	relação a produtividade	75
Tabela 10	-Valor singular, proporção explicada e relação da informação (IR) dos qua-	
	tro componentes principais (PCs)	75
Tabela 11	-Médias dos genótipos, ambientes e posição das médias em relação a pro-	
	dutividade	84
Tabela 12	-Produtividade média com os valores centrada nos ambientes de testes com	
	15 genótipos (G1 a G15) em quatro ambientes (A1 a A4) $\ \ldots \ \ldots \ \ldots$	85
Tabela 13	-Coordenadas dos eixos das abscissas $(x)$ e de coordenadas $(y)$ , para os dois	
	gráficos Biplot	87
Tabela 14	-Produção de dados de ambiente-centrados com os 15 genótipos (G1 a G15)	
	testadas em quatro ambientes	89
Tabela 15	-Valor singular, proporção explicada e relação da informação (IR) dos qua-	
	tro componentes principais (PCs)	90
Tabela 16	-Valores singulares dos quatro componentes principais (PCs) para os mod-	
	elos AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot	99

Tabela 17 - A matriz de correlação dos valores singulares de PCs dos modelos AMMI,
AMMI-ponderado, GGE biplot e W-GGE biplot
Tabela 18 - Proporções explicadas e IR dos componentes principais dos modelos
AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot 100
Tabela 19 - Análise de variância conjunta para um conjunto de dados com 9 genótipos
avaliados em 7 ambientes com 4 blocos
Tabela 20 - Médias dos genótipos, ambientes e posição das médias em relação a pro-
dutividade
Tabela 21 - Valores estimados da interação dupla de 9 genótipos e 7 ambientes para a
produção em $kg\ ha^{-1}$
Tabela 22 - Porcentagem do desdobramento da soma de quadrados da interação (G $\times$
E) captada por componente principal (PC)
Tabela 23 - Coordenadas dos eixos das abscissas $(x)$ e de coordenadas $(y)$ , para os dois
gráficos Biplot
Tabela 24 - Médias dos nove genótipos (G1 a G9) testadas em sete ambientes (A1 a
A7) em relação a produtividade média $(kg\ ha^{-1})$ 108
Tabela 25 - Valor singular, proporção explicada e relação da informação (IR) dos qua-
tro componentes principais (PCs)
Tabela 26 - Médias dos genótipos, ambientes e posição das médias em relação a pro-
dutividade (AMMI-ponderado)
Tabela 27 - Produtividade média com os valores centrados nos ambientes de testes
com nove genótipos (G1 a G9) em sete ambientes (A1 a A7) pela análise
AMMI-ponderado
Tabela 28 - Coordenadas dos eixos das abscissas $(x)$ e de coordenadas $(y)$ , para os dois
gráficos Biplot
Tabela 29 - Médias dos nove genótipos (G1 a G9) testadas em sete ambientes (A1 a
A7) em relação a produtividade médio $(kg\ ha^{-1})$
Tabela 30 -Valor singular, proporção explicada e relação da informação (IR) dos qua-
tro componentes principais (PCs)
Tabela 31 - Valores singulares das sete componentes principais (PCs) para os modelos
AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot
Tabela 32 - A matriz de correlação dos valores singulares de PCs dos modelos AMMI,
AMMI-ponderado, GGE biplot e W-GGE biplot

Tabela 33 - Proporções explicada e IR dos componentes principais dos modelos AMMI,	
AMMI-ponderado, GGE biplot e W-GGE biplot	129

## 1 INTRODUÇÃO

A resposta diferencial de genótipos em ambientes é frequente em estudos de experimentos multi-ambientais (MET) e é conhecido como interação genótipo × ambiente (GEI). Dados de METs são conduzidos de vários anos para os principais produtos agrícolas no mundo, frequentemente resumidos em tabelas de duas entradas com genótipos nas linhas e ambientes (ou locais) nas colunas. Os METs são essenciais porque a presença da GEI, ou seja, a mudança na performance relativa de genótipos por meio de diferentes ambientes, complica a avaliação de genótipo. Se não existisse a GEI, um único cultivar prevaleceria no mundo inteiro e um único experimento bastaria para avaliação de genótipo (GAUCH; ZOBEL, 1996; GAUCH, 2013; RODRIGUES et al., 2014). A GEI ocorre em várias maneiras, com a maneira mais extrema que consiste em interações cruzadas, na classificação de genótipos com mudanças entre ambientes, por exemplo, um genótipo responde de maneira diferente às variações ambientais e nesta situação, os melhores genótipos em um local, podem não o ser, necessariamente os melhores em outros locais.

Enquanto o cruzamento de GEI induz uma necessidade de adaptação específica e requer a existência de pelo menos dois mega-ambientes (GAUCH; ZOBEL, 1997; GAUCH, 2013), ausência de cruzamento de interação enfatiza a utilidade do efeito principal genotípico como uma medida de desempenho e pontos para adaptação ampla. Uma vez que a soma dos quadrados de interação  $(SQ_{GEI})$  é muitas vezes maior do que a soma dos quadrados para os efeitos principais genotípicos, o estudo e compreensão de adaptações específicas oferecem uma oportunidade importante para melhorar traços fenotípicos em relação a gradientes ambientais.

No melhoramento de plantas, os principais objetivos de ensaios multiambiente (METs) são: (i) estudo de GEI, (ii) avaliar a adaptabilidade e estabilidade
genotípica, (iii) estabelecer relações entre os testes de ambientes, entre os genótipos, entre genótipos e ambientes (ou locais) simultaneamente e (iv) fazer predições do valor da
reprodução de genótipos que vai permitir fazer uma seleção precisa dos pais para o próximo
ciclo de reprodução. A presença da GEI complica este processo e normalmente é expresso
como respostas inconsistentes de alguns genótipos em relação a outros, devido à mudança
de classificação genotípica, ou como mudanças nas diferenças absolutas entre genótipos sem
mudança no classificação (CROSSA; YANG; CORNELIUS, 2004).

Modelos linear-bilineares (também chamados de multiplicativo) para o estudo de interações de dados de duas entradas foi estudado por Mandel (1961, 1969). Modelos de

efeito fixo linear-bilineares, tais como modelo de SREG (Sites Regression) (CORNELIUS; CROSSA; SEYEDSADR, 1996; CROSSA; CORNELIUS, 1997), o modelo AMMI (Additive Main Effects and Multiplicative Interaction Model) (GAUCH, 1992; GAUCH; ZOBEL, 1996; GAUCH, 2013), o GGEbiplot (Genotype main effects + Genotype environment interaction) (YAN; KANG, 2003; YAN; HOLLAND, 2010) e W-AMMI (Weighted Additive Main Effects and Multiplicative Interactions) (RODRIGUES, 2012; RODRIGUES et al., 2014), são utilizados para estudar os padrões genotípicos das respostas por meio de ambientes. Estes modelos são parcimoniosos, uma vez que os parâmetros de interação são estimados a partir da decomposição por valor singular (DVS) da matriz de GE (AMMI) ou a partir da matriz (SREG) de genótipo (G) mais GEI (GGE), e os padrões de resposta de genótipos e ambientes pode ser visualizados graficamente usando Biplots (GABRIEL, 1971; KEMPTON, 1984). O W-AMMI é uma generalização do modelo AMMI que permite ter em consideração a heterogeneidade da variância do erro (error variance) ao longo dos ambientes (RODRIGUES, 2012; RODRIGUES, et al., 2014).

Para os melhoristas, modelos linear-bilineares, tais como os efeitos fixos SREG e/ou AMMI podem oferecer mais oportunidades para a modelagem GGE ou GE do que a regressão simples dos genótipos sobre a média local que foi anteriormente sugerido por Finlay e Wilkinson (1963) e Eberhart e Russell (1966). Várias revisões recentes têm estudados os modelos AMMI e GGE biplot para a análise GEI (GAUCH, 2006; DIAS; KRZANOWSKI, 2006; Yan et al., 2007; GAUCH; PIEPHO; ANNICCHIARICO, 2008; YAN; HOLLAND, 2010; RODRIGUES, et al., 2014; HONGYU, et al., 2014). Denis e Gower (1994, 1996) e Denis e Pazman (1999) propuseram regiões de confiança assintóticas para os escores genotípicos e ambientais que ajudam os melhoristas a tomar decisões mais confiáveis sobre a seleção de genótipos e recomendação, no entanto, essas regiões de confiança não são facilmente implementadas para modelos com mais de dois termos bilineares, e eles exigem suposições restritivas, como a normalidade assintótica. Yang et al. (2009) defenderam o uso de uma técnica de reamostragem não paramétrica (bootstrapping) para a construção de regiões de confiança para os escores genotípicos e ambientais que podem ser aplicadas para modelos linear-bilineares de efeito fixo e misto.

A única diferença entre as metodologias GGE e AMMI está na etapa inicial da análise, o modelo GGE analisa diretamente o efeito de G+GEI, enquanto que a AMMI separa G da GEI e na etapa final da análise com formação dos gráficos biplot para as duas metodologias. No entanto, essa separação não é capaz de conferir superioridade à análise

AMMI (YAN; KANG, 2003; YAN et al., 2007). Comparando as metodologias AMMI e GGE Biplot na formação de mega-ambientes, CAMARGO-BUITRAGO et al. (2011) relataram 95,2% de coincidência nos resultados entre as metodologias. Assim, o modelo GGE biplot é mais indicado para a identificação de mega-ambientes, seleção de ambientes representativos e discriminativos e indicação de cultivares mais adaptadas e estáveis a ambientes específicos (YAN et al., 2007; GAUCH et al., 2008; ASFAW et al., 2009; YAN, 2011). Por outro lado, a análise AMMI pode ser utilizada com eficiência na identificação de condições ambientais superiores para a exploração agrícola (seleção de locais de cultivo) e genótipos de superior desempenho médio (GAUCH et al., 2008; YAN, 2011). Vários estudos recentes (YAN; HOLLAND, 2010; CAMARGO-BUITRAGO et al., 2011; YAN, 2011; GAUCH, 2013; RO-DRIGUES, et al., 2014; HONGYU, et al., 2014) demonstraram que ambas as metodologias, quando utilizadas adequadamente, são eficientes na definição de mega-ambientes.

#### 1.1 Objetivos

Os objetivos desta pesquisa são 1) investigar o mega-ambiente para a compreensão do ambiente alvo; 2) avaliar os genótipos dentro de cada mega-ambiente; 3) avaliar o ambiente de teste dentro de cada mega-ambiente; 4) compreender as causas da interação genótipo × ambiente (GE) e 5) propor novos modelos W-GGE biplot e AMMI-ponderado para análise de dados MET e além de fazer uma comparação entre os modelos existentes como AMMI (Additive Main effects and Multiplicative Interaction) (GAUCH, 1992; GAUCH; ZOBEL, 1996) e GGE biplot (Genotype main effects + Genotype environment interaction) (YAN; KANG, 2003).

### 2 REVISÃO BIBLIOGRÁFICA

#### 2.1 Interação genótipo × ambiente

A interação genótipo × ambiente (GEI) foi definida por Shelbourne (1972) como sendo a variação entre genótipos em resposta a diferentes condições ambientais. Neste sentido, na presença da interação, os resultados das avaliações podem mudar de um ambiente para outro, ocasionando mudanças na posição relativa dos genótipos ou mesmo na magnitude das suas diferenças (FALCONER; MACKAY, 1996). A GEI deve ser encarada como um fenômeno biológico com suas implicações no melhoramento de plantas, e não como um fenômeno de simples efeito estatístico, cumprindo buscar a explicação evolutiva do evento se quiser tirar proveito de seus efeitos benéficos ou indesejáveis sobre a avaliação de genótipos e recomendação de cultivares (CHAVES, 2001).

A seleção e recomendação de genótipos mais produtivos é objetivo básico dos programas de melhoramento genético de qualquer espécie cultivada (CARVALHO et al., 2002). O processo de seleção é frequentemente realizado avaliando-se o desempenho dos genótipos em diferentes ambientes (ano, local, época de semeadura). Contudo, a decisão de lançamento de novas cultivares normalmente é dificultado pela ocorrência da GEI.

A GEI ocorre quando há respostas diferenciadas dos genótipos testados em diferentes ambientes (CARVALHO et al., 2002). Ela pode ser simples, quando é proporcionada pela diferença de variabilidade entre genótipos nos ambientes, e complexa, quando denota a falta de correlação entre medidas de um mesmo genótipo em ambientes distintos e indica haver inconsistência na superioridade de genótipos com a variação ambiental (ROBERTSON, 1959). Somente quando ocorre a interação complexa haverá dificuldades no melhoramento (CRUZ; REGAZZI, 1997).

As características relacionadas à produção vegetal estão condicionadas ao controle genético do organismo, ao ambiente em que é cultivado e à interação entre esses dois fatores. As diferentes respostas fenotípicas frente a mudanças nas condições ambientais resultam em comportamentos distintos dos genótipos, caracterizando a interação. Se a expressão de um determinado genótipo depende dos genes e do ambiente em que é avaliado, a GEI deve ser mais um fator a ser considerado na análise. Assim, no melhoramento, o processo de seleção depende também da estimação dessa interação, para que não ocorra a queda inesperada de desempenho de um material testado.

Em um determinado ambiente, a manifestação fenotípica (P) é o resultado

da ação do genótipo (G) sob influência do meio (E). Entretanto, quando se considerar uma série de ambiente, detecta-se, além dos efeitos genéticos e ambientais, um efeito adicional, proporcionado pela interação destes (GE) (CRUZ e REGAZZI, 1997). Pode-se escreve:

- i) P = G + E; quando da ausência de interação.
- ii) P = G + E + GE; quando da presença de GEI.

A avaliação da GEI torna-se de grande importância no melhoramento, pois, no caso de sua existência, há possibilidades de o melhor genótipo em um ambiente não o ser em outro. Este fato influencia o ganho de seleção e dificulta a recomendação de cultivares com ampla adaptabilidade. Pela importância desta interação, cabe ao melhorista avaliar sua magnitude e tecnologia e fornecer subsídio que possibilitem adotar procedimentos para sua minimização e, ou, seu aproveitamento.

O potencial genético de determinado cultivar depende das condições ambientais o qual está sujeito, tais como, condições climáticas, tipo de solo, mudanças e melhorias no ambiente. A GEI reflete as diferentes respostas dos genótipos às condições ambientais, sendo um dos aspectos mais importantes na seleção e recomendação dos melhores genótipos com alta produtividade e qualidade pelos melhoristas (FALCONER; MACKAY, 1996; WARZECHA et al., 2011; GAUCH, 2013).

No melhoramento genético de plantas, os ensaios multiambientais são importantes para testar a adaptação geral e específica das cultivares. Uma cultivar desenvolvendose em diferentes ambientes mostrará uma flutuação significativa na produtividade em relação a outras cultivares. Essas mudanças são influenciadas por diferentes condições ambientais e são referidas como interação  $(G \times E)$  (DIAS; KRZANOWSKI, 2006; ARCINIEGAS-ALARCÓN; DIAS, 2009).

Cockerham (1963) atribuiu o aparecimento de GEI como sendo devido a respostas diferenciais do mesmo conjunto gênico em ambientes distintos ou pela expressão de diferentes conjuntos gênicos em diferentes ambientes. Quando um mesmo conjunto de genes se expressa em diferentes ambientes, as diferenças nas respostas podem ser explicadas pela heterogeneidade das variâncias genéticas e experimentais ou por ambas e quando diferentes conjuntos de genes se expressam em ambientes distintos, as diferenças nas respostas explicam-se pela inconsistência das correlações genéticas entre os valores de um mesmo caráter em dois ambientes (FALCONER, 1989). A GEI também pode surgir em função de fatores fisiológicos e bioquímicos próprios de cada genótipo cultivado (CRUZ; REGAZZI,

1997). Ainda que a falta de ajuste do modelo estatístico adotado ao conjunto de dados pode ser uma das causas da GEI significativa (CHAVES; VENCOVSKY; GERALDI, 1989).

A GEI pode ser reduzida, utilizando-se cultivares com ampla adaptabilidade e boa estabilidade, ou estratificando-se a região considerada em sub-regiões com características ambientais semelhantes, dentro das quais a interação passa a ser não-significativa (RAMALHO; SANTOS; ZIMMERMANN, 1993).

#### 2.1.1 Classificação da GEI

Mazer e Schick (1991) explicaram várias magnitudes da interação, conforme Figura 1, em que observa-se que populações distintas (A a E) são submetidas a dois ambientes distintos (1 e 2) para a manifestação de uma variável métrica qualquer (SQUILASSI, 2003).

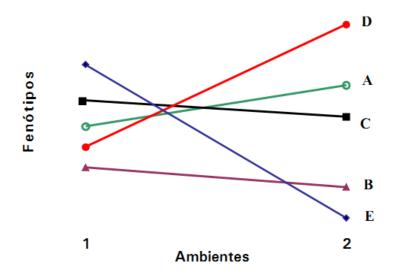


Figura 1 - Interações Genótipos × Ambientes (GEI) (SQUILASSI, 2003)

Nota-se que a relação entre fenótipo e genótipo pode ser alterada de acordo com o ambiente, de dois modos principais, que não são mutuamente excludentes, a saber: i) a magnitude da variação (diferença) entre os genótipos e ii) o ranking dos fenótipos devido aos diferentes genótipos. Pela figura 1, podem ser observadas três situações:

a) Ausência de interação, pois a mudança nas condições ambientais não alterou o comportamento dos genótipos (B × C), afetando igualmente o comportamento destes.

- b) Nas duas outras situações, a alteração no ambiente resultou em resposta diferenciada dos genótipos, evidenciando a ocorrência de GEI, que pode ser classificada como:
  - b.1) Interação Simples ou Quantitativa, onde há mudança na magnitude de performance dos genótipos ( $A \times B$  ou  $B \times D$ ), mas o ordenamento destes nos diferentes ambientes permanece inalterado, ou seja, um genótipo que era superior no ambiente 1 permanece superior no ambiente 2. Este tipo de interação indica que as populações são geneticamente heterogêneas e os ambientes homogêneos, ou vice-versa.
  - b.2) Interação Cruzada ou Qualitativa, na qual há uma resposta diferenciada das populações (ou genótipos) a diferentes ambientes, havendo uma mudança na classificação destas ( $C \times D$ ,  $A \times C$  ou  $B \times E$ ).

As situações a e b.1 indicam que uma população (ou genótipo) é melhor adaptada aos dois ambientes de teste. Assim, uma seleção baseada na média de ambientes beneficiará sempre o melhor genótipo. Na situação b.1, a manifestação da interação é consequência, apenas, da conceituação aditiva do fenômeno e poderia ser explicada, por exemplo, pela ação multiplicativa dos ambientes sobre os genótipos (CHAVES; VENCOV-SKY; GERALDI, 1989). Na situação b.2, utilizar a média dos ambientes como indicador de seleção não é suficiente e poderá induzir a erros, levando à escolha de genótipos mal adaptados a uma situação particular.

Quando se consideram diversos genótipos avaliados em diferentes ambientes, muitas combinações das situações acima discutidas ocorrem, o que torna a interpretação muito difícil. Daí a necessidade de se utilizar métodos e procedimentos adequados de análise. Para n genótipos e m ambientes, pode-se obter nm!/n!m! combinações diferentes (ALLARD; BRADSHAW, 1964), o que torna o trabalho praticamente inviável ou impraticável, mesmo para valores baixos de n e m (SQUILASSI, 2003).

#### 2.2 Análise de variância

A Análise de Variância (ANOVA) é um procedimento utilizado para verificar o efeito de dois ou mais níveis de tratamentos. Existem muitas variações da ANOVA devido aos diferentes tipos de experimentos que podem ser realizados.

Em um experimento, cada observação  $Y_{ij}$  pode ser decomposta conforme o modelo a seguir, para um experimento com delineamento inteiramente aleatorizado (FER-

REIRA, 2000):

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i = 1, \dots, I \quad e \quad j = 1, \dots, J \tag{1}$$

em que:  $Y_{ij}$  é a observação do i-ésimo tratamento na j-ésima repetição;  $\mu$  é o efeito constante (média geral);  $\tau_i$  é o efeito do i-ésimo tratamento;  $\varepsilon_{ij}$  é o erro associado ao i-ésimo tratamento na j-ésima repetição assumido como:  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2)$ . Aqui, IID significa que os erros devem ser independentes e identicamente distribuídos.

Em um experimento, existe o interesse em testar se há diferenças entre as médias dos tratamentos, o que equivale a testar as hipóteses:

$$\begin{cases} H_0: \tau_1 = \tau_2 = \dots = \tau_I = 0 \\ H_1: \tau_i \neq 0 \quad Pelo \ menos \ um \ contraste \ nos \ efeitos \ \tau_i \ \acute{e} \ significativo. \end{cases}$$

De forma equivalente, podemos escrever tais hipóteses da seguinte forma:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_I \\ H_1: \mu_i \neq \mu_i' \quad Pelo \ menos \ um \ contraste \ nas \ médias \ \mu_i \ \acute{e} \ significativo. \end{cases}$$

em que

$$\mu_i = \mu + \tau_i$$
  $i = 1, 2, \dots, I.$ 

Note que, se a hipótese nula for verdadeira, todos os tratamentos terão uma média comum  $\mu$ . A análise de variância, baseia-se na decomposição da variação total da variável resposta em partes que podem ser atribuídas aos tratamentos (variação entre) e ao erro experimental (variação dentro). Essas variações podem ser medidas por meio das somas de quadrados definidas para cada um dos seguintes componentes:

$$SQTotal = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}^{2} - C$$

em que

$$C = \frac{\sum\limits_{i=1}^{I} \sum\limits_{j=1}^{J} y_{ij}^2}{IJ}$$

$$SQTrat = \frac{\sum_{i=1}^{I} y_{i.}^{2}}{J} - C$$

em que  $y_{i.} = \sum_{j=1}^{J} y_{ij}$  o total do *i*-ésimo tratamento.

E a soma de quadrados dos resíduos pode ser obtida por diferença:

$$SQRes = SQTotal - SQTrat$$

.

A SQTrat também é chamada de "variação Entre", que é a variação existente entre os diferentes tratamentos e a SQRes é chamada de "variação Dentro" que é função das diferenças existentes entre as repetições de um mesmo tratamento.

Essas soma de quadrados podem ser organizadas em uma tabela, denominada tabela da análise de variância, como apresentado na Tabela 1.

Para testar a hipótese  $H_0$ , utiliza-se o teste F apresentado na tabela da Análise de Variância (Tabela 1). Convém lembrar que esse teste é válido se os pressupostos assumidos para os erros do modelo estiverem satisfeitos (FERREIRA, 2000).

Tabela 1 - Esquema da análise de variância para um experimento inteiramente aleatorizado

Fonte de	Graus de	Somas de	Quadrados	F calculado
Variação	Liberdade	Quadrados	Médios	
Tratamentos	I-1	SQTrat	QMTrat	QMTrat/QMRes
Resíduo	I(J-1)	SQRes	QMRes	
Total	IJ-1	SQTotal		

em que QMTrat = SQTrat/(I-1) e QMRes = SQRes/[I(J-1)].

Pode-se mostrar que o quociente QMTrat/QMRes tem distribuição F com (I-1) e I(J-1) graus de liberdade, supondo que,  $y_{ij}$  são variáveis aleatórias independentes, todos os tratamentos têm variância iguais a  $\sigma^2$  e  $Y_{ij} \sim N(\mu_i, \sigma^2)$ . Por esses motivos, os pressupostos da ANOVA devem ser testados ou avaliados em qualquer análise.

Se Fcalculado para tratamentos for maior que Ftabelado, ao nível  $\alpha$  de significância:  $F_{[\alpha,(I-1),I(J-1)]}$ , logo rejeitamos a hipótese de nulidade  $H_0$ , ou seja, existem evidências de diferenças significativas em pelo menos um contraste nas médias de tratamentos, ao nível  $\alpha$  de significância escolhido. Caso contrário, não rejeitamos a hipótese de nulidade  $H_0$ , ou seja, não há evidências de diferença significativa entre os níveis de tratamentos, ao nível  $\alpha$  de significância escolhido.

Em um experimento, cada observação  $Y_{ij}$  pode ser decomposta conforme o modelo a seguir, para um experimento com delineamento em blocos aleatorizados:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$
  $i = 1, \dots, I \ e \ j = 1, \dots, J$ 

em que:  $Y_{ij}$  é a observação do *i*-ésimo tratamento no *j*-ésimo bloco;  $\mu$  é a média geral;  $\tau_i$  é o efeito do *i*-ésimo tratamento;  $\beta_j$  é o efeito do *j*-ésimo bloco;  $\varepsilon_{ij}$  é o erro associado ao *i*-ésimo tratamento no *j*-ésimo bloco assumido como:  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2)$ .

# 2.3 Os modelos estatísticos com base em regressão e decomposição por valores singulares

O modelo mais simples para descrever observações fenotípicas ao longo de ambientes é o modelo de termo aditivo sem interação. Neste caso, a resposta fenotípica esperada para o genótipo i, (i = 1, ..., I), no ambiente j, (j = 1, ..., J), é igual à média geral mais os efeitos principais de genótipos e ambiente dado por

$$y_{i,j} = \mu + g_i + e_j + \varepsilon_{i,j} \tag{2}$$

O modelo aditivo é a base de todos os modelos com interação, mas é apenas aplicável quando não há GEI numa tabela de dupla entradas com genótipos nas linhas e ambientes nas colunas, isto é, quando a resposta fenotípica em ambientes é um conjunto de linhas paralelas. Se houver GEI, o modelo (2) pode ser escrito da forma

$$y_{i,j} = \mu + g_i + e_j + (ge)_{i,j} + \varepsilon_{i,j} \tag{3}$$

em que  $(ge)_{ij}$  representa o termo Genótipo  $\times$  ambiente  $(G \times E)$  para o genótipo i e o ambiente j. O modelo de interação completo (3) tem o maior número de parâmetros a serem estimados como combinações  $G \times E$ , que está associado com os testes menos precisos devido a falta de graus de liberdade, e representa um modelo menos parcimonioso.

Uma proposta alternativa ao modelo aditivo (2), foi proposto pela primeira vez por Finlay e Wilkinson (1963), em que as respostas fenotípicas em ambientes são regredido na média fenotípica sobre os ambientes (uma medida de produtividade ou de qualidade

biológica na ausência de outras caracterizações ambientais). O modelo pode ser escrito com a GEI como

$$y_{i,j} = \mu + g_i + e_j + \beta_i e_j + \varepsilon_{i,j} \tag{4}$$

Outro modelo baseado em regressão foi apresentada por Gusmão (1985), em que a informação de bloco (física) foi usado para corrigir os efeitos espaciais. Deste modo, as respostas fenotípicas por bloco são regredido em ambientes resultando na regressão  $i \times b$ , em que b é o número de blocos.

Uma outra alternativa para o modelo de interação completo é o modelo de efeitos principais aditivos e interação multiplicativa (AMMI) (GAUCH, 1992, GAUCH; ZOBEL, 1996; DUARTE; VENCOVSKY, 1999, GAUCH; PIEPHO; ANNICCHIARICO, 2008; GAUCH, 2013; RODRIGUES et al., 2014), que é mais flexível do que a regressão linear simples de Finlay e Wilkinson (1963), pois, pode dividir a interação em  $N = \min(I-1, J-1)$  termos. Ela combina a análise de variância (ANOVA) e análise de componentes principais (PCA), com ANOVA realizada inicialmente e, em seguida, DVS (decomposição em valores singulares) aplicada à matriz resultante de GEI (Gauch, 1992).

#### 2.4 Decomposição do valores singulares

A decomposição por valores singulares (singular value decomposition, DVS) é um dos resultados mais importantes da álgebra linear, tanto computacional quanto teórico, que é aplicada em diversas áreas de conhecimento e tem um importante papel na solução de sistemas de equações e portanto no método dos quadrados mínimos. Dada uma matriz  $\boldsymbol{A}$ ,  $(m \times n)$ , a matriz  $\boldsymbol{A}^T \boldsymbol{A}$ ,  $(n \times n)$ , é uma matriz simétrica com n valores próprios (ou autovalores) reais não negativos, iguais ou distintos,  $\lambda_1, \lambda_2, \ldots, \lambda_n$ . Designam-se por valores singulares da matriz  $\boldsymbol{A}$  a raíz quadrada dos n valores próprios da matriz  $\boldsymbol{A}^T \boldsymbol{A}$ ,  $\lambda_i^* = \sqrt{\lambda_i}$ .

O teorema da DVS é uma generalização do teorema de Jordan ou da decomposição espectral. Sua utilidade é bastante grande em algumas aplicações da análise multivariada como os modelos FAMMI (factor additive main effects and multiplicative interaction) e AMMI (additive main effects and multiplicative interaction).

Qualquer matriz X de ordem  $n \times m$ , com posto p < min(n, m), é possível

expressar cada elemento de X da seguinte maneira (MANDEL, 1982):

$$x_{ij} = \lambda_1 \gamma_{1i} \alpha_{1j} + \lambda_2 \gamma_{2i} \alpha_{2j} + \ldots + \lambda_p \gamma_{pi} \alpha_{pj} = \sum_{k=1}^p \lambda_k \gamma_{ki} \alpha_{kj},$$
 (5)

em que  $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \ldots \geq \lambda_p$ . Esta é a definição de DVS da matriz  $\boldsymbol{X}$ .

Os p vetores  $\gamma_i$  são ortogonais a cada vetor  $\gamma_j$ , para  $i \neq j$  e i, j = 1, 2, ..., p, assim como os p vetores  $\alpha_i$ . Além disso, considera-se que cada um desses vetores tem tamanho unitário (vetores normalizados), ou seja

$$\sum_{i} \gamma_{ki}^2 = \sum_{j} \alpha_{kj}^2 = 1, \qquad \forall k.$$
 (6)

Assim, podemos anunciar o seguinte teorema (JOHNSON; WICHERN, 1998; ANDERSON, 2003):

Teorema da Decomposição do Valor Singular: Toda matriz  $\boldsymbol{X}$   $(n \times m)$  de posto p pode ser decomposta em

$${}_{n}\boldsymbol{X}_{m} = {}_{n}\boldsymbol{U}_{p}\boldsymbol{\Lambda}_{p}\boldsymbol{V}_{m}^{T} \tag{7}$$

em que  $\boldsymbol{U}$   $(n \times p)$  e  $\boldsymbol{V}$   $(m \times p)$  são ortonormais por coluna, ou seja,  $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_p$  e  $\boldsymbol{\Lambda} = diag(\lambda_i^{1/2}), \ \lambda_i > 0$ . As quantidades  $\lambda_1, \lambda_2, \ldots, \lambda_p$  são os autovalores não nulos das matrizes  $\boldsymbol{X}\boldsymbol{X}^T$  ou  $\boldsymbol{X}^T\boldsymbol{X}$ ,  $\boldsymbol{U}$  e  $\boldsymbol{V}$  correspondem às matrizes formadas pelos p autovalores das matrizes  $\boldsymbol{X}\boldsymbol{X}^T$  ou  $\boldsymbol{X}^T\boldsymbol{X}$  dispostos em suas colunas, respectivamente.

**Demonstração**: Suponhamos que podemos decompor a matriz X, considerando

$$X = U\Lambda V^T$$
.

da mesma forma que foi apresentada no teorema, mas com U  $(n \times p)$  e V  $(m \times p)$  são ortonormais, ou seja,  $U^TU = UU^T = I$  e  $Q^TQ = QQ^T = I$ . A matriz  $\Lambda$ , agora é uma matriz  $(n \times m)$  contendo na sua diagonal os valores  $\sqrt{\lambda_i}$ ,  $\lambda_i \geq 0$ , i = 1, 2, ..., m. Sem perda de generalidade, vamos assumir que  $n \geq m$ . Obviamente, podemos construir as matrizes ortogonais U e V acrescentando n - p e m - p vetores, respectivamente, às colunas de U e V, ortonormais aos primeiros p termos.

Para uma maior compreensão da nova apresentação, vamos explicitar a matriz  $\pmb{\Lambda}$  por

$$\mathbf{\Lambda} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

A matriz  $\boldsymbol{X}^T\boldsymbol{X}$  possui autovalores  $\lambda_1,\lambda_2,\ldots,\lambda_m$ . Como  $\boldsymbol{X}^T\boldsymbol{X}$  é simétrica, pois  $(\boldsymbol{X}^T\boldsymbol{X})^T=\boldsymbol{X}^T\boldsymbol{X}$ , então existe  $\boldsymbol{V}$  ortogonal tal que

$$oldsymbol{V}^Toldsymbol{X}^Toldsymbol{X}oldsymbol{V} = oldsymbol{\Lambda}_m^2 = \left[egin{array}{cccc} \lambda_1 & 0 & \cdots & 0 \ 0 & \lambda_2 & \cdots & 0 \ dots & dots & \ddots & dots \ 0 & 0 & \cdots & \lambda_m \end{array}
ight]$$

em que V é a matriz dos autovetores de  $X^TX$  formando suas colunas.

Podemos assumir, ainda, uma situação mais restritiva, em que apenas r dos  $\lambda_i$ 's sejam não nulos e a matriz  $\Lambda$  (n × m) é dado por

$$\mathbf{\Lambda} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_p} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Essencialmente,  $\Lambda_p$  e  $\Lambda_m$  são idênticas nas primeiras r linhas e colunas. O que ocorre é que a matriz  $\Lambda_m$  possui m-r linhas e m-r colunas de zeros a mais que a matriz  $\Lambda_p$ . Mais especialmente, podemos ver facilmente que  $\Lambda^T \Lambda = \Lambda_m^2$ , dada por

$$\mathbf{\Lambda}_{m}^{2} = \begin{bmatrix} \lambda_{1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \lambda_{2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{p} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

Assim, podemos verificar que  $(\boldsymbol{V}^T\boldsymbol{X}^T)\boldsymbol{X}\boldsymbol{V}=\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}$ , ou seja,  $(\boldsymbol{X}\boldsymbol{V})^T(\boldsymbol{X}\boldsymbol{V})=\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}$ . Fazendo  $\boldsymbol{W}=\boldsymbol{X}\boldsymbol{V}$ , temos que  $\boldsymbol{W}^T\boldsymbol{W}=\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}$ . sendo que  $\boldsymbol{W}=[\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_p,\cdots,\boldsymbol{w}_m]$  e concluirmos que

$$\boldsymbol{w}_i^T \boldsymbol{w}_i = \begin{cases} \lambda_i & se \ i \le p \\ 0 & se \ i > p \end{cases}$$

em que  $\boldsymbol{w}_i^T \boldsymbol{w}_j$  se  $i \neq j = 1, 2, \cdots, m$ .

Se  $\boldsymbol{w}_i=0$  para i>p, as primeiras colunas de  $\boldsymbol{W}$  são linearmente independentes. Logo, concluímos que  $p\leq n.$  Se definirmos

$$oldsymbol{u}_i = rac{1}{\sqrt{\lambda_i}} oldsymbol{w}_i = rac{1}{\sqrt{\lambda_i}} oldsymbol{X} oldsymbol{v}_i$$

em que  $i = 1, 2, \dots, p$ , então devemos tomar  $\boldsymbol{u}_{p+1}, \boldsymbol{u}_{p+2}, \dots, \boldsymbol{u}_m$ , se p < n, ortogonais entre si e aos demais  $\boldsymbol{u}_i$ 's tais que a matriz  $(n \times n)$  seja ortogonal:

$$oldsymbol{U} = [oldsymbol{u}_1, oldsymbol{u}_2, \cdots, oldsymbol{u}_n]$$

Podemos verificar pela definição de  $u_i$ , que a expressão  $u_i\sqrt{\lambda_i}=\boldsymbol{X}\boldsymbol{v}_i$  pode ser expressa matricialmente por  $U\boldsymbol{\Lambda}=\boldsymbol{X}\boldsymbol{V}$ . Como  $\boldsymbol{V}$  é ortogonal, temos que  $\boldsymbol{V}^{-1}=\boldsymbol{V}^T$ , resultando em

$$egin{aligned} oldsymbol{X}oldsymbol{V} & oldsymbol{U}oldsymbol{\Lambda} oldsymbol{V}^T & oldsymbol{U}oldsymbol{\Lambda}oldsymbol{V}^T \end{aligned}$$

resultando em  $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T$ . Podemos, então considerar apenas as p primeiras colunas de  $\boldsymbol{U}$ , as p primeiras linhas e colunas de  $\boldsymbol{\Lambda}$  e as p primeiras colunas de  $\boldsymbol{V}$  e finalizamos a demonstração do teorema (JOHNSON; WICHERN, 1998).

Esse teorema pode ser utilizado para resolver problemas que envolvem o método dos quadrados mínimos em muitas aplicações da estatística univariada e multivariada, também pode ser utilizado para definirmos outra forma de obter inversas generalizadas de uma matriz  $\boldsymbol{X}$   $(n \times m)$ . Assim, podemos definir a matriz inversa generalizada  $\boldsymbol{X}^ (m \times n)$  da seguinte forma:

$$\boldsymbol{X}^{-} = \boldsymbol{U}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{T}$$

em que  $X^-$  não é única e segue a propriedade  $XX^-X=X$  e  $\Lambda^{-1}=diag(\lambda_i^{-1/2})=diag(\frac{1}{\sqrt{\lambda_i}}).$ 

### 2.5 Análise de Componentes Principais

A análise de componentes principais é uma técnica multivariada de modelagem da estrutura de covariância. A técnica foi inicialmente descrita por Pearson (1901) e uma descrição de métodos computacionais práticos veio muito mais tarde com Hotelling (1933, 1936) quem a usou com o propósito determinado de analisar as estruturas de correlação.

A análise de componentes principais (ACP) tem como objetivo reduzir a dimensionalidade do conjunto de variáveis e facilitar a interpretação da independência entre elas. Para isso, são obtidas combinações lineares das variáveis originais que, geometricamente representam a seleção de novos sistemas de coordenadas obtidos pela rotação do sistema original que tem p variáveis aleatórias como eixos das coordenadas. Esses novos eixos ortogonais (novas variáveis) são chamados de componentes principais e os valores das novas variáveis são chamadas de escores dos componentes principais ou coordenadas principais. Essas novas variáveis são não correlacionadas (JOHNSON; WICHERN, 1998). Algumas característica devem ser observadas nas modelagem da matriz de covariâncias. Se tivermos estudando um fenômeno observando p variáveis relacionadas, teremos p variáveis latentes que devem ser mutuamente não-correlacionadas e é comum ordenarmos essas variáveis latentes em função de suas variâncias.

Em geral escolhemos o componente principal de maior importância (o primeiro componente principal) como sendo aquele de maior variância  $(\max_i \lambda_i)$ , que explique o máximo de variabilidade dos dados, o segundo componente de maior importância, o que apresenta a segunda maior variância e assim sucessivamente, até o componente principal de menor importância. Por outro lado, os últimos componentes principais serão responsáveis por direções que não estão associadas a muita variabilidade. Em outras palavras, esses últimos componentes principais identificarão relações lineares entre as variáveis originais próximo de constante (JOHNSON; WICHERN, 1998; ANDERSON, 2003; FERREIRA, 2011).

A obtenção dos componentes principais é realizada por meio da diagonalização de matrizes simétricas positivas semi-definidas. Então, podemos calcular os componentes principais facilmente e utilizá-los em diferentes aplicações nas mais variadas áreas científicas. Esta facilidade é função da existência de inúmeros programas capazes de realizar cálculos matriciais para diagonalizar uma matriz simétrica positiva semi-definida. Muitos pesquisadores têm utilizado a análise de componentes principais para resolver problemas

como da multicolinearidade em regressão linear, para estimar fatores, que representam outra técnica multivariada de modelagem da matriz de covariâncias, para realizar a modelagem da interação entre fatores em experimentos sem repetição, estudos de divergência e agrupamento entre genótipos em estudo de genética e melhoramento de plantas e animais, entre outras possibilidades (JOHNSON; WICHERN, 1998).

Para obter os componentes principais de uma forma geral, seja um conjunto de p variáveis  $X_1, X_2, \dots, X_p$  com médias  $\mu_1, \mu_2, \dots, \mu_p$  e variância  $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$ , respectivamente. Estas variáveis não são independentes e portanto, possuem covariância entre a i-ésima e k-ésima variável definida por  $\sigma_{ik}$ , para  $i \neq k = 1, 2, \dots, p$ . Então as p variáveis podem ser expressas na forma vetorial por:  $\mathbf{X} = [X_1, X_2, \dots, X_p]^T$ , com vetor de médias  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]^T$  e matriz de covariância  $\boldsymbol{\Sigma}$ . Encontram-se os pares de autovalores e autovetores  $(\lambda_1, \boldsymbol{e}_1), (\lambda_2, \boldsymbol{e}_2), \dots, (\lambda_p, \boldsymbol{e}_p)$ , em que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , associados a  $\boldsymbol{\Sigma}$  e então o i-ésimo componente principal é definido por:

$$\boldsymbol{Z}_i = \boldsymbol{e}_i^T \boldsymbol{X} = e_{i1} X_1 + e_{i2} X_2 + \dots + e_{ip} X_p$$
(8)

em que  $i = 1, 2, \dots, p$ .

A variável  $Z_i$ , é uma variável latente, ou seja, não é mensurada a partir do experimento ou levantamento amostral (JOHNSON; WICHERN, 1998). O objetivo é determiná-la a partir das p variáveis contidas no vetor  $\boldsymbol{X}$ . A idéia é projetar os pontos coordenados originais em um plano maximizando a distância entre eles, o que equivale a maximizar a variabilidade da variável latente  $\boldsymbol{Z}_i$ . A variância de  $\boldsymbol{Z}_i$  é dada por

$$Var(\mathbf{Z}_i) = Var(\mathbf{e}_i^T \mathbf{X}) = \mathbf{e}_i^T Var(\mathbf{X}) \mathbf{e}_i = \mathbf{e}_i^T \mathbf{\Sigma} \mathbf{e}_i$$
(9)

e a covariância entre  $\mathbf{Z}_i$  e  $\mathbf{Z}_k$   $(i \neq k)$ , dada por

$$Cov(\boldsymbol{Z}_i, \boldsymbol{Z}_k) = Cov(\boldsymbol{e}_i^T \boldsymbol{X}, \boldsymbol{e}_k^T \boldsymbol{X}) = \boldsymbol{e}_i^T Var(\boldsymbol{X}) \boldsymbol{e}_k = \boldsymbol{e}_i^T \boldsymbol{\Sigma} \boldsymbol{e}_k$$
 (10)

A definição dos componentes principais é baseada na maximização de sua variância. No entanto, o máximo da variância do componente principal não existe, pois na medida em que os componentes do vetor de coeficientes  $\mathbf{e}_i$  crescem, a variância crescerá para infinito. Assim, devemos impor a restrição de que a soma ao quadrado dos coeficientes do vetor, que estabelece a combinação linear dos componentes principais, seja iguais à unidade. Assim, devemos maximizar a variância  $\operatorname{Var}(Y_i) = \mathbf{e}_i^T \mathbf{\Sigma} \mathbf{e}_i$  com relação ao vetor  $\mathbf{e}_i$ , sujeito à restrição de  $\mathbf{e}_i^T \mathbf{e}_i = 1$ . Para isso, podemos utilizar a técnica de multiplicadores de Lagrange, na qual devemos maximizar

$$\max_{\boldsymbol{e}_i} \left[ \boldsymbol{e}_i^T \boldsymbol{\Sigma} \boldsymbol{e}_i - \lambda_i (\boldsymbol{e}_i^T \boldsymbol{e}_i - 1) \right]$$

em relação a  $e_i$ , sendo  $\lambda_i$  o multiplicador de Lagrange.

Uma abordagem mais tradicional impõe a restrição de outra forma dividindo a função original por  $e_i^T e_i$ , resultando em uma função que devemos maximizar da seguinte forma:

$$\lambda_i = \max_{oldsymbol{e}_i} rac{oldsymbol{e}_i^T oldsymbol{\Sigma} oldsymbol{e}_i}{oldsymbol{e}_i^T oldsymbol{e}_i}$$

Para obtermos o máximo, em ambos os casos, derivamos a função resultante em relação a  $e_i$  e igualamos o sistema de equações a zero, resulta em

$$(\mathbf{\Sigma} - \lambda_i \mathbf{I}) \mathbf{e}_i = \mathbf{0} \tag{11}$$

em que I é uma matriz identidade  $(p \times p)$ . Pela equação (11), verificamos:

$$\Sigma e_i = \lambda_i e_i \tag{12}$$

em que podemos mostrar que

$$Var(\mathbf{Z}_i) = \mathbf{e}_i^T \mathbf{\Sigma} \mathbf{e} = \mathbf{e}_i^T \lambda_i \mathbf{e}_i = \lambda_i \mathbf{e}_i^T \mathbf{e}_i = \lambda_i$$

pode ser também

$$Cov(\boldsymbol{Z}_i, \boldsymbol{Z}_k) = \boldsymbol{e}_i^T \boldsymbol{\Sigma} \boldsymbol{e}_k = \boldsymbol{e}_i^T \lambda_k \boldsymbol{e}_k = \lambda_k \boldsymbol{e}_i^T \boldsymbol{e}_k = 0$$

com  $i \neq k$ , uma vez que  $e_i$  e  $e_k$  são ortogonais.

Podemos observar que a definição dos componentes principais é a obtenção dos autovalores e autovetores da matriz  $\Sigma$ . Os autovetores  $e_i$   $(i=1,\cdots,p)$ , definem as variâncias desses novos eixos coordenados. Como as covariâncias são nulas entre os diferentes pares de eixos, os novos eixos, que são os componentes principais, são ortogonais e, portanto, representam uma rotação rígida dos eixos originais. Se fizemos  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ , definirmos os componentes principais  $\mathbf{Y}_1 = \mathbf{e}_1^T \mathbf{X}$ ,  $\mathbf{Y}_2 = \mathbf{e}_2^T \mathbf{X}$ ,  $\cdots$ ,  $\mathbf{Y}_p = \mathbf{e}_p^T \mathbf{X}$ , respectivamente. Dessa forma, temos tantos componentes principais quanto variáveis originais (JOHNSON; WICHERN, 1998; ANDERSON, 2003; FERREIRA, 2011).

Se utilizarmos a decomposição espectral da matriz  $\Sigma$ , dado por  $\Sigma = P\Lambda P^T$ , em que P é a matriz composta pelos autovetores de  $\Sigma$  em suas colunas e  $\Lambda$ , a matriz diagonal de autovalores de  $\Sigma$ , então, podemos observar que (JOHNSON; WICHERN, 1998):

$$tr(\mathbf{\Sigma}) = tr(\mathbf{P}\mathbf{\Lambda}\mathbf{P}^T) = tr(\mathbf{\Lambda}\mathbf{P}^T\mathbf{P}) = tr(\mathbf{\Lambda}\mathbf{I}) = tr(\mathbf{\Lambda}) = \sum_{i=1}^p \lambda_i$$

Mas  $tr(\Sigma)$  é dada pela soma dos elementos da diagonal, ou seja,

$$tr(\Sigma) = \sum_{i=1}^{p} \sigma_{ii}$$

portanto,

$$\sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i \tag{13}$$

que significa a variabilidade total contida nas variáveis originais é igual à variabilidade total contida nos componentes principais.

Assim, como a variância total é igual ao traço da matriz de covariância, então a variância total contida nas variáveis originais é igual ao traço de  $\Sigma$  que é  $tr\Sigma = \sum_{i=1}^p \sigma_{ii}$ . Da mesma forma, a variância total de Z é igual a  $tr\Lambda = \sum_{i=1}^p \lambda_i$ . Pela equação (13), a variabilidade total contida nas variáveis originais é preservada nas variáveis transformadas, componentes principais. Quando adotamos o modelo parcimonioso, com k < p componentes principais, temos que a covariância do vetor Z ( $k \times 1$ ):

$$Cov(\boldsymbol{Z}) = Cov(\boldsymbol{P}_k^T \boldsymbol{X}) = \boldsymbol{P}_k^T Cov(\boldsymbol{X}) \boldsymbol{P}_k = \boldsymbol{P}_k^T \boldsymbol{\Sigma} \boldsymbol{P}_k = \boldsymbol{P}_k^T \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}^T \boldsymbol{P}_k = \boldsymbol{\Lambda}_k$$

em que

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$$

Portanto, a variância total do vetor  $\boldsymbol{Y}$   $(k \times 1)$  é  $tr(\boldsymbol{\Lambda}_k) = \sum_{i=1}^k \lambda_i$ . Podemos estabelecer que a explicação do modelo reduzido em relação ao modelo completo pode ser obtida pela relação das variâncias dos dois modelos, ou seja, quanto da variação total das variáveis originais é explicada pelo modelo de k componentes principais. Essa proporção acumulada da explicação da variação total, expressa em porcentagem é dada seguinte:

$$\rho_k^2 = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \sigma_{ii}} \times 100\%$$
(14)

A explicação individual de cada componente pode ser calculado, por exemplo, para k-ésimo componente principal a proporção da explicação é dada por

$$P_k^2 = \frac{\lambda_k}{\sum\limits_{i=1}^p \sigma_{ii}} \times 100\%$$

Pela proporção de explicação da variância total, que o modelo de k componentes principais é responsável, podemos determinar o número de componentes devemos reter. Muitos casos, adotamos modelos que expliquem pelos menos 80% da variação total (JOHNSON; WICHERN, 1998) Se o número mínimo de componentes k para explicar a proporção mínima da variação total for muito grande, em muitas aplicações práticas não haverá vantagem em utilizar o método dos componentes principais.

#### 2.6 Histórico dos modelos AMMI

Em melhoramento de plantas, ensaios multi-ambientes (MET-multienvironment trials) são importantes para testar adaptação geral e específica de cultivar. Um cultivar crescendo em diferentes ambientes irá freqüentemente mostrar flutuações significantes na performance de produção relativa a outras cultivares. Essas mudanças são influenciadas por diferentes condições experimentais e são referidas como interação  $G \times E$  (DIAS, 2005; HONGYU et al., 2014).

O termo G × E como sendo linearmente relacionados ao efeito de ambiente, isto é colocar  $(ge)_{ij} = \xi_i e_j + d_{ij}$  em que  $\xi_i$  é o coeficiente de regressão linear do *i*-ésimo genótipo na média ambiental  $e_j$  e  $d_{ij}$  é um desvio (YATES; COCHRAN, 1938). Este procedimento foi posteriormente usado por Finlay e Wilkinson (1963) e modificado por Eberhart e Russell (1966). Tukey (1949) propôs um teste para G × E usando  $(ge)_{ij} = Kg_ie_j$  (K é uma constante). Mandel (1961) generalizou o modelo de Tukey colocando  $(ge)_{ij} = \lambda \gamma_i e_j$  para genótipos e  $(ge)_{ij} = \lambda g_i \alpha_j$  para ambientes e assim obtendo um feixe de linhas retas que podem ser testadas para concorrência (isto é, se os  $\gamma_i$  ou os  $\alpha_j$  são todos os mesmos) ou não concorrência (DIAS, 2005).

Gollob (1968) e Mandel (1969, 1971) propuseram um termo bilinear GEI  $(ge)_{ij} = \sum\limits_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}$  no qual  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$  e  $\gamma_{ik}$ ,  $\alpha_{jk}$  satisfazem a restrição de ortonormalidade  $\sum\limits_i \gamma_{ik} \gamma_{ik'} = \sum\limits_j \alpha_{jk} \alpha_{jk'} = 0$  para  $k \neq k'$  e  $\sum\limits_i \gamma_{ik}^2 = \sum\limits_j \alpha_{jk}^2 = 1$ .

Isto conduz ao modelo linear-bilinear  $\overline{Y}_{ij} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \overline{\varepsilon}_{ij}$ , o qual é uma generalização da regressão no modelo de média, com mais flexibilidade para descrever  $G \times E$  porque mais que uma dimensão genótipo ambiente é considerada (DIAS, 2005; DIAS; KRZANOWSKI, 2006). Gauch (1988, 1992) chamou esse modelo de "AMMI" (Additive Main Effects and Multiplicative Interaction Model).

Uma família de modelos multiplicativos pode então ser gerada retirando-se o efeito principal de genótipo (modelo de regressão para ambientes), (SREG – Site Regression Model), o efeito principal de ambientes (GREG – Genotype Regression Model), ou ambos efeitos (COMM – Complete Multiplicative Model) e GGE (Genotype main effects + genotype environment interaction) proposto por Yan e Kang (2003). Um outro modelo multiplicativo (SHMM – Shifted Multiplicative Model), é útil para estudar crossover G × E (CROSSA; CORNELIUS; YAN, 2002; DIAS; KRZANOWSKI, 2003).

#### 2.7 Histórico Biplot

O método biplot foi desenvolvido por GABRIEL (1971) para representar graficamente resultados de análise de componentes principais ou de decomposição de valores singulares (DVS), no qual o valor de cada elemento de uma tabela de dupla entrada pode ser visualizado pelo produto de vetores e pelo co-seno do ângulo entre dois vetores (YAN; KANG, 2003). Quando duas matrizes apresentarem o mesmo número de linhas e colunas, é possível multiplicá-las. A nova matriz gerada da multiplicação das duas anteriores assume o mesmo número de linhas e colunas de ambas as matrizes (GABRIEL, 1971; YAN; KANG, 2003; SILVA, BENIN, 2012). Dessa forma, qualquer matriz de n observações e p variáveis pode ser representada por um biplot. Os valores singulares da DVS são distribuídos para os marcadores de linhas e colunas de diferentes maneiras, dependendo se o interesse está na relação entre linhas (JK-Biplot), ou entre colunas (GH-Biplot) ou entre linhas e colunas (SQRT-Biplot).

Alguma técnica de redução de dimensão é normalmente usada para representar as amostras como pontos, muitas vezes, análise de componentes principais

(ACP) (PEARSON, 1901; HOTELLING, 1933) ou análise de variáveis canônicas (AVC) (HOTELLING, 1935, 1936). De modo mais geral, as técnicas de dimensionamento, como análise de coordenadas principais (PCO) (TORGERSON, 1952; GOWER, 1966) ou escalonamento multidimensional métrica ou não-métrica (MDS) (KRUSKAL, 1964a, 1964b; SAMMON, 1969) são usados. Jolliffe (2002) dedica uma monografia ao ACP, enquanto Krzanowski (2000) abrange temas mais gerais com multivariados. Cox e Cox (2001) e Borg e Groenen (2005) são referências padrão para técnicas de escala.

A colinearidade entre os vetores linhas e colunas em tabela de dupla entrada, aproximada sobre o gráfico Biplot, pode dar indícios para o ajuste de alguns modelos bilineares do tipo multiplicativo (BRADU; GABRIEL, 1978). Foi estudado por Gabriel e Zamir (1979) a interação dos modelos bilineares por meio de uma aproximação de posto inferior. Também foi demonstrado que os biplots são melhores para descrever e analisar um conjunto de dados comparando com a análise exploratória de dados (COX; GABRIEL, 1982). As vantagens do método Biplot em relação à seleção e diagnóstico de modelos que se ajustam a matriz de dados multivariados foi descrito por Gabriel (1981). Galindo (1986) propõe uma alternativa de representação simultânea para as linhas e colunas (HJ-Biplot), mas com a desvantagens de não reproduzir os dados originais.

Os Biplots não lineares foram desenvolvidos, utilizados para obter o ajuste de trajetórias não lineares e representar as variáveis que se projetam sobre as representações obtidas por meio de coordenadas principais (GOWER; HARDING, 1988; GOWER, 1992). Yan (2001) utilizou o Biplot para estudo de GEI, com os modelos bilineares para a análise de GEI (linhas-colunas) na área da agronomia, possibilitando o estudo da estabilidade dos genótipos a serem testados em vários ambientes (SOUZA, 2010). O Meta Biplot foi desenvolvido por Martín-Rodriguez et al. (2002) para comparação de diversos Biplots, sendo aplicado ao estudo com bastantes indivíduos e variáveis ao longo do tempo em diversas situações experimentais.

Varela (2002) utilizou o Biplot como ferramenta para explicar a interação de ordem superior associada a um modelo linear, dando ênfase às interações de segunda e terceira ordem, para as tabelas de dupla e tripla entrada, respectivamente, e também para identificar as linhas e as colunas responsáveis pela ausência/presença de interação e no diagnóstico de modelos, identificando os diferentes tipos de interações presentes em uma tabela de múltiplas entradas (SOUZA, 2010). O Biplot para detectar multicolinearidade entre as variáveis baseado na decomposição espectral da inversa da matriz de correlações foi

estudado por Ramíres et al. (2005). Para as variáveis que são binárias (presença/ausência), foi sugerido para a utilização do Biplot Logístico (VICENTE-VILLARDÓN; GALINDO-VILLARDÓN; BLÁSQUEZ-ZABALLOS, 2006), o qual é uma representação baseada em modelos de regressão logística.

Muitos pacotes estatísticos podem ser utilizados para produzir, pelo menos, o mais simples de biplots da abordagem tradicional. Estes incluem os principais pacotes estatísticos Minitab (MINITAB, 2007), SPSS (SPSS, 2008), Stata (STATACORP, 2007) e vários produtos do SAS (SAS, 2009). Muitas vezes, a funcionalidade é limitada e difíceis de obter os resultados. Maior funcionalidade é fornecida pelos três programas dedicados ao biplot como: XLS-Biplot (UDINA, 2005a, 2005b), o GGEBiplot (YAN; KANG, 2006) e biplot (LIPKOVICH; SMITH, 2002a, 2002b). O XLS-Biplot é baseado em XLisp-Stat (TIERNEY, 1990) e tem muitas características úteis, incluindo um servidor-web relacionada que pode ser usado para construir biplots online. GGEbiplot é voltado principalmente para engenheiros agrônomos, fitotecnistas e geneticistas. Ela complementa o livro de Yan e Kang (2003), que é a utilização da metodologia GGE (Genotype main effects + Genotype environment interaction) para o estudo de GEI. O software possui várias características especiais para análise de dados, principalmente para o método GGE Biplot e construção de vários tipos de Biplots.

Além do software **GGEBiplot**, existem outros também para construção de Biplot, como software **R** (R DEVELOPMENT, 2009). O software **R** é livre e desenvolvido por vários pesquisadores e possui diversos pacotes para análise de dados, nesse caso para estudo de Biplot, tem como pacote **Agricolae** desenvolvido por Felipe de Mendiburu e Reinhard Simon, utilizado em pesquisas agrícolas e em vários tipos de análises de delineamentos como, análise do modelo AMMI com gráfico Biplot. O Pacote **Bpca** desenvolvido por José Cláudio Faria com implementação dos métodos clássicos dos Biplots: SQRT, JK, GH e HJ. A apresentação dos gráficos podem ser na forma de 2D ou 3D.

O pacote **GGEBiplotGUI** (2013) fornece uma interface gráfica para a construção de GGE biplots em R. Algumas das funções do pacote são: (i) classificação das cultivares com base em seu desempenho em qualquer ambiente; (ii) classificação dos ambientes com base no desempenho relativo das cultivares; (iii) comparar o desempenho de qualquer par de cultivares em ambientes diferente; (iv) identificar a melhor cultivar em cada ambiente; (v) o agrupamento dos ambientes baseados nas melhores cultivares; (vi) a avaliação das cultivares com base em ambos rendimento médio e a estabilidade e (vii) a

avaliação dos ambientes baseados em ambos capacidade de discriminar e representatividade.

# 2.8 Tipos de Biplot

A escolha dos marcadores de linhas e colunas a partir da aproximação de posto inferior obtido, pode ser feita de diversas maneiras, por essa razão vários autores propõem distintas escolhas e estudam suas propriedades conforme a fatoração escolhida (SOUZA,2010). A representação é realizada pela fatoração da matriz **Y** em um produto de matrizes, tem-se:

$$Y \approx (U\Lambda^s)(\Lambda^{1-s}V^T) = AB^T$$
(15)

em que as matrizes  $\boldsymbol{A}$  e  $\boldsymbol{B}$  possuem posto q com as respectivas coordenadas dos n vetores marcadores de linhas  $\boldsymbol{a}_i$  e dos p vetores marcadores de colunas  $\boldsymbol{b}_j$ , com  $i=1,\ldots,n$  e  $j=1,\ldots,p$ .

Na equação (15), a constante  $s \in [0,1]$ . Assim, Gabriel (1971) propõe três representações de Biplots: GH-Biplot, JK-Biplot e SQRT-Biplot e Galindo (1986) propõe uma modificação com o HJ-Biplot (SOUZA, 2010).

i) Se s=0 na equação (15), temos o **GH-Biplot** ou **CMP-Biplot** (Column Metric Preserving). Este Biplot é obtido pela escolha dos marcadores de colunas, conforme as expressões  $\mathbf{A} = \mathbf{U}$  e  $\mathbf{B} = \mathbf{V}\mathbf{\Lambda}$ . Pode-se fazer um ajuste nos marcadores associados (multiplicando e dividindo por um fator de escala) dada abaixo:

$$\boldsymbol{A}_{(q)} = \sqrt{(n-1)}\boldsymbol{U}_{(q)}$$

е

$$oldsymbol{B}_{(q)} = rac{1}{\sqrt{(n-1)}} oldsymbol{V}_{(q)} oldsymbol{\Lambda}_{(q)}$$

em que a restrição métrica  $\boldsymbol{A}^T\boldsymbol{A}=\boldsymbol{I}$  e  $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T=\boldsymbol{A}\boldsymbol{A}^T.$ 

Os produtos escalares das colunas da matriz  $\mathbf{Y}^T\mathbf{Y}$  coincidem com os produtos escalares dos marcadores  $\mathbf{B}\mathbf{B}^T$ . Para que a matriz de variância e covariância  $\mathbf{S}$  seja proporcional a matriz  $(\mathbf{Y}^T\mathbf{Y})_{p\times p}$ , temos:

$$egin{aligned} oldsymbol{S} &= rac{1}{(n-1)} oldsymbol{Y}^T oldsymbol{Y} = rac{1}{(n-1)} (oldsymbol{A} oldsymbol{B}^T)^T (oldsymbol{A} oldsymbol{B}^T) = rac{1}{(n-1)} oldsymbol{B} oldsymbol{A}^T oldsymbol{A} oldsymbol{B}^T \ &= oldsymbol{B} oldsymbol{U}^T oldsymbol{U} oldsymbol{B}^T = oldsymbol{B} oldsymbol{B}^T \end{aligned}$$

em que a decomposição espectral da matriz de covariância é também uma decomposição em valores singulares:

$$oldsymbol{S} = rac{1}{(n-1)} oldsymbol{Y}^T oldsymbol{Y} = rac{1}{(n-1)} oldsymbol{V} oldsymbol{\Lambda}^2 oldsymbol{V}^T$$

Pela equação acima, está aproximando uma matriz  $Y_{n\times p}$  de posto (r) para uma matriz de menor posto (q), logo, perde-se informação e a representação do Biplot é aproximada (SOUZA, 2010). Para medir essas perdas consiste em calcular a qualidade de representação do ajuste para a aproximação da matriz de variância e covariância que é definida por:

$$\left(\sum_{k=1}^{q} \lambda_k^2 / \sum_{k=1}^{r} \lambda_k^2\right).100\%$$

em que  $\lambda_k$  são os autovalores da matriz  $\boldsymbol{Y}$ .

No GH-Biplot obtém-se uma máxima qualidade de representação para as colunas da matriz  $\boldsymbol{Y}$  e uma mínima qualidade de representação para as linhas.

ii) Se s=1 na equqação (15), temos o **JK-Biplot** ou RMP-Biplot (Row Metric Preserving). Considera-se a matriz  $\mathbf{Y}$  centrada por colunas. Este Biplot é obtido pela escolha dos marcadores de linhas, conforme as expressões abaixo:

$$egin{aligned} oldsymbol{A}_{(q)} &= oldsymbol{U}_{(q)} oldsymbol{\Lambda}_{(q)} \ &e \ oldsymbol{B}_{(q)} &= oldsymbol{V}_{(q)} \end{aligned}$$

em que a restrição métrica  $\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{I}$  e  $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T = \boldsymbol{B}\boldsymbol{B}^T$ .

A aproximação dos produtos escalares em dimensão reduzida é considerada ótima no sentido dos mínimos quadrados de:

$$YY^T = AB^TBA^T = AV^TVA^T = AA^T$$

A decomposição espectral da matriz de produtos escalares entre as linhas é também sua decomposição em valores singulares:

$$\boldsymbol{Y}\boldsymbol{Y}^T = \boldsymbol{U}\boldsymbol{V}^2\boldsymbol{U}^T$$

portanto, a melhor aproximação é a de posto q que coincide com a aproximação que se obtém da matriz Y:

$$oldsymbol{Y}oldsymbol{Y}^Tpprox oldsymbol{U}_{(q)}oldsymbol{\Lambda}_{(q)}^2oldsymbol{U}_{(q)}^T=oldsymbol{A}_{(q)}oldsymbol{A}_{(q)}^T$$

Entretanto, a distância euclidiana entre duas linhas de Y coincide com a distância euclidiana entre os marcadores de A. A qualidade da representação é melhor para as linhas do que para as colunas. Os marcadores para as colunas são as projeções dos eixos originais (base canônica em espaço p-dimensional) no espaço das componentes principais (SOUZA, 2010).

iii) Se s=1/2 na equ<br/>qação (15), temos o **SQRT-Biplot** (Square Root Biplot) ou Biplot simétrico: este Biplot é obtido pe<br/>la escolha dos marcadores de linhas e colunas, conforme as expressões abaixo:

$$oldsymbol{A}_{(q)} = oldsymbol{U}_{(q)} oldsymbol{\Lambda}_{(q)}^{1/2}$$

е

$$oldsymbol{B}_{(q)} = oldsymbol{V}_{(q)} oldsymbol{\Lambda}_{(q)}^{1/2}$$

com restrição métrica  $\mathbf{A}^T \mathbf{A} \neq \mathbf{I}$  e  $\mathbf{B}^T \mathbf{B} \neq \mathbf{I}$ . Este tipo de Biplot é utilizado quando o objetivo da análise consiste na aproximação dos elementos  $Y_{ij}$  da matriz  $\mathbf{Y}$ , por exemplo, no caso de modelos de diagnóstico em tabelas de contingência, em que as linhas e as colunas têm as mesmas importâncias (SOUZA, 2010).

iv) O HJ-Biplot (Row-Column Metric Preserving): Quando as linhas e colunas têm as mesmas importâncias e querem interpretar as características de ambas, mantendo a relação entre elas, são utilizadas interpretações simétricas como Análise Fatorial de Correspondência, em que interpretam as posições das linhas, as posições das colunas e as relações linha-coluna por meio dos fatores. Galindo (1986) propõe por meio da DVS a escolha dos marcadores:

$$oldsymbol{A}_{(q)} = oldsymbol{U}_{(q)} oldsymbol{\Lambda}_{(q)}$$

е

$$oldsymbol{B}_{(q)} = oldsymbol{V}_{(q)} oldsymbol{\Lambda}_{(q)}$$

com as expressões dada acima, os dados originais não se reproduzem, ou seja,  $Y \neq AB^T$ . Neste caso, o objetivo é ter uma máxima qualidade de representação do ajuste para as linhas e colunas de Y, em que é incorporado a cada matriz marcadores, a matriz diagonal  $\Lambda$ , que possibilita a decomposição espectral tanto para a matriz de variâncias e covariâncias entre colunas como para a matriz de distâncias euclidianas entre linhas, coincida com a DVS de Y (SOUZA, 2010).

Como as coordenadas das linhas do HJ-Biplot coincidem com as do JK-Biplot e as coordenadas das colunas são as mesmas do GH-Biplot, o HJ-Biplot possui as mesmas propriedades dos respectivos Biplot desenvolvidos por Gabriel (1971). A proximidade entre indivíduos é interpretada como similaridade, os ângulos entre dois vetores que unem duas variáveis com a origem, como correlação. Projetando os indivíduos sobre os vetores que representam as variáveis é possível determinar a distribuição aproximada que os indivíduos apresentam sobre estas variáveis (CÁRDENAS; GALINDO; VICENTE-VILLARDÓN, 2007; SOUZA, 2010). A vantagem de se utilizar HJ-Biplot, é que ele proporciona uma melhor qualidade na representação simultânea tanto para as linhas (indivíduos) como para as colunas (variáveis).

### 2.9 Teoria e conceito do Biplot AMMI

O método AMMI está na possibilidade de representar, num único gráfico, os efeitos de interação para cada genótipo e cada ambiente no biplot baseando-se na aproximação de Decomposição por valores singulares (DVS) da matriz de  $\mathbf{GE}$  (interação genótipo  $\times$  ambiente).

Para permitir a construção do gráfico o posto da matriz aproximada deverá, na realidade, ser igual a um, dois ou três, resultando num biplot em uma, duas ou três dimensões, respectivamente. Na análise AMMI o uso do biplot é garantido em razão da DVS aplicada à matriz de interações, que será representada, na maioria das aplicações, por meio de sua aproximação de posto um ou dois. Seja, então, a aproximação DVS para  $\mathbf{GE}$  (matriz de posto conhecido p) é da seguinte forma,  $\widehat{\mathbf{GE}}_{(g \times e)} = \sum_{k=1}^{n} \lambda_k \gamma_k \alpha_k' = \widetilde{\mathbf{USV}}'$  (GAUCH, 1992; RODRIGUES, et al., 2014).

em que,  $\widetilde{\mathbf{U}}_{g\times e}$  tem em suas colunas apenas os n primeiros vetores  $\gamma_g$ ;

 $\widetilde{\mathbf{S}}_n$  é a matriz diagonal com os primeiros valores singulares:  $\lambda_1, \lambda_2, \dots, \lambda_n$  com n < p; e  $\widetilde{\mathbf{V}}'_{n \times e}$  tem em suas linhas os n primeiros vetores  $\alpha'_{1 \times e}$ . A expressão pode também ser escrita como:

$$\widehat{GE}_{(q \times e)} = (\widetilde{U}\widetilde{S}^{1/2})(\widetilde{S}^{1/2}\widetilde{V}') = (\widetilde{G})(\widetilde{H}')$$
(16)

em que  $\widetilde{\mathbf{G}}_{(g\times n)}=(\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}^{1/2})$  e  $\widetilde{\mathbf{H}}'_{n\times e}=(\widetilde{\mathbf{S}}^{1/2}\widetilde{\mathbf{V}}')$ . Temos a matriz  $\mathbf{G}$  com g vetores denominados marcadores de linhas, cada um com n componentes das coordenadas dos genótipos para cada eixo singular selecionado. Da mesma forma,  $\widetilde{\mathbf{H}}'$  terá a marcadores de colunas, cada um com n componentes, correspondendo às coordenadas de ambientes para os mesmos n eixos. Portanto, a representação em biplot, no caso da análise AMMI utilizando a decomposição aproximada por n componentes, pode ser escrito da seguinte forma (GAUCH, 1992; DIAS, KRZANOWSKI, 2006; HONGYU et al., 2014):

$$\widehat{GE}_{(g \times e)} = \widetilde{U}\widetilde{S}\widetilde{V}' = \widetilde{G}\widetilde{H}' = \sum_{k=1}^{n} (\lambda_k^{1/2} \gamma_k) (\lambda_k^{1/2} \alpha_k')$$
(17)

Na análise AMMI, os gráficos Biplot utilizados para o estudo da GEI, construídos em duas dimensões, geralmente apresentam dois tipos de gráficos, o gráfico Biplot1 é apresentada no eixo das abscissas os efeitos principais (médias dos genótipos e dos locais) e no eixo das ordenadas os escores do primeiro eixo de interação (ACP1) e o Biplot2 apresenta somente os efeitos da interação, no eixo das abscissas os escores do primeiro eixo de interação (ACP1) e no eixo das ordenadas os escores do segundo eixo de interação (ACP2).

# 3 MATERIAL E MÉTODOS

#### 3.1 Características dos dados

Foram utilizados dois conjunto de dados MET sobre avaliação de produtividade de milho. O primeiro conjunto de dados provenientes da empresa Criagene SK que fica no município de Casa Branca no estado de São Paulo. Os dados utilizados são relativos a experimentos com 15 genótipos de milho, em 4 locais distribuídos nas localidades e altitudes (o milho é uma planta sensível a altitude e temperaturas noturnas altas), dos municípios de Casa Branca (680m), Guaíra (520m) no estado de São Paulo, nos municípios de Vazante (650m) e (Vazante 720m). Os ensaios foram conduzidos na safrinha de 2011. Têm-se ainda que em cada experimento foi utilizado um delineamento aleatorizado em blocos ao acaso, com 2 blocos em cada experimento.

O segundo conjunto de dados utilizados foram obtidos pelo CIMMYT (INTERNATIONAZ MAIZE AND WHEAT IMPROVEMENT CENTER) em experimentos realizados em 7 países, com 9 genótipos de milho, cada genótipo foi avaliado em 7 ambientes com 4 blocos, caracterizando assim, um delineamento aleatorizado em blocos.

#### 3.2 Análise de variância conjunta

Em muitas áreas da pesquisa agronômica, o pesquisador dificilmente teria condições de conduzir um experimento em um determinado local com o objetivo de tirar conclusões mais amplas dos seus resultados para toda uma região, tendo em vista que as condições climáticas variam muito de um local para outro, afetando, sobremaneira, o comportamento dos seres vivos. Tal fato ocorre, por exemplo, nos experimentos com vegetais, nos quais o solo e as condições climáticas têm influência muito grande (FERREIRA, 2000).

Com o objetivo de verificar se existe a interação entre genótipos e ambiente, realiza-se uma análise de variância conjunta que envolve o estudo de todos os genótipos em todos os ambientes. De acordo com Annichiarico (1992), pode-se assumir que o efeito de genótipos seja fixo e o efeito dos ambientes como aleatório, obtendo o efeito da GEI aleatório. Os dados serão representados pelo seguinte modelo matemático:

$$Y_{ijh} = \mu + g_i + e_j + (ge)_{ij} + b_{h(j)} + \varepsilon_{ijh}$$

$$\tag{18}$$

em que

 $Y_{ijh}$  é a resposta do i-ésimo genótipo no j-ésimo ambiente e no h-ésimo bloco, com i=1

 $1, 2, \ldots, g, j = 1, 2, \ldots, e \in h = 1, 2, \ldots, r;$ 

 $\mu$ : é uma constante, geralmente a média geral;

 $g_i$ : é o efeito do *i*-ésimo genótipo;

 $e_i$ : é o efeito do j-ésimo ambiente;

 $(ge)_{ij}$ : é o efeito da interação do *i*-ésimo genótipo com o *j*-ésimo ambiente;

 $b_{h(j)}$ : é o efeito do h-ésimo bloco dentro do j-ésimo ambiente;

 $\varepsilon_{ijh}$ : é o erro experimental associado ao *i*-ésimo genótipo, no *j*-ésimo ambiente e no *h*-ésimo bloco, assumido ser independente e  $\varepsilon_{ijh} \stackrel{IID}{\sim} N(0, \sigma^2)$ .

O quadro para análise de variância para este modelo é como mostrado na Tabela 2 a seguir:

Tabela 2 - Esquema da análise de variância para experimentos de um mesmo grupo de genótipos avaliado em e locais com r blocos

Fonte de Variação	Graus de Liberdade $(GL)$	Quadrado Médio $(QM)$
Blocos d. ambiente $(Bd.E)$	e(r-1)	$QM_{Bd.E}$
$Gen\'otipos(G)$	(g - 1)	$QM_G$
Ambiente $(E)$	(e-1)	$QM_E$
Interação (G $\times$ E)	(g-1)(e-1)	$QM_{G \times E}$
Resíduos $(Res)$	e(g-1)(r-1)	$QM_{EM}$
Total	(ger-1)	-

O modelo mais simples e comum para a análise estatística de um conjunto de dados, com essa estrutura, é dado por (DUARTE; VENCOVSKY, 1999; GAUCH, 2013; RODRIGUES, et al., 2014):

$$Y_{ij} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ij}$$
(19)

em que

 $Y_{ij}$  é a resposta média do genótipo i no ambiente j; com  $i=1,2,\ldots,g$  e  $j=1,2,\ldots,e$ ;

 $\mu$ : é uma constante, geralmente a média geral;

 $g_i$ : é o efeito fixo do genótipo i;

 $e_i$ : é o efeito fixo do ambiente j;

 $(ge)_{ij}$ : é o efeito da interação específica do genótipo i com o ambiente j;

e  $\varepsilon_{ij}$ : é o erro experimental médio, assumido independente e  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2)$ .

O quadro de ANOVA correspondente tem, portanto, a estrutura mostrada na Tabela 3, assumindo um delineamento experimental inteiramente aleatorizado:

Tabela 3 - Esquema da análise de variância para experimentos de um mesmo grupo de genótipos avaliado em e locais com r repetições

Fonte de Variação	Graus de Liberdade $(GL)$	Quadrado Média $(QM)$
$Gen\'otipos(G)$	(g - 1)	$QM_G$
Ambiente $(E)$	(e - 1)	$QM_E$
Interação (G×E)	(g-1)(e-1)	$QM_{G \times E}$
Erro Médio/r	ge(r-1)	$QM_{EM}$
Total	(ger-1)	-

A análise efetuada com média exige cálculo a parte da estimativa da variância residual média  $(QM_{EM})$ , assumida comum a toda a série de experimentos. Isto é feito obtendo-se a média ponderada dos Quadrados Médios Residuais  $(QM_{Res})$  obtida de todas as ANOVA's individuais de experimentos, onde os graus de liberdade são os pesos, ou seja:

$$QM_{EM} = \frac{\sum_{j} SQ_{Res_j}}{\sum_{j=1} GL_{Res_j}};$$
(20)

 $com j = 1, 2, \dots, e.$ 

### 3.3 Modelo AMMI

Em uma análise conjunta, com dados de avaliação de g genótipos em e ambientes, existem  $g \times e$  efeitos de interação que são estimáveis. Cada um desses efeitos, portanto, é desdobrado nas mencionadas componentes. Uma dada componente, por sua vez, é formada por uma contribuição do genótipo e outra do ambiente, contribuições essas que são multiplicadas entre si, além de uma constante, para gerar a parte da interação explicada pela componente em questão. Essa ideia de quantificar um efeito de interação pela ação conjunta do genótipo e do ambiente, pela multiplicação das respectivas contribuições, é o fundamento básico deste modelo.

Vale atentar para o significado da sigla AMMI (GAUCH, 1992). Ela especifica que a média de uma variável, obtida num genótipo e num certo ambiente, é resultado dos efeitos principais, genéticos e ambientais, que se somam, ou são aditivos (AM: Additive Main effects) no modelo geral. Os efeitos de interação, por outro lado, são resultado da multiplicação entre uma contribuição do genótipo e outra do ambiente (Multiplicative Interaction effects). Gauch e Zobel (1996) informam que essa análise pode ajudar tanto na identificação de genótipos de alta produtividade e largamente adaptados, como na realização do chamado zoneamento agronômico, com fins de recomendação regionalizada e seleção de locais de teste.

Suponha que um conjunto de g genótipos tenha sido testado experimentalmente em e ambientes (locais, anos etc). Assim, as médias de cada combinação de genótipo e ambiente obtidas de r repetições por experimento (um conjunto balanceado de dados), podem ser representadas pela matriz:

$$\mathbf{Y}_{g \times e} = \begin{pmatrix} \bar{Y}_{11} & \bar{Y}_{12} & \cdots & \bar{Y}_{1e} \\ \bar{Y}_{21} & \bar{Y}_{22} & \cdots & \bar{Y}_{2e} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{Y}_{g1} & \bar{Y}_{g2} & \cdots & \bar{Y}_{ge} \end{pmatrix}$$
(21)

Sendo a interação significativa, o próximo passo é fazer a decomposição da  $SQ_{GEI}$ , para descartar um resíduo adicional presente nessa soma de quadrados. Essa decomposição é feita utilizando o fator analítico proposto por Gollob (1968) e Mandel (1961, 1971) e tem a seguinte expressão:

$$(ge)_{ij} = \sum_{k=1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}, \tag{22}$$

em que  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ , e  $\gamma_{ik}$ ,  $\alpha_{jk}$  satisfazem o contraste de ortonormalização  $\sum_i \gamma_{ik} \gamma'_{ik} = \sum_j \alpha_{jk} \alpha'_{jk} = 0 \text{ para } k \neq k' \text{ e } \sum_i \gamma^2_{ik} = \sum_j \alpha^2_{jk} = 1.$ 

O modelo AMMI pressupõe componentes aditivos para os efeitos principais de genótipos  $(g_i)$  e de ambientes  $(e_j)$  e componentes multiplicativos para o efeito da interação  $(ge)_{ij}$ . Assim, a resposta média sobre r repetições ou blocos do i-ésimo genótipo no j-ésimo ambiente é representada por (GAUCH, 1992; GAUCH, ZOBEL, 1996; GAUCH, 2013; RODRIGUES et al., 2014; HONGYU, et al., 2014):

$$Y_{ijr} = \mu + g_i + e_j + \sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + b_{r(j)} + \varepsilon_{ijr}$$
(23)

Os componentes aditivos e multiplicativos do modelo podem ser escritos da seguinte forma (DUARTE; VENCOVSKY, 1999):

$$Y_{ijr} = \overbrace{\mu + g_i + e_j}^{p} + \sum_{k=1}^{p} \lambda_k \gamma_{ik} \alpha_{jk} + b_{r(j)} + \varepsilon_{ijr}$$

parte multiplicativa

 $Y_{ijr} = \mu + g_i + e_j + \sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk} + \sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk} + b_{r(j)} + \varepsilon_{ijr}$ 

em que n < p,  $\rho_{ij} = \sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}$  e n é o número de eixos ou componentes principais selecionados para descrever o  $padr\tilde{a}o$  da GEI;

 $Y_{ijr}$  é a resposta do *i*-ésimo genótipo no *j*-ésimo ambiente e no *r*-ésimo bloco, com  $i=1,2,\ldots,g$ ,  $j=1,2,\ldots,e$  e  $r=1,2,\ldots,h$ ;

 $\mu$  é uma constante, geralmente a média;

 $g_i$  é o efeito do *i*-ésimo genótipo;

 $e_j$  é o efeito do j-ésimo ambiente;

 $\lambda_k$  é a raiz quadrada do k-ésimo autovalor da matriz  $(GE)(GE^T)$  ou  $(GE^T)(GE)$ , com  $k=1,2,\ldots,p$  e  $p=\min\{g-1,e-1\}$ ;

 $\gamma_{ik}$  é o *i*-ésimo elemento do vetor coluna  $\gamma_k$  associado a  $\lambda_k$ ;

 $\alpha_{jk}$  é o j-ésimo elemento do vetor linha  $\alpha_k$  associado a  $\lambda_k$ ;

 $\rho_{ij}$  é o resíduo adicional contendo os dois termos multiplicativos não incluídas no modelo;

 $b_{r(j)}$ : é o efeito do r-ésimo bloco dentro do j-ésimo ambiente;

 $\varepsilon_{ijr}$  é erro experimental associado ao *i*-ésimo genótipo e *j*-ésimo ambiente no *r*-ésimo bloco, assumido ser independente e  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \frac{\sigma^2}{r})$ , sendo *r* o número de repetições ou blocos.

O termo  $(ge)_{ij}$  (interação no modelo tradicional), na metodologia AMMI é representado pela soma de p parcelas, cada uma da multiplicação de  $\lambda_k$ , expresso na mesma unidade de  $\mathbf{Y}_{ij}$ , por um efeito de genótipo  $(\gamma_{ik})$  e um efeito ambiental  $(\alpha_{jk})$ , ambos adimensionais, ou seja,  $\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$  (n: termos da interação). O termo  $\lambda_k$  traz uma informação relativa à variação devida à  $G \times E$ , na k-ésima parcela. De forma que a soma das p parcelas recompõe toda a variação  $(SQ_{GEI} = \sum_{k=1}^{p} \lambda_k^2)$ , e os efeitos  $\gamma_{ik}$  e  $\alpha_{jk}$  representam os pesos do genótipo i e do ambiente j, naquela parcela da interação (GAUCH, 1992; GAUCH; ZOBEL, 1996; GAUCH, 2013).

Pela abordagem AMMI não se busca recuperar toda a  $SQ_{GEI}$ , mas apenas a parcela mais fortemente determinada por genótipo e ambiente (linhas e colunas da matriz GE), ou seja, o padrão (parte determinística ou sistemática). Assim, a interação do genótipo i com o ambiente j ( $\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$ ), descartando-se o resíduo adicional  $\rho_{ij}$  dado por:  $\sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}$ . Aqui, como em ACP, esses eixos captam, sucessivamente, porções cada vez menores da variação presente na matriz GE ( $\lambda_1^2 \ge \lambda_2^2 \ge ... \ge \lambda_p^2$ ). Por isso, o método AMMI é visto como um procedimento capaz de separar padrão e ruido na análise da  $SQ_{GEI}$ :  $\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$  e  $\sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}$ , respectivamente (DUARTE; VENCOVSKY, 1999; GAUCH, 2013).

A matriz GE é matriz de interação entre genótipo e ambiente (matriz de resíduos) obtida pela matriz (21), ou seja, cada elemento  $(ge)_{ij}$  da matriz GE é encontrado pela seguinte relação:

$$(\widehat{ge})_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{.i}$$

em que  $Y_{ij}$  é a média das repetições do genótipo i no ambiente j, com  $i=1,2,\ldots,g$  e  $j=1,2,\ldots,e$ ;  $\overline{Y}_{i}$  é a média do genótipo i;  $\overline{Y}_{.j}$  é a média do amiente j;  $\overline{Y}_{.i}$  é a média geral do experimento.

Uma propriedade da DVS que determina a partição da  $SQ_{GEI}$  na análise AMMI é como segue. Considere-se, a princípio, a decomposição exata da matriz  $GE_{(g\times e)}$  por DVS:

$$\underbrace{GE}_{B} = \underbrace{U}_{(P_{BB'})} \underbrace{S}_{(\Lambda^{1/2})} \underbrace{V'}_{(P_{B'B})} = \sum_{k=1}^{p} \lambda_k \gamma_k \alpha'_k;$$
(24)

com  $p = min\{(g-1), (e-1)\}.$ 

Portanto, a DVS de GE pode ser escrita como uma soma de p matrizes de posto unitário, podemos obter a partição da soma de quadrados dos elementos da matriz GE:

$$\underbrace{\mathbf{GE}}_{\substack{[\mathbf{B}=[b_{ij}]\\(g\times e)}} = \underbrace{\lambda_{1}\gamma_{1}\alpha'_{1}}_{\substack{(q\times e)}} + \underbrace{\lambda_{2}\gamma_{2}\alpha'_{2}}_{\substack{(q\times e)}} + \dots + \underbrace{\lambda_{p}\gamma_{p}\alpha'_{p}}_{\substack{(q\times e)}} \\
\underbrace{\mathbf{B}_{1}=[b_{1ij}]}_{\substack{(g\times e)}} - \underbrace{\mathbf{B}_{2}=[b_{2ij}]}_{\substack{(g\times e)}} - \underbrace{\mathbf{B}_{p}=[bp_{ij}]}_{\substack{(q\times e)}}$$
(25)

$$\sum_{i,j} b_{ij}^2 = \sum_{i,j} (g\hat{e})_{ij}^2 = \sum_{i,j} b1_{ij}^2 + \sum_{i,j} b2_{ij}^2 + \dots + \sum_{i,j} bp_{ij}^2 = SQ_{G \times E}$$
 (26)

$$\sum_{k} \lambda_{k}^{2} = \sum_{i,j} (g\hat{e})_{ij}^{2} = \lambda_{1}^{2} + \lambda_{2}^{2} + \dots + \lambda_{p}^{2} = \sum_{k} SQ_{IPCA_{k}} = SQ_{GEI}$$
 (27)

em que o termo IPCA vem de "Incremental Principal Component Analysis", ou seja,  $IPCA_k$  é o efeito de interação do componente principal associado ao k-ésimo eixo.

O modelo AMMI procura explicar a  $SQ_{GEI}$  por meio de uma aproximação de posto n para a matriz GE (com n bem menor do que p). Ou seja, o objetivo da análise é descrever a GEI por um número reduzido de eixos (n=1 a 3), resultando num modelo informativo, mas que consuma poucos graus de liberdade. Com isso, o modelo é capaz de captar o padrão subjacente à interação e descartar os ruidos presentes nos dados e sem interesse agronômico, os quais só prejudicam a capacidade preditiva do modelo. Essa separação em padrão e ruidos pode ser expressa em termos de soma de quadrados da seguinte forma (DUARTE; VENCOVSKY, 1999; GAUCH, 2013):

$$SQ_{GEI} = \sum_{k=1}^{p} \lambda_k^2 = (\sum_{k=1}^{n} \lambda_k^2) + (\sum_{k=n+1}^{p} \lambda_k^2) = (SQ_{GEI[padr\tilde{a}o]}) + (SQ_{GEI[ruidos]})$$
(28)

O número de termos de interação no modelo ou componentes da ACP, n, com o qual se pretende reduzir a dimensão do sistema e fornecer uma descrição mais parcimoniosa relativo à estrutura da interação, deve ser escolhido da forma mais adequada para que não afete todos os resultados na análise dos dados (GAUCH et al., 2008; YANG et al., 2009; GAUCH, 2013).

Um dos procedimentos usuais adotados para a definição do número de eixos a serem retidos, de maneira a explicar o padrão relacionado à interação, consiste em determinar graus de liberdade associados à parcela da  $SQ_{GEI}$  relacionada a cada membro da família AMMI. Obtém-se, então, o quadrado médio (QM) correspondente a cada parcela (ou modelo). Em seguida é obtido um teste F, avaliando-se a significância de cada componente em relação ao  $QM_{erro}$  médio.

Dessa forma o ponto de parada que determina a seleção do modelo  $(AMMI_0, AMMI_1, \cdots$  ou  $AMMI_n)$  baseia-se na significância do teste F para os sucessivos eixos da interação. O resíduo AMMI, reunindo os eixos descartados da interação, também pode ser testado de maneira a assegurar o seu caráter desprezível. Existem várias técnicas estatísticas para selecionar o número de componentes adequado na decomposição da  $SQ_{GEI}$ . Entre estes métodos destaca-se o teste  $F_r$  proposto por Cornelius (1993) que é considerado um método robusto. A estatística  $F_r$ , sob a hipótese nula de que não haja mais do que n termos significativos para a interação, tem uma distribuição F aproximada com  $f_2 = (g - 1 - n)(e - 1 - n)$  graus de liberdade e os graus de liberdade do resíduo. Sob essa hipótese, o numerador da expressão (30) é aproximadamente uma variável qui-quadrado com  $f_2$  graus de liberdade (PIEPHO, 1995):

$$F_r = \frac{SQ_{GEI} - \sum_{k=1}^n \lambda_k^2}{f_2 QM_{Res}}$$
 (29)

em que,  $\lambda_k$  é a raiz quadrada do k-ésimo autovalor da matriz  $(\mathbf{GE})(\mathbf{GE})^t$  e  $QM_{\mathrm{Res}}$  é o quadrado médio do resíduo. Assim, um resultado significativo pelo teste sugere que pelo menos um termo multiplicativo ainda deve ser adicionado aos n já ajustados. Logo,  $F_r$  pode ser visto como um

teste para a significância dos n+1 primeiros termos da interação.

Yan e Tinker (2006) propôs uma "relação de informação (IR)" para avaliar a adequação de um biplot em exibir os padrões de uma tabela de dupla entrada. Suponha que esta tabela de dupla entrada tem g genótipos e e ambientes. O número máximo de PCs são requeridos para representar completamente esta tabela é  $k = \min(e, g - 1)$ . Se não há correlação entre os ambientes, todos os k PCs devem ser completamente independente e a proporção da variação total explicada por cada PC deve ser exatamente 1/k.

Quando existir alguma correlação entre os ambientes, a proporção da variação explicada pelos primeiros PCs deve ser maior do que 1/k, e a variação explicada por outros PCs deve ser inferior ou igual a 1/k (Yan; Tinker, 2006; Yan, 2011). A IR pode ser calculado para cada um PC, que é a proporção da variação total explicada por cada PC multiplicado por k. A interpretação é a seguinte: um PC com IR > 1 contém padrões (associações entre ambientes), um PC com IR  $\approx 1$  não contém padrões, mas pode conter alguma informação independente, e um PC com IR < 1 não contém qualquer padrão ou informação; o biplot de dimensão 2 representa adequadamente os padrões nos dados, se apenas os dois primeiros PCs tem um IR > 1 (Yan; Tinker, 2006).

### 3.4 Análise AMMI ponderado (W-AMMI)

Quando a variância do erro  $(error\ variance)$  numa tabela de dupla entrada de dados interação genótipo  $\times$  ambiente (Y) é heterogêneo entre os ambientes, os valores da tabela devem ter pesos diferentes para os seus resíduos quadrados no processo de estimação dos parâmetros do modelo. Da forma mais geral, quando a estrutura de variância e covariância do erro para genotípico difere significativamente entre ensaios ou ambientes, por causa da heterogeneidade dos ambientes, então essa heterogeneidade deve ser levada em consideração na análise de GEI (RO-DRIGUES et al., 2014).

Neste trabalho, foi proposta uma generalização do modelo AMMI: AMMIponderado. Este novo modelo é uma solução de problema para os dados de multi-ambientes com heterogeneidade de ambientes.

O modelo AMMI pressupõe componentes aditivos para os efeitos principais de genótipos  $(g_i)$  e de ambientes  $(e_j)$  e componentes multiplicativos para o efeito da interação  $(ge)_{ij}$ . Assim, a resposta média sobre r repetições ou blocos do i-ésimo genótipo no j-ésimo ambiente é representada por (GAUCH, 1992; GAUCH, ZOBEL, 1996; GAUCH, 2013; RODRIGUES et al., 2014; HONGYU, et al., 2014):

$$\frac{\overline{Y}_{ijr}}{s_j^2} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + b_{r(j)} + \varepsilon_{ijr}$$
(30)

em que  $\overline{Y}_{ijr}$  é a resposta do *i*-ésimo genótipo no *j*-ésimo ambiente e no *r*-ésimo bloco, com  $i=1,2,\ldots,g,\ j=1,2,\ldots,e$  e  $r=1,2,\ldots,h;$ 

 $s_j^2$  é a variância do erro experimental do ambiente j ou Quadrado médio do resíduo do ambiente j, obtida pela análise da variância conjunta da expressão (18);

 $\mu$  é uma constante, geralmente a média;

 $g_i$  é o efeito do *i*-ésimo genótipo;

 $e_j$  é o efeito do j-ésimo ambiente;

 $\lambda_k$  é a raiz quadrada do k-ésimo autovalor da matriz  $(GE)(GE^T)$  ou  $(GE^T)(GE)$ , com  $k = 1, 2, \dots, p$  e  $p = \min\{g - 1, e - 1\}$ ;

 $\gamma_{ik}$  é o *i*-ésimo elemento do vetor coluna  $\gamma_k$  associado a  $\lambda_k$ ;

 $\alpha_{jk}$  é o j-ésimo elemento do vetor linha  $\alpha_k$  associado a  $\lambda_k$ ;

 $\rho_{ij}$  é o resíduo adicional contendo os dois termos multiplicativos não incluídos no modelo;

 $b_{r(j)}$ : é o efeito do r-ésimo bloco dentro do j-ésimo ambiente;

 $\varepsilon_{ijr}$  é erro experimental associado ao *i*-ésimo genótipo e *j*-ésimo ambiente no *r*-ésimo bloco, assumido ser independente e  $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \frac{\sigma^2}{r})$ , sendo *r* o é o número de repetições ou blocos.

## 3.5 Modelo SREG

O conceito SREG (Sites Regression), origina-se da análise de ensaios multiambientais (MultiEnvironment Trials - MET) no melhoramento de plantas para testar adaptação geral e específica de cultivares. A variação fenotípica observada ou desempenho de um cultivar (P) consiste do efeito principal do genótipo (G), do efeito principal do ambiente (E) e efeitos ambientais nos genótipos ou GEI (YAN; KANG, 2003; YAN; HOLLAND, 2010),

$$P = G + E + GE \quad \Rightarrow \quad P - E = G + GE \tag{31}$$

O modelo SREG utiliza o gráfico biplot GGE (Genotype-Genotype Environment), o qual foi proposto para explorar as respostas dos genótipos a ambientes específicos e é semelhante ao modelo AMMI. A diferença no modelo SREG, é que os efeitos principais dos genótipos são considerados junto com o efeito da GEI e no modelo AMMI são estimados como efeitos aditivos. A equação (31) está em termos de componentes de variância, quando os efeitos são medidos, pode ser escrito como (YAN; KANG, 2003):

$$\overline{Y}_{ij} = \mu + G_i + E_j + (GE)_{ij}$$
ou
$$\overline{Y}_{ij} - \mu - E_j = G_i + (GE)_{ij}$$

em que,  $\overline{Y}_{ij}$  é a resposta média do *i*-ésimo genótipo e *j*-ésimo ambiente,  $\mu$  é a média geral,  $G_i$  é o efeito principal do *i*-ésimo genótipo com  $i = (1, \dots, g)$ ,  $E_j$  é o efeito principal do *j*-ésimo ambiente com  $j = (1, \dots, e)$  e  $(GE)_{ij}$  é o efeito da interação do *i*-ésimo genótipo no *j*-ésimo ambiente. Podemos reescrever GE em dois termos multiplicativos, temos:

$$\overline{Y}_{ij} - \mu - E_j = g_{i1}e_{1j} + g_{i2}e_{2j} + \varepsilon_{ij}$$
(32)

em que,  $g_{i1}$  e  $e_{1j}$  são denominados escores principais do genótipo i e ambiente j, respectivamente;  $g_{i2}$  e  $e_{2j}$  são os escores secundários para genótipo i e ambiente j, respectivamente e  $\varepsilon_{ij}$  é o resíduo não explicado por ambos os efeitos. Aplicando a técnica de decomposição em valores singulares (DVS) na equação (32), tem-se (YAN; KANG, 2003):

$$\overline{Y}_{ij} - \mu - E_j = \lambda_1 \mathbf{u}_{i1} \mathbf{v}_{j1} + \lambda_2 \mathbf{u}_{i2} \mathbf{v}_{j2} + \varepsilon_{ij}$$
(33)

em que  $\lambda_1$  e  $\lambda_2$  são os maiores autovalores do primeiro e segundo componentes principais (ACP1 e ACP2), respectivamente;  $\boldsymbol{u}_{i1}$  e  $\boldsymbol{u}_{i2}$  são os autovetores do i-ésimo genótipo para ACP1 e ACP2, respectivamente e  $\boldsymbol{v}_{j1}$  e  $\boldsymbol{v}_{j2}$  são os autovetores do j-ésimo ambiente para ACP1 e ACP2, respectivamente.

### 3.6 GGE biplot

O modelo GGE biplot (genotype main effects + genotype environment interaction) (YAN; KANG, 2003; YAN; TINKER, 2006; YAN; HOLLAND, 2010; YAN, 2011), que considera o efeito principal de genótipo mais a GEI, é baseadas em gráficos biplot, representa graficamente uma matriz de dados. Este Biplot é construído nos dois primeiros componentes principais de uma Análise de Componentes Principais (ACP) utilizando Modelos de Regressões Locais (SREG). A primeira componente, quando se encontra altamente correlacionada com o efeito principal do genótipo, representa a proporção do rendimento que se deve somente às características do genótipo. A segunda componente representa a parte da variável devida a GEI (YAN; HOLLAND, 2010; YAN, 2011). Neste gráfico pode-se visualizar três aspectos importantes (YAN; KANG, 2003):

a) a relação G × E, permitindo o agrupamento de genótipos e ambientes com comportamentos

semelhantes, mostrando o genótipo com maior potencial e sua identificação em cada subgrupo de ambientes (mega-ambientes);

- b) a inter-relação entre ambientes, facilitando a identificação do melhor ambiente na avaliação das cultivares e indicando qual ambiente pode ser menos favorável;
- c) a inter-relação entre genótipos, facilitando a comparação dos genótipos e do ordenamento para os parâmetros de rendimento e estabilidade.

Quando diferentes cultivares estão adaptadas as diferentes grupos de ambientes e a variação entre grupos é maior do que dentro do grupo, tem-se a formação de um mega-ambiente (YAN; KANG, 2003). A definição de mega-ambientes e a relação entre os ambientes auxiliam aos melhoristas de plantas na identificação de genótipos que possuam adaptação ampla ou específica a determinados ambientes ou grupos de ambientes (TRETHOWAN, et al., 2001; SILVA; BENIN, 2012). Na análise GGE biplot, quando se estuda mega-ambientes, a média no gráfico não está relacionada à média geral, mas sim à média do mega-ambiente e esta abordagem auxilia na identificação de genótipos que possuam adaptação ampla ou específica a determinados ambientes ou grupos de ambientes (TRETHOWAN et al., 2001; YAN; KANG, 2003; YAN; TINKER, 2006).

No modelo GGE biplot, o cosseno do ângulo entre dois ambientes corresponde à correlação genética entre eles. Outros tipos de biplots não apresentam essa propriedade, o que torna essa metodologia mais eficiente em relação a outras técnicas baseadas em biplot (YAN et al., 2007; YAN, 2011). Tal metodologia pode ser descrita da seguinte forma (YAN; KANG, 2003; YAN, 2011):

$$Y_{ij} = \mu + G_i + E_j + GE_{ij} \tag{34}$$

em que,  $Y_{ij}$  é o desempenho esperado do genótipo i no ambiente j;  $\mu$  é a média geral das observações;  $G_i$  é o efeito principal do genótipo i;  $E_j$  é o efeito principal do ambiente j e  $GE_{ij}$  é a interação entre o genótipo i e o ambiente j.

Neste modelo GGE biplot, variação fenotípica é resultante do efeito genotípico  $(G_i)$ , do ambiente  $(E_j)$  e da interação entre genótipo e ambiente  $(GE_{ij})$  (YAN; KANG, 2003). Portanto, apenas o efeito principal de genótipo e da GEI são os mais importantes (YAN et al., 2007; YAN; HOLLAND, 2010). Considera-se que o efeito principal do ambiente não é relevante na seleção de cultivares. Isso se deve ao fato de que, na seleção de cultivares e formação de mega-ambientes, a capacidade adaptativa dos genótipos é mais importante em relação às condições agroclimáticas (CAMARGO-BUITRAGO; INTIRE; GORDÓN-MENDOZA, 2011; SILVA; BENIN, 2012).

O modelo GGE biplot não separa os efeitos do genótipo e da GEI, mantendo-os juntos em dois termos multiplicativos, que podem ser visualizados na equação (34) da metodologia SREG. Assim sendo, a construção do gráfico biplot no modelo GGE se dá por meio da dispersão

simples de  $g_{i1}$  e  $g_{i2}$  para genótipos,  $e_{1j}$  e  $e_{2j}$  para ambientes. A equação (34), refere-se ao modelo para ambiente centrado, pois  $(\mu + E_j)$  são subtraídos de cada um dos valores observados  $Y_{ij}$ . Existe dois problemas em utilizar este modelo (YAN et al., 2001; YAN; KANG, 2003):

- i) Este modelo só pode ser utilizado para analisar dados MET (multi-ambientes), que tenham uma única característica, por exemplo, produção é uma característica que está em todos os ambientes;
- ii) No conjunto de dados, cujos ambientes são heterogêneos.

Para solucionar estes problemas, Yan e Kang (2003) propuseram duas transformações. A primeira é a transformação de dados para dimensionar o ambiente centralizado, tem-se:

$$\frac{\overline{Y}_{ij} - \mu - E_j}{d_j} = g_{i1}e_{1j} + g_{i2}e_{2j} + \varepsilon_{ij}$$
(35)

em que  $d_j$  é o desvio fenotípico padronizado do ambiente j:

$$d_{j} = \sqrt{\frac{1}{g-1} \sum_{i=1}^{g} (Y_{ij} - \overline{Y}_{.j})^{2}}$$
(36)

em que,  $Y_{ij}$  é a resposta do genótipo i e ambiente j,  $\overline{Y}_{.j}$  é a média do ambiente j e g é o número de genótipos testados. Para dados MET que tenham uma única característica, este modelo pressupõe que todos os ambientes tem a mesma importância. Opõe-se, portanto, a possibilidade de detectar diferenças entre os ambientes em sua capacidade de discriminação, um aspecto importante para a avaliação do teste-ambiente.

A segunda transformação de dados é escalar os dados centrados no ambiente, com o erro padrão dentro do ambiente  $(s_i)$ :(YAN; KANG, 2003):

$$\frac{\overline{Y}_{ij} - \mu - E_j}{s_j} = g_{i1}e_{1j} + g_{i2}e_{2j} + \varepsilon_{ij}$$
(37)

em que  $s_j$  é o erro padrão do ambiente j:

$$s_{j} = \sqrt{\frac{1}{m(r-1)} \sum_{i=1}^{m} \sum_{k=1}^{r} (Y_{ijk} - \overline{Y}_{ij.})^{2}}$$
(38)

em que,  $\overline{Y}_{ij}$  é a média do genótipo i no ambiente j,  $Y_{ijk}$  é a resposta do genótipo i e ambiente j na k-ésima repetição e r é o número de repetições (k = 1, ..., r) no ambiente j.

A metodologia SREG utilizada na construção do GGE biplot tem sido reconhecida como uma metodologia revolucionária na abordagem com análise gráfica bidimensional para melhoramento e pesquisa genética, incluindo, também, análise de estabilidade e adaptabilidade na

recomendação de cultivares, destacando-se das demais técnicas pela sua alta eficiência estatística na explicação da GEI (YAN, et al., 2000; YAN; KANG, 2003; YAN; HOLLAND, 2010; YAN, 2011).

### 3.7 DVS em GGE biplot

A técnica de decomposição por valores sigulares (DVS) é importante na análise biplot, pois permite computar os componentes principais e também, fornece uma medida da variabilidade capturada por cada um dos componentes (YAN; KANG, 2003). A DVS é dada por meio de:  $\mathbf{P}_{ge} = \mathbf{U}_{g,r}\mathbf{S}_{r,r}\mathbf{V}_{e,r}^T$  [ $\mathbf{r} \leq \min(\mathbf{g},\mathbf{e})$ ], em que a matriz  $\mathbf{P}_{ge}$  é uma matriz de interação (G+GEI) composta pela matriz  $\mathbf{U}$  que tem g linhas e r colunas, caracterizando g genótipos, matriz  $\mathbf{V}$  que tem r linhas e e colunas, que caracterizam os ambientes e a matriz  $\mathbf{S}$  é a matriz diagonal contendo r valores singulares (YAN; KANG, 2003; YAN; TINKER, 2006). Em geral, a DVS decompõe os valores da matriz ou tabela de dados em r componentes principais, cada um contendo o vetor genótipo ( $u_i$ ), o vetor ambiente ( $v_i$ ) e os valores singulares ( $\lambda$ ) que compõem a seguinte equação:

$$P_{ij} = \sum_{k=1}^{r} \boldsymbol{u}_{ik} \lambda_k \boldsymbol{v}_{kj}$$

em que  $\lambda_k \geq \lambda_{k+1}$ , r é a classificação na tabela de dupla entrada (genótipos e ambientes), ou seja, o número de componentes principais necessários para representar P, com  $r \leq \min(g,e)$ . Se r < g é indicativo de que há associações entre os genótipos e se r < e há associações entre os ambientes. Os autovetores PCA1 e PCA2 não podem ser computados diretamente antes dos valores singulares serem particionados em genótipos e ambientes (YAN; KANG, 2003; SILVA; BENIN, 2012). A partição é dada a seguir:

$$P_{ij} = \sum_{k=1}^{r} \boldsymbol{u}_{ik} \boldsymbol{v}_{ij} = \sum_{k=1}^{r} (\boldsymbol{u}_{ik} \lambda_k^s) (\lambda_k^{1-s} \boldsymbol{v}_{kj})$$

em que, s é o fator de partição, que pode ser qualquer valor entre 0 e 1.

Pode ser apresentada graficamente em um biplot de dimensão 2 após uma particionamento de valor singular (PVS) apropriado.

$$P_{ij} = (\lambda_1^s \boldsymbol{u}_{i1}) \left(\lambda_1^{1-s} \boldsymbol{v}_{j1}\right) + (\lambda_2^s \boldsymbol{u}_{i2}) \left(\lambda_2^{1-s} \boldsymbol{v}_{j2}\right) + \varepsilon_{ij}$$

em que s = [0; 0,5; 1] é o fator de partição de valor singular (PVS). O biplot é construído por meio da representação gráfica  $\lambda_1^s \boldsymbol{u}_{i1}$  como abcissa,  $\lambda_2^s \boldsymbol{u}_{i2}$  como ordenada para cada genótipo, e ao mesmo tempo traçando  $\lambda_1^{1-s} \boldsymbol{v}_{j1}$  como abcissa e  $\lambda_2^{1-s} \boldsymbol{v}_{j2}$  como ordenada para cada ambiente (YAN; KANG, 2003; YAN, 2011).

O expoente s é usado para redimensionar os escores de linhas e colunas para melhorar a interpretação visual do biplot para um propósito particular. No contexto de dados MET, os valores singulares são alocados inteiramente nos escores de genótipo (linha) se s=1, isto é "particionamento em valor singular centrada-genótipo" ou PVS = 1 (YAN, 2002), ou inteiramente aos escores de ambiente (coluna) se s=0 ("particionamento em valor singular centrada-ambiente" ou PVS = 2) e s=0,5 alocará as raízes quadradas dos valores singulares  $\lambda_k$  para ambos os escores de genótipo e ambiente ("particionamento em valor singular simétrico" ou PVS = 3). Portanto, diferentes valores de s são capazes de alterar a visualização das inter-relações entre genótipos e ambientes (YAN; KANG, 2003; YAN; HOLLAND, 2010).

Na análise biplot GGE, o PVS de genótipo-centrada e o ambiente-centrada são utilizados para avaliação de genótipos e ambiente de teste, respectivamente (YAN, 2011). Uma propriedade importante do biplot é que a aproximação de qualquer elemento da matriz original **P** de posto 2 pode ser estimada visualmente por produto interno de vetores de genótipo e ambiente, respectivamente, e o cosseno do ângulo entre si. Isto é conhecido como a propriedade do produto interno do biplot (YAN; KANG, 2003; YAN; HOLLAND, 2010; YAN, 2011).

Yan e Tinker (2006) propôs uma "relação de informação (IR)" para avaliar a adequação de um biplot em exibir os padrões de uma tabela de dupla entrada. Suponha que esta tabela de dupla entrada tem g genótipos e e ambientes. O número máximo de PCs são requeridos para representar completamente esta tabela é  $k = \min (e, g - 1)$ . Se não há correlação entre os ambientes, todos os k PCs devem ser completamente independente e a proporção da variação total explicada por cada PC deve ser exatamente 1/k.

Quando existir alguma correlação entre os ambientes, a proporção da variação explicada pelos primeiros PCs deve ser maior do que 1/k, e a variação explicada por outros PCs deve ser inferior ou igual a 1/k (Yan; Tinker, 2006; Yan, 2011). A IR pode ser calculada para cada um PC, que é a proporção da variação total explicada por cada PC multiplicado por k. A interpretação é a seguinte: um PC com IR > 1 contém padrões (associações entre ambientes), um PC com IR  $\approx 1$  não contém padrões, mas pode conter alguma informação independente, e um PC com IR < 1 não contém qualquer padrão ou informação; o biplot de dimensão 2 representa adequadamente os padrões nos dados, se apenas os dois primeiros PCs tem um IR > 1 (Yan; Tinker, 2006).

### 3.8 W-GGE Biplot (GGE biplot ponderado)

Existem dois problemas no modelo GGE biplot, em um desses problemas é quando na análise de dados multi-ambientes (MET), os ambientes são heterogêneos (YAN et al., 2000;

YAN; KANG, 2003). Para solucionar estes problemas, neste trabalho desenvolvemos uma generalização do modelo GGE biplot: W-GGE biplot (Weighted Genotype main effects + Genotype environment interaction).

O W-GGE biplot, ou pode ser chamado de GGE bipot ponderado, é uma genelarização do GGE biplot para análise de dados MET, que permite ter consideração a heterogeneidade da variância do erro ao longo dos ambientes. De forma geral, a estrutura de variância e covariância do erro para genótipo significa que difere entre ensaios ou ambientes por causa da heterogeneidade dos ambientes, logo essa heterogeneidade deve ser levada em consideração na análise de GEI (RODRIGUES et al., 2014). Para solucionar este problema de heterogeneidade, propusemos uma transformação de dados para dimensionar o ambiente centralizado, tem-se:

$$\frac{\overline{Y}_{ij} - \mu - E_j}{s_j^2} = g_{i1}e_{1j} + g_{i2}e_{2j} + \varepsilon_{ij}$$
(39)

em que  $s_j^2$  é a variância do erro experimental do ambiente j ou Quadrado médio do resíduo do ambiente j, pelo modelo abaixo da análise da variância conjunta:

$$Y_{ijh} = \mu + g_i + e_j + (ge)_{ij} + b_{h(j)} + \varepsilon_{ijh}$$

$$\tag{40}$$

em que

 $Y_{ijh}$  é a resposta do *i*-ésimo genótipo no *j*-ésimo ambiente e no *h*-ésimo bloco, com  $i=1,2,\ldots,g$ ,  $j=1,2,\ldots,e$  e  $h=1,2,\ldots,r$ ;

 $\mu$ : é uma constante, geralmente a média geral;

 $g_i$ : é o efeito do *i*-ésimo genótipo;

 $e_j$ : é o efeito do j-ésimo ambiente;

 $(ge)_{ij}$ : é o efeito da interação do *i*-ésimo genótipo com o *j*-ésimo ambiente;

 $b_{h(j)}$ : é o efeito do h-ésimo bloco dentro do j-ésimo ambiente;

 $\varepsilon_{ijh}$ : é o erro experimental associado ao *i*-ésimo genótipo, no *j*-ésimo ambiente e no *h*-ésimo bloco, assumido ser independente e  $\varepsilon_{ijh} \stackrel{IID}{\sim} N(0, \sigma^2)$ .

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise AMMI e Biplot para 1º conjunto dados multi-ambientes (MET)

### 4.1.1 Análise de variância conjunta para dois fatores

A análise de variância conjunta de dados MET de produtividade de milho, provenientes da empresa Criagene SK, sediada no município de Casa Branca no estado de São Paulo, considerou 15 genótipos em 4 ambientes com 2 blocos e é apresentada na Tabela 4. Os resultados da análise de variância conjunta relativa à produtividade de milho  $(kg\ ha^{-1})$  mostraram que os efeitos de genótipos (G), ambientes (E) e GEI são significativos (p < 0,01), como era de se esperar, pois o milho é uma planta sensível a altitude e temperatura. Esses resultados indicam que os genótipos apresentaram comportamento diferenciado nos ambientes, o que justifica estudo mais aprofundado sobre o comportamento dos genótipos para identificar suas magnitudes de interação com os ambientes. Um dos principais resultados de interesse apresentado na tabela é a soma de quadrados da interação ( $SQ_{GEI}=109478996$ ), que representa 15,75% da variabilidade total dos dados.

Tabela 4 - Análise de variância conjunta para um conjunto de dados com 15 genótipos avaliados em 4 ambientes com 2 blocos

Fonte de Variação	GL	SQ	QM	F	Valor-p
Blocos/Ambiente	4	1555492	388873	0,4721	0,756
Ambiente (E)	3	304154907	101384969	260,7149	4,846e-05 ***
Genótipo (G)	14	124525440	8894674	10,7984	2,712e-11 ***
GEI	42	109478996	2606643	3,1645	3,370e-05 ***
Resíduo	56	46127517	823706	-	-
Total	119	695321348	-	-	-

NS: Não-significativo; \*\*Significativo a 0,01 de probabilidade.

O coeficiente de variação (CV%) do experimento pela análise de variância conjunta mostrou-se de baixa magnitude (8,72%), permitindo inferir que os experimentos foram bem conduzidos e não ocorrem muita influência de fatores não controlados. As estimativas das médias dos genótipos e dos ambientes, com relação a produtividade em  $kg \ ha^{-1}$ , são apresentados na Tabela 5. Observa-se que o genótipo G2 apresentou a maior produtividade, seguidos dos genótipos G15,

<sup>\*\*\*</sup>Significativo a 0,001 de probabilidade.

G14, G5 e G3, respectivamente. Já os genótipos G11, G9 e G12 apresentaram as menores produtividades médias. Com relação aos ambientes, têm-se que o ambiente A2 (município de Casa Branca/SP 680m) apresentou a maior produtividade média, seguido o ambiente A1 (município de Vazante/MG 720m) e as menores produtividades médias foram os ambientes A4 (município de Vazante/MG 650m) e A3 (município de Guaíra/SP 520m), respectivamente.

Tabela 5 - Médias dos genótipos, ambientes e posição das médias em relação a produtividade

Genótipo	Média	Posição	Ambiente	Média	Posição
$G_1$	9593,750	12	$A_1$	11461,167	2
$G_2$	11695,875	1	$A_2$	12347,867	1
$G_3$	11178,750	5	$A_3$	9539,733	3
$G_4$	10466,000	7	$A_4$	8283,767	4
$G_5$	11572,250	4			
$G_6$	10071,250	10			
$G_7$	11074,750	6			
$G_8$	9790,000	11			
$G_9$	8810,375	14			
$G_{10}$	10377,750	8			
$G_{11}$	8508,875	15			
$G_{12}$	9376,625	13			
$G_{13}$	10294,000	9			
$G_{14}$	11646,750	3			
$G_{15}$	11665,000	2			

Considerando a expressão (22) aplicada à matriz de médias da expressão (21), obtém-se a matriz de interação GE, apresentada na Tabela 6. A próxima etapa da análise AMMI corresponde ao ajuste da interação pela decomposição em valores singulares ( $GE = USV^T$ ) aplicado na matriz de interação GE. Esta matriz tem posto igual p = min(15, 3) = 3, portanto a  $SQ_{GEI}$  pode ser decomposta em até 3 componentes principais. A decomposição da matriz GE é dada da seguinte forma (Tabela 6):

Tabela 6 - Valores estimados da interação dupla de 15 genótipos e 4 ambientes para a produção em  $kq\ ha^{-1}$ 

Ambiente/	A1	A2	A3	A4
Genótipo				
G1	1277,22	975,02	-178,85	-2073,38
G2	-348,41	641,89	-503,48	209,99
G3	-96,28	$389,\!52$	219,15	-512,38
G4	-1543,53	-250,73	1435,40	358,87
G5	-516,78	775,02	517,65	-775,88
G6	34,72	2324,02	-1606,85	-751,88
G7	530,72	502,02	-698,85	-333,88
G8	-474,03	-2134,73	590,90	2017,87
G9	154,59	-420,61	351,02	-85,01
G10	98,22	-677,98	-290,35	870,12
G11	$1676,\!59$	343,39	-1066,48	-953,51
G12	184,84	-1093,36	-1111,73	2020,24
G13	1101,97	-256,23	-74,60	-771,13
G14	-1289,78	-357,98	1213,15	434,62
G15	-790,03	-759,23	1203,90	345,37

A matriz U contêm os vetores singulares à esquerda da interação, S é uma matriz diagonal contendo os valores singulares e V é uma matriz cujas colunas contêm os vetores singulares à direita da interação e são apresentadas da seguinte forma:

$$\mathbf{S} = \left(\begin{array}{ccc} 5844, 298 & 0 & 0\\ 0 & 3671, 743 & 0\\ 0 & 0 & 2664, 954 \end{array}\right)$$

$$V = \begin{pmatrix} -0,392 & -0,533 & -0,558 \\ -0,582 & 0,309 & 0,562 \\ 0,359 & 0,658 & -0,433 \\ 0,615 & -0,433 & 0,429 \end{pmatrix}$$

$$U = \begin{pmatrix} -0,412 & 0,109 & -0,367 \\ -0,049 & -0,010 & 0,324 \\ -0,073 & 0,146 & -0,016 \\ 0,255 & 0,418 & 0,095 \\ -0,092 & 0,324 & 0,062 \\ -0,412 & -0,009 & 0,623 \\ -0,164 & -0,121 & 0,054 \\ 0,493 & -0,243 & -0,121 \\ 0,044 & 0,015 & -0,192 \\ 0,135 & -0,226 & 0,024 \\ -0,313 & -0,293 & -0,259 \\ 0,241 & -0,556 & 0,237 \\ -0,134 & -0,104 & -0,397 \\ 0,243 & 0,323 & 0,068 \\ 0,239 & 0,226 & -0,135 \end{pmatrix}$$

Na análise AMMI, a soma de quadrados da interação  $(SQ_{GEI})$  corresponde exatamente à soma dos autovalores ao quadrado dos valores singulares que estão na diagonal da matriz  $S\left(\sum_{k=1}^{n}\lambda_{k}^{2}\right)$  (Tabela 7), que pode estar inflacionado devido à presença de ruídos (variação não explicável) na variável resposta. Em virtude disto, faz-se necessário realizar um ajuste da interação por decomposição por valores singulares, aplicada à matriz de GEI. Esta matriz, por sua vez tem posto p=3, ou seja, pode ser decomposta em três componentes principais, que são a soma de quadrados parciais na análise de variância.

Observando a decomposição da GEI dos experimentos por meio do modelo AMMI (Tabela 7), constatou-se que apenas o primeiro componente principal (PC1) foi significativo com (p<0,01) pelo teste F de Gollob (1968) e explicou 62,4% da variação da  $SQ_{GEI}$ , que é considerada uma resposta padrão presente na  $SQ_{GEI}$  com 16 graus de liberdade. As não-significâncias para o PC2 e PC3 são desprezíveis e contém apenas ruído (variação aleatória não relacionada com o fenômeno da interação), que podem diminuir a eficiência da interpretação da estabilidade dos genótipos e ambientes na análise gráfica. Assim, a interpretação gráfica, considerando apenas a variação contida no primeiro eixo da componente principal foi suficiente para avaliar a estabilidade dos genótipos e ambientes.

Tabela 7 - Porcentagem da soma de quadrados da GEI captada por componente principal (PC)

Componente	Proporção	Proporção	$_{ m SQ}$	QM	F	Valor-p
Principal	$\mathrm{SQ}_{PC}/SQ_{GEI}$	Acumulada(%)				
PC1	0,624	62,4	68311642	4269478	5,18	<0,0001 ***
PC2	0,246	87,0	26963397	1925957	2,32	0,0126
PC3	0,130	100,0	14203957	1183663	1,44	0,1758
Total	1,00	-	109478996	-	-	-

NS: Não-significativo; \*\*\*Significativo a 0,001 de probabilidade.

Tabela 8 - Coordenadas dos eixos das abscissas (x) e de coordenadas (y), para os dois gráficos Biplot

Genótipos	Biplo	t AMMI1	Biplo	t AMMI2
Ambientes	Média $(x)$	PC1(y)	PC1(x)	PC2(y)
G1	9593,750	-31,488	-31,488	-6,594
G2	11695,875	-3,784	-3,784	0,631
G3	11178,750	-5,562	-5,562	-8,871
G4	10466,000	19,465	19,465	-25,331
G5	$11572,\!250$	-7,056	-7,056	-19,659
G6	$10071,\!250$	-31,488	-31,488	0,545
G7	11074,750	-12,519	-12,519	7,317
G8	9790,000	37,696	37,696	14,699
G9	8810,375	3,379	3,379	-0,916
G10	10377,750	10,290	10,290	13,686
G11	8508,875	-23,898	-23,898	17,776
G12	9376,625	18,392	18,392	33,699
G13	10294,000	-10,250	-10,250	6,305
G14	11646,750	18,544	18,544	-19,597
G15	11665,000	18,277	18,277	-13,692
A1	11461,167	-29,971	-29,971	32,323
A2	$12347,\!867$	-44,519	-44,519	-18,695
A3	9539,733	27,502	27,502	-39,864
A4	8283,767	46,987	46,987	26,236

Os resultados para o desdobramento da GEI, correspondente a cada membro da família de modelos AMMI, ou seja, os modelos AMMI0, sendo toda a GEI, com 42 GL; modelo AMMI1 é o restante da interação, depois de retirados os 16 GL e a soma de quadrados atribuída ao primeiro eixo do componente principal (PC1); modelo AMMI2 com 14 GL, a soma de quadrados atribuída ao segundo eixo do componente principal (PC2) e modelo AMMI3, com 12 GL e a soma de quadrados atribuída ao terceiro eixo do componente principal (PC3) (Tabela 7 e Tabela 8).

Avaliando a família dos modelos AMMI (AMMIO, AMMII, AMMI2 e AMMI3) verificou-se pelo teste F, com os graus de liberdade ajustados pelo método de Gollob (1968) significância para o primeiro componente principal da interação, o que levaria à seleção do modelo AMMII para explicar a interação de forma significativa e este apresentou o melhor padrão de resposta para os genótipos aos diferentes ambientes (Tabela 7). Nesse caso, seria possível a construção de dois gráficos biplot: i) AMMII - Média da produtividade dos genótipos vs PC1 (Figura 2); ii) AMM2 - PC1 vs PC2 (Figura 3).

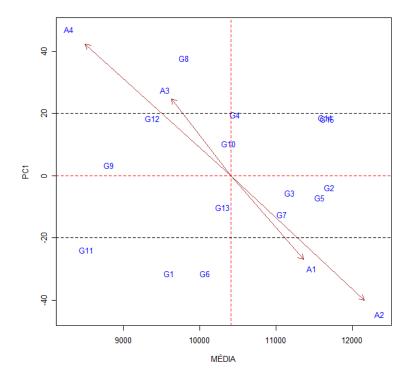


Figura 2 - Biplot AMMI1 (Médias vs PC1) para os dados de produtividade de milho  $(kg \ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)

Os gráficos (Figura 2 e Figura 3) biplots (AMMI1 e AMMI2) foram utilizados para analisar a dispersão da variação dos genótipos e ambientes e da interação entre estes. O biplot AMMI1 contém a variação dos efeitos aditivos principais de genótipos e ambientes, visualizado

na horizontal do gráfico (eixo das abscissas), e a variação dos efeitos multiplicativos da GEI, visualizado na vertical (eixo das ordenadas). O biplot AMMI2 é visualizado apenas os efeitos multiplicativos da GEI contidos nos dois primeiros PCs e dispersos em ambas as ordenadas do gráfico. A interpretação do gráfico biplot permitiu avaliar a estabilidade dos genótipos nos ambientes por meio da observação da magnitude e do sinal dos escores dos genótipos e ambientes para o eixo da interação (Tabela 8). Portanto, os escores baixos (próximos de zero) representam os genótipos e ambientes mais estáveis (GAUCH, 2013).

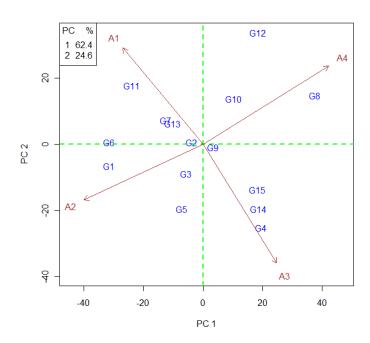


Figura 3 - Biplot AMMI2 (PC1 vs PC2) para os dados de produtividade de milho  $(kg \ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)

Os genótipos que apresentaram as produtividades acima da média foram G2, G15, G14, G5, G3, G7 e G4 em ordem decrescente (Figura 2; Tabela 8) e os mais estáveis foram G2, G9, G5, e G3 (Figura 2; Figura 3 e Tabela 3), estes podem ser cultivados em todos os locais onde foram feitos os estudos. Entre estes, apenas os genótipos G2, G5 e G3 apresentaram produtividade acima da média geral e estão entre os dois grupos (produtivos e estáveis), indicando que esses genótipos associaram adaptabilidade com estabilidade. O G2 apresentou o melhor desempenho entre todos os genótipos, em relação de estabilidade e adaptabilidade (Figura 2 e Figura 3), portanto pode ser escolhido para recomendação pelos pesquisadores.

Os ambientes mais favoráveis para a produtividade de milhos foram A2 (Casa Branca/SP 680m) em seguida A1 (Varante/MG 720m) e o pior ambiente para produtividade foi

A4 (Varante/MG 650m). E também há formação de dois grupos de ambientes: Grupo 1 formado pelos locais A4 e A3 e outro por A1 e A2, sendo que estes grupos apresentaram características opostas, pois os vetores dos ambientes A4 e A3 apontam para a mesma direção, o mesmo para A1 e A2, mas entre o grupo A4 e A3, os vetores têm sentidos opostos do grupo A1 e A2 (Figura 2).

Em relação às adaptações específicas dos genótipos em ambientes, verifica-se que os genótipos G4, G14 e G15 foram mais adaptados ao ambiente A3, G8 e G10 ao ambiente A4, G11, G7 e G13 ao ambiente A1, G1 e G3 ao ambiente A2 (Figura 3). Observa-se que os vetores relativos a esses genótipos apontam para uma direção comum, que pode ser orientada pelos ambientes. Além disso, constata-se falta de adaptação dos genótipos G4, G14 e G15 ao ambiente A1, dos genótipos G8 e G12 ao ambiente A2, do genótipo G1 ao ambiente A4 e G11 ao A3 por causa dos vetores de genótipos e ambientes apontam em direções opostos (Figura 3).

#### 4.2 GGE biplot para análise de 1º conjunto de dados MET

Com a técnica da análise de GGE biplot pode ser alcançado os quatro principais objetivos para os dados multi-ambientes (MET): investigar o mega-ambiente para a compreensão do ambiente de destino; avaliar os genótipos, ambiente de teste dentro de cada mega-ambiente e compreender as causas da GEI. Tabela 9 mostra a produtividade média  $(kg\ ha^{-1})$  de cada genótipo em cada ambiente.

De acordo com a relação de informação (IR) dos quatro componentes (Tabela 10), apenas os dois primeiros PCs contêm padrões (IR  $\geq$  1). Portanto, o biplot é considerado adequado para representar os padrões dos dados. O GGE biplot com base neste conjunto de dados é apresentado na Figura 4, na abscissa do biplot apresentam-se os escores de PC1 e na ordenada os escores do PC2, dos genótipos e ambientes. O GGE biplot mostra a propriedade do produto interno do GGE biplot e permite a visualização do desempenho relativo de cada genótipo em cada ambiente. O biplot explicou 63,96% da variação total G + GE e foi construído no ambiente centrado (centralização = 2), não-dimensionado (escala = 0) e particionamento em valor singular com foco no meio ambiente (PVS = 2).

Os 15 genótipos são rotulados como G1 a G15 e os quatro ambientes como A1 a A4. A linha reta traçada da origem do biplot para a colocação de um ambiente ou genótipo é chamado "vetor de ambiente" ou "vetor de genótipo", de modo que a partir desses vetores, as interações específicas entre um genótipo e um ambiente (por exemplo, o desempenho de cada um dos genótipos em cada ambiente) possam ser visualizadas. A regra de interpretação é: 1) o

desempenho de um genótipo de um ambiente é melhor do que a média, se o ângulo entre o seu vetor e do vetor do ambiente é  $< 90^{\circ}$ ; é pior do que a média se o ângulo  $> 90^{\circ}$  (porque o cosseno de um ângulo obtuso é menor do que 0); e que está próximo da média se o ângulo é aproximadamente igual a 90  $^{\circ}$  (YAN; TINKER, 2006; YAN, 2011).

Tabela 9 - Médias dos 15 genótipos (G1 a G15) testadas em quatro ambientes em relação a produtividade

Genótipo	Ambientes					
	A1	A2	A3	A4		
G1	11924	12508,5	8546,5	5396		
G2	12400,5	14277,5	10324	9781,5		
G3	12135,5	13508	10529,5	8542		
G4	9975,5	12155	11033	8700,5		
G5	12108,5	14287	11221,5	8672		
G6	11159	14335	7596	7195		
G7	12658,5	13516,5	9507,5	8616,5		
G8	10369	9595	9512,5	9683,5		
G9	10018	10329,5	8293	6601		
G10	11529	11639,5	9219	9123,5		
G11	11238,5	10792	6574	5431		
G12	G12 10614,5		7396,5	9272,5		
G13	12449	11977,5	9351	7398,5		
G14	11410	13228,5	11991,5	9957		
G15	11928	12845,5	12000,5	9886		

Tabela 10 - Valor singular, proporção explicada e relação da informação (IR) dos quatro componentes principais (PCs)

PC	Valor singular	Variação explicado (%)	IR
1	20,994	39,35	1,57
2	16,601	24,61	1,00
3	15,634	21,82	0,87
4	12,619	14,22	0,57

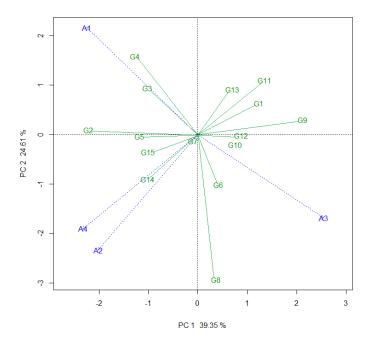


Figura 4 - O GGE Biplot para os dados de produtividade de milho  $(kg \ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)

A Figura 4 pode ser utilizada: 1) para classificar os genótipos com base no desempenho, em qualquer ambiente, e 2) para classificar ambientes sobre o desempenho relativo de qualquer genótipo. Por exemplo, os genótipos G9, G12, G1 e G11 tem ângulos obtusos com três dos quatro ambientes (A1, A2 e A4), isto significa que esses genótipos tiveram menor produtividades médias nesses ambientes. Essa informação pode ser confirmada pela Tabela 9. Os genótipos G11 e G9 apresentaram vetores maiores do que G12 e G1, portanto, devem ter produtividades mais baixas em todos os ambientes, exceto A3. Podem ocorrer erros sobre as interpretações no biplot, devido os dois genótipos não têm exatamente os mesmos ângulos com os ambientes e também a qualidade do ajuste do biplot é de 63,96%, em vez de 100%. Como outro exemplo, os genótipos G2, G15, G5, G14 e G3 formaram ângulos agudos com todos os ambientes, exceto A3, portanto, esse grupo de genótipos produziu mais do que a média em todos os ambientes, exceto no A3, em que devem produzir inferior ou igual às médias ambientais (Figura 4). O genótipo G2 apresentou o maior vetor que o restante dos genótipos neste grupo (Figura 4), portanto, G2 deve ter maior produtividade em todos os ambientes. Isto também pode ser confirmado a partir da Tabela 9.

Devido à propriedade de produto interno, quando um conjunto de dados MET é suficientemente aproximado por uma matriz de posto-2, pode ser graficamente estudado em um GGE biplot de duas dimensões por três aspectos: 1) diferenças e semelhanças entre genótipos; 2) as relações entre ambientes de teste e 3) interações específicas entre genótipo × ambiente (YAN, 2011).

Estes correspondem aos três objetivos da análise de dados MET, ou seja, avaliação de genótipos, avaliação de ambiente de teste e análise de mega-ambiente. O sistema de análise GGE biplot consiste em um conjunto de pontos de vista que foram projetados para atender especificamente cada uma dessas questões.

### 4.2.1 Análise Mega-ambiente (1º conjunto de dados)

Um Mega-ambiente é um grupo de ambientes ou sub-regiões em que um genótipo ou um grupo de genótipos semelhantes são especificamente adaptados e tiveram melhor desempenho. O objetivo da análise de mega-ambiente é tentar dividir uma região de corte alvo, em sub-regiões significativas de modo que G × E possa ser explorada (YAN; KANG, 2003; YAN; HOLLAND, 2010; YAN, 2011). Quando um GGE biplot de dimensão 2 é julgado como uma aproximação suficiente dos dados (Tabela 10), o GGE biplot (Figura 5) (também é chamado como "Whichwon-where") é uma ferramenta eficaz para análise de mega-ambiente.

Este ponto de vista é constituído por um polígono irregular e um conjunto de linhas retas que irradiam a partir da origem do biplot e intersectam cada um dos lados do polígono perpendicularmente. Os vértices do polígono são os marcadores de genótipos localizados mais distante da origem do biplot em todas as direções, de modo a que todos os genótipos estão contidos dentro do polígono (YAN, 2011). Uma linha que cruza perpendicularmente um lado do polígono representa ambientes hipotéticos, em que os dois genótipos que definem esse lado do polígono apresentaram produtividades boas; a posição relativa dos dois genótipos seria invertida em ambientes em lados opostos da linha ("GE cruzado"). Assim, as linhas que se irradiam dividem o biplot em setores, para cada setor, existe um genótipo no vértice, que apresentou melhor desempenho para ambientes que se enquadram nesse setor (YAN; KANG, 2003; YAN; HOLLAND, 2010; YAN, 2011).

A Figura 5 permite o agrupamento visual dos ambientes de teste com base em cruzamento G × E entre os melhores genótipos. Este é o mesmo GGE biplot apresentado na Figura 4, exceto para as linhas complementares. O biplot foi construído no ambiente centrado (centralização = 2), não-dimensionado (escala = 0) e (PVS = 2). Os vértices do polígono são formados pelos genótipos: G8, G9, G11, G4 e G2. Os quatros ambientes foram cortados em 3 grupos pelas linhas que saíram da origem do biplot, os grupos são formados por i) A1; ii) A4; iii) A2 e A3.

O genótipo G2 é o vértice do setor em que o ambiente A4 é colocado, portanto, é o

genótipo que teve melhor desempenho neste ambiente; o G8 é o vértice no setor em que A2 e A3 são colocados, logo, é o genótipo mais adaptado nestes ambientes e é o mesmo caso do genótipo G4 no ambiente A1 (Figura 5). Nos setores de G9, G1, G11, G12 G10 e G13 que não contém ambientes, significa que estes genótipos não eram produtivos em nenhum ambiente, ou seja, estes genótipos são os piores genótipos em relação à produtividade em alguns ou em todos os ambientes.

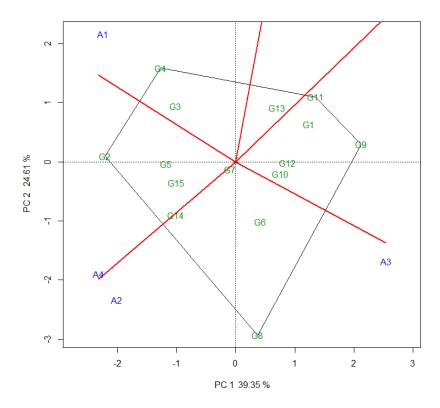


Figura 5 - O GGE Biplot ("Which-won-where") mostra que para os dados de produtividade de milho  $(kg \ ha^{-1})$ , quais dos genótipos apresentaram melhor desempenho e em qual ambientes

O padrão da GE cruzado sugere que os ambientes alvos podem ser divididos em três mega-ambientes diferentes. Este padrão parece ser repetitivo ao longo dos anos e o delineamento mega-ambiente é, portanto, significativo. Se um GGE biplot dimensão 2 é considerado como não adequado em exibir os padrões da GE, uma opção é a construção de um GGE biplot baseado em um subconjunto de dados, removendo genótipos que não produziram bem em todos ou na maioria dos ambientes de teste. Assim, com a exclusão de genótipos de baixo rendimento, levando a menor proporção de G em relação ao GE no novo biplot e, portanto, uma maior separação dos ambientes de teste (YAN, 2011).

#### 4.2.2 Avaliação de genótipos baseada em GGE biplot (1º conjunto de dados)

Um genótipo "ideal" deve ter alto desempenho médio e alta estabilidade por meio de um mega-ambiente. A visualização do GGE biplot "Média versus Estabilidade" é uma ferramenta eficaz para a avaliação de genótipos em ambos aspectos (YAN; TINKER, 2006; YAN et al., 2007; YAN, 2011). Na Figura 6, o pequeno círculo representa o "ambiente-média" e é definido pelas coordenadas médias de todos os ambientes de teste no biplot.

A linha reta com uma única seta que passa pela origem do biplot e do ambientemédia é referido como o "eixo do ambiente-média" ou EAM. A seta aponta para um maior desempenho médio para os genótipos. A linha com duas setas que passa pela origem do biplot e é perpendicular ao EAM, com as setas apontam para a maior variabilidade de desempenho (menor estabilidade) em ambas as direções. Este biplot é baseado em (PVS = 1), ou seja, os valores singulares são totalmente compartimentados para as escores genotípicos (YAN, 2002; YAN, 2011).

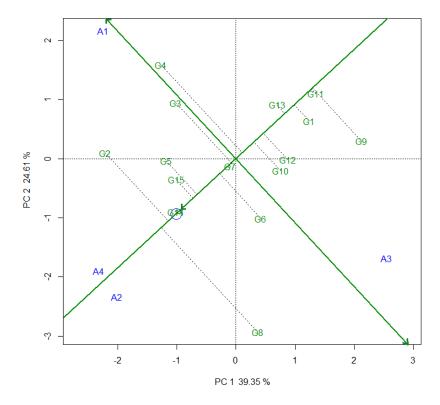


Figura 6 - O GGE Biplot ("Média versus Estabilidade") com eixo do ambiente-média (EAM) para mostrar o desempenho médio e estabilidade dos genótipos para os dados de produtividade de milho  $(kg\ ha^{-1})$ 

Assim, os genótipos são classificados de acordo com sua produtividade média da seguinte forma: G8 > G2 > G14 > G15 > G5 > ... > G3 > média geral > G4 > G10 > ... > G1 > G11 > G9; O G8 foi altamente instável, pois tinha rendimento inferior ao esperado em ambiente A1, enquanto produziu relativamente bem em A2, A3 e A4; O G2 não foi estável, mas apresentou um desempenho bem superior em comparação aos outros genótipos em relação à produtividade, produzindo bem nos ambientes A1, A2 e A4, exceto no A3. O G7 foi estável e com produtividade pouco acima da média, já o Genótipo G14 teve produtividade bem acima da média (terceiro melhor) e foi o mais estável, portanto, o G14 foi o genótipo "ideal" para recomendação (Figura 6). Note-se que, se o biplot explica apenas uma proporção da variação total, alguns genótipos aparentemente estáveis podem não ser verdadeiramente estáveis como as suas variações não podem ser completamente explicadas neste biplot.

A Figura 6 é útil apenas quando o G é considerável. Quando G é muito pequeno em relação à GE, essa interpretação não será significativa. Mas isso não compromete a utilidade do GGE biplot. Isso reflete o senso comum que geralmente genótipos não adaptados podem ser esperados e genótipos adaptados significativamente devem ser procurado quando G é desprezível (YAN; TINKER, 2006; YAN, 2011).

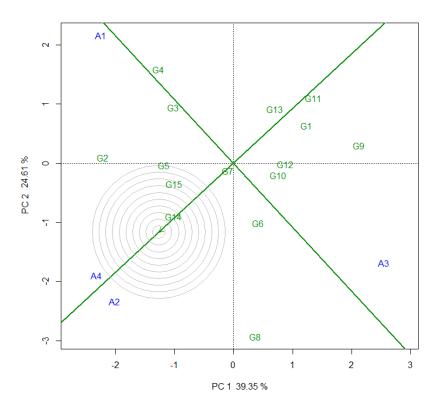


Figura 7 - O GGE Biplot com eixo do ambiente-média (EAM) para classificar os genótipos em relação ao genótipo ideal (no centro dos círculos concêntricos)

Um "genótipo ideal" (o centro dos círculos concêntricos) pode ser um ponto sobre o EAM, no sentido positivo e tem um comprimento do vetor igual aos vetores mais longos dos genótipos sobre o lado positivo do EAM, isto é, mais alto desempenho médio (Figura 7). Portanto, os genótipos localizados mais perto do genótipo ideal são mais desejáveis do que outros. Assim, G14 foi o "genótipo ideal" neste conjunto de dados, em seguida, os melhores foram os genótipos G15 e G5. Embora G8 e G2 tiveram rendimento médio superior, não foram estáveis, de acordo com Figura 7, o G12 foi mais desejável que o G8; os piores genótipos para recomendação foram G11 e G9. Os genótipos G13 e G11 são altamente estáveis (Figura 6 e 7), isso não significa que estes genótipos tiveram rendimentos bons, significa apenas que o desempenho relativo do G13 e G11 foram consistentes, mas estão longe de serem genótipos ideais.

A Figura 7 ilustra um conceito importante em relação a estabilidade, o termo "estabilidade elevada" só tem sentido quando associado ao desempenho médio; o genótipo estável é desejado apenas quando apresenta alta performance média (YAN; TINKER, 2006; YAN, 2011).

### 4.2.3 Avaliação do ambiente de teste a teste-ambiente (1º conjunto de dados)

O objetivo da avaliação do "teste-ambiente" é identificar ambientes de teste que podem ser utilizados para selecionar genótipos superiores de forma eficaz para um mega-ambiente. Um ambiente de teste "ideal" deve ser tanto de discriminação dos genótipos e representação do ambiente alvo. A visualização no GGE biplot "Discriminação contra Representatividade" (Figura 8) foi projetado para esta finalidade.

Quando o biplot GGE baseia-se fora de escala (não padronizado), dados centrados no ambiente (Escala = 0), o comprimento do vetor de um ambiente é proporcional ao desvio padrão das médias de genótipos (SD), que é igual à raiz quadrada da variância fenotípica  $(\sigma_p)$  em ambiente de teste, o qual pode ser utilizado como uma medida do poder de discriminação do ambiente (YAN, 2011). Os ambientes de teste com vetores mais longos são mais discriminantes em relação aos genótipos. Aqueles ambientes com um vetor curto é menos discriminante, o que significa que todos os genótipos tendem a executar de forma semelhante e pouca ou nenhuma informação sobre as diferenças genotípicas podem ser reveladas em tal ambiente, portanto, não devem ser utilizados como ambientes de teste. Um curto vetor também pode significar que o ambiente não é bem representado por PC1 e PC2 se o biplot não exibir adequadamente o G + GE dos dados. Todos os ambientes (A1, A2, A3 e A4) apresentaram vetores longos, significando que são mais discriminantes em relação aos genótipos (Figura 8).

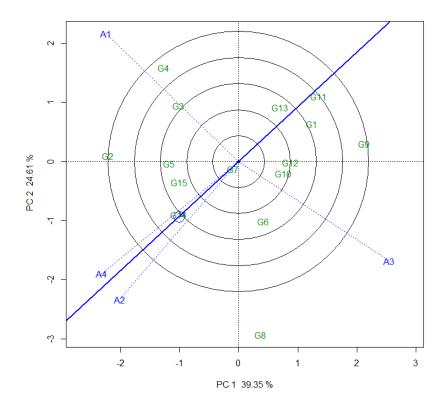


Figura 8 - O GGE biplot "discriminação e representatividade" para mostrar a capacidade discriminação e de representatividade dos ambientes de teste

Uma segunda utilidade da Figura 8 é indicar ambientes de representatividade do ambiente alvo. Uma vez que os ambientes de teste que têm pequenos ângulos com EAM (eixo do ambiente-média), por exemplo, A4 e A2, são mais representativos do ambiente alvo do que aqueles que têm ângulos maiores, por exemplo, A1 e A3. Portanto, os ambientes de teste que são discriminante e representativos (por exemplo, A4 e A2) são ambientes de teste ideal para a seleção de genótipos adaptados. Os ambientes de teste discriminante mas não representativos (por exemplo, A1 e A3) são úteis para a seleção de genótipos adaptados especificamente se os ambientes alvos podem ser divididos em mega-ambientes; ou para seleção de genótipos instáveis se o ambiente alvo é um único mega-ambiente (Figura 8).

Dentro de um único mega-ambiente, o ambiente de teste ideal deve ser mais discriminante e também mais representativo do ambiente alvo. Um "ambiente de teste ideal" é o centro dos círculos concêntricos (Figura 9). É um ponto da EAM no sentido positivo, com uma distância da origem biplot igual ao vetor mais longo de todos os ambientes. A4 é mais próximo a este ponto, portanto, é o "ambiente de teste ideal", o ambiente A2 é o segundo melhor. O ambiente A3 foi o pior para selecionar genótipos adaptados a toda a região (Figura 9).

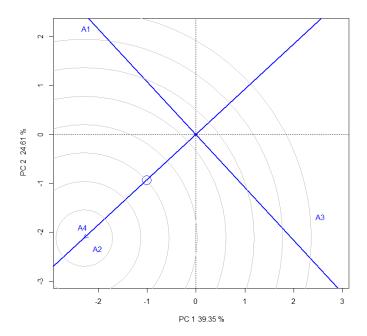


Figura 9 - O GGE biplot discriminação e representatividade para classificar ambientes de teste em relação a um ambiente de teste ideal (representado por centro dos círculos concêntricos)

Para resumir, GGE biplot pode efetivamente resolver os três principais objetivos da análise de dados MET: mega-ambiente delimitação, avaliação de genótipos e avaliação ambiente de teste. A Análise GGE biplot para os dados de MET pode ajudar os pesquisadores a entender melhor seu ambiente alvo, para estabelecer estratégias e testes de melhoramento de baixo custo, além de identificar genótipos superiores que são amplamente ou especificamente adaptados (YAN, 2011).

# 4.3 Análise AMMI-ponderado e Biplot para $1^o$ conjunto dados multi-ambientes (MET)

Quando existe heterogeneidade nos ambientes de testes nos dados multi-ambientes (MET), a análise AMMI não é eficiente para alcançar os objetivos da avaliação de genótipos e ambientes e compreender as causas da interação  $G \times E$ . Quando a variância do erro (error variance) numa tabela de dupla entrada da  $GE(\mathbf{Y})$  é heterogênea entre os ambientes, os valores da tabela devem ter pesos diferentes para os seus resíduos quadrados no processo de estimação dos parâmetros

do modelo. Da forma mais geral, quando a estrutura de variância e covariância do erro para genotípica difere significativamente entre ensaios ou ambientes, por causa da heterogeneidade entre os ambientes, portanto, essa heterogeneidade deve ser levada em consideração na análise de GEI (RODRIGUES et al., 2014).

Tabela 11 - Médias dos genótipos, ambientes e posição das médias em relação a produtividade

Genótipo	Média-centrado	Posição	Ambiente	Média-centrado	Posição
$G_1$	0,0064	10	$A_1$	0,0147	1
$G_2$	0,0074	1	$A_2$	0,0050	2
$G_3$	0,0071	6	$A_3$	0,0034	4
$G_4$	0,0064	11	$A_4$	0,0035	3
$G_5$	0,0073	3			
$G_6$	0,0065	9			
$G_7$	0,0072	4			
$G_8$	0,0062	12			
$G_9$	0,0057	15			
$G_{10}$	0,0067	8			
$G_{11}$	0,0059	14			
$G_{12}$	0,0061	13			
$G_{13}$	0,0068	7			
$G_{14}$	0,0071	5			
$G_{15}$	0,0073	2			

Para solucionar este problema com ambientes heterogêneos, neste trabalho foi criada uma generalização do modelo AMMI (AMMI-ponderado com variância dos ambientes). Com AMMI ponderado para análise de dados MET, permite levar em consideração a heterogeneidade da variância do erro ao longo dos ambientes. Os dados centrados nos ambientes foram obtidos dividindo-se a variância do erro de cada ambiente a partir do valor de rendimento original de cada genótipo em cada ambiente. A variância do erro experimental de cada ambiente ou os quadrados médios dos resíduos de cada ambiente  $(s_j^2)$  pela análise da variância conjunta são: (A1:  $s_1^2 = 215985,3 \ (kg \ ha^{-1})^2$ ; A2:  $s_2^2 = 93179,33 \ (kg \ ha^{-1})^2$ ; A3:  $s_3^3 = 2835058 \ (kg \ ha^{-1})^2$  e A4:  $s_4^2 = 34644,63 \ (kg \ ha^{-1})^2$ ).

A análise de variância conjunta dos dados considerando 15 genótipos em quatro ambientes com 2 blocos é a mesma utilizando AMMI ponderado e o AMMI (Tabela 4). As

estimativas das médias dos genótipos e dos ambientes centradas em relação aos ambientes são apresentados na Tabela 11. O genótipo G2 apresentou maior valor da estimativa da produtividade, seguidos dos genótipos G15, G5, G7 e G14, respectivamente. Já os genótipos G9, G11 e G12 apresentaram as piores estimativas da produtividade. Com relação aos ambientes, têm-se que o ambiente A1 apresentou a maior estimativa de produtividade média e a pior foi A3 (Tabela 11).

Os dados da Tabela 12, que são valores de rendimento centrados nos ambientes de 15 genótipos de milho testados em quatro ambientes. A partir desta Tabela 12 que é uma matriz de dupla entrada de genótipos e ambientes foi construída a matriz de interação GE e também a construção dos gráficos biplot (Figura 10 e 11). O PC1 explicou 75,5% da variação total da GEI e o PC2 explicou 16,7% (Figura 11). com isso, os dois componentes explicaram um total de 92,2% da GEI, enquanto a análise AMMI explicou apenas 87% da GEI com o PC1 e PC2. Portanto, o biplot PC1 vs PC2 do AMMI-ponderado explicou uma maior variação de GEI em comparação com biplot AMMI, consequentemente, as conclusões do biplot AMMI-ponderado (Figura 11) são mais confiáveis do que as do biplot AMMI (Figura 3).

Tabela 12 - Produtividade média com os valores centrada nos ambientes de testes com 15 genótipos (G1 a G15) em quatro ambientes (A1 a A4)

Genótipo		Ambi	ientes	
	A1	A2	A3	A4
G1	0,0153	0,0051	0,0031	0,0023
G2	0,0159	0,0058	0,0038	0,0042
G3	0,0155	0,0055	0,0038	0,0036
G4	0,0128	0,0049	0,0040	0,0037
G5	0,0155	0,0058	0,0040	0,0037
G6	0,0143	0,0058	0,0028	0,0031
G7	0,0162	0,0055	0,0035	0,0037
G8	0,0133	0,0039	0,0035	0,0041
G9	0,0128	0,0042	0,0030	0,0028
G10	0,0148	0,0047	0,0034	0,0039
G11	0,0144	0,0044	0,0024	0,0023
G12	0,0136	0,0042	0,0027	0,0039
G13	0,0159	0,0049	0,0034	0,0031
G14	0,0146	0,0054	0,0044	0,0042
G15	0,0153	0,0052	0,0044	0,0042

O biplot AMMI1-ponderado contém a variação dos efeitos aditivos principais de genótipos e ambientes, visualizado na horizontal do gráfico (eixo das abscissas) e a variação dos efeitos multiplicativos da GEI, visualizado na vertical (eixo das ordenadas). No biplot AMMI2-ponderado é visualizado apenas os efeitos multiplicativos da GEI contidos nos PC1 e PC2. A interpretação do gráfico biplot permitiu avaliar a estabilidade dos genótipos nos ambientes por meio da observação da magnitude e do sinal dos escores dos genótipos e ambientes para o eixo da interação (Tabela 13). Portanto, os escores baixos (próximos de zero) representam os genótipos e ambientes mais estáveis (GAUCH, 2013).

Os genótipos são classificados de acordo com sua produtividade média da seguinte forma: G2 > G15 > G5 > G7 > G14 > G3 > G13 > G10 > média geral dos genótipos > G6 > G1 > G4 > ... > G11 > G9; A classificação dos ambientes são: A1 > A2 > A4 > A3 (Figura 10 e Tabela 11). O G2 foi o genótipo que apresentou maior produtividade e o A1 foi o melhor entre os ambientes de testes.

Em relação às adaptações específicas dos genótipos em ambientes, verifica-se que os genótipos G13, G7, G2 foram mais adaptados ao ambiente A1; os genótipos G8, G12, G15 e G10 foram mais adaptados no A4; os genótipos G4, G9 e G14 no ambiente A3; G6, G5 e G1 no ambiente A2 (Figura 11). O genótipo G6 foi o genótipo que apresentou maior produtividade no A2, G8 no ambiente A4; G13 no ambiente A1 e G14 no ambiente A3. Essa afirmação pode ser confirmada também na Tabela 11. Observa-se que os vetores relativos a esses genótipos apontam para uma direção comum, que pode ser orientada pelos ambientes. Além disso, constata-se a falta de adaptação dos genótipos G4 e G9 ao ambiente A1; dos genótipos G8 e G12 ao ambiente A2; do genótipo G13 ao ambiente A3 e os genótipos G1 e G6 ao ambiente A4 (Figura 11).

O ambiente A1 foi altamente instável, mas foi o ambiente que apresentou a maior produtividade média, mas A1 não foi o "ideal" ambiente de teste por falta de estabilidade. O ambiente A2 foi o segundo melhor em relação a produtividade e foi o mais estável, neste conjunto de dados, por isso A2 foi o "ambiente de teste ideal" (produtividade e estabilidade). O ambiente A3 foi estável, mas não teve desempenho em produtividade; o pior ambiente foi A4, pois foi instável e apresentou produtividade bem abaixo da média (Figura 10 e 11).

O genótipo G8 foi altamente instável, pois teve rendimento inferior ao esperado em ambiente A1 e A2, enquanto produziu relativamente bem em A4 e A3 (Figura 11). Os genótipos G4, G1, G7 e G13 não foram estáveis; o G1 produziu bem nos ambientes A1 e A2, mas, foi bem abaixa da média em A4; o G4 produziu bem no ambiente A3 e foi bem abaixo da média no ambientes A1; os genótipos G13 e G7 apresentaram um desempenho muito semelhantes nos ambientes de testes, ambos apresentaram produtividade acima da média geral, são genótipos que adaptaram bem no ambiente A1 e com falta de adaptação nos restantes ambientes (Figura 10 e

11).

Para avaliação de genótipos mais estáveis, temos PC1 com 75,5% da variação total da GEI e PC2 com somente 16,7%, portanto, o PC1 apresentou a maior variabilidade total de GEI. Os genótipos mais estáveis (PC1) foram G10, G15, G6 e G5, G3 e G2 (Figura 10 e 11; Tabela 13), estes podem ser cultivados em todos os locais onde foram feitos os estudos. Entre estes, apenas o genótipo G6 teve produtividade abaixo da média geral.

Os "ideais" genótipos (Estabilidade e produtividade) foram G5, G3 e G2, que podem ser considerados como os melhores entre todos os genótipos de testes. Como o biplot AMMI-ponderado explicou 92,2% da variação total de GEI, logo, com esse novo método, as interpretações dos gráficos de biplot são altamente confiáveis para recomendação de cultivares.

Tabela 13 - Coordenadas dos eixos das abscissas (x) e de coordenadas (y), para os dois gráficos Biplot

Genótipos	Biplo	t AMMI1	Biplot AMMI2		
Ambientes	Média $(x)$	PC1(y)	PC1(x)	PC2(y)	
G1	0,0064	-0,0204	-0,0204	-0,0094	
G2	0,0074	-0,0095	-0,0095	0,0033	
G3	0,0071	-0,0088	-0,0088	-0,0004	
G4	0,0064	0,0292	0,0292	-0,0154	
G5	0,0073	-0,0069	-0,0069	-0,0066	
G6	0,0065	-0,0044	-0,0044	-0,0226	
G7	0,0072	-0,0193	-0,0193	0,0070	
G8	0,0062	0,0248	0,0248	0,0151	
G9	0,0057	0,0162	0,0162	-0,0091	
G10	0,0067	0,0018	0,0018	0,0107	
G11	0,0059	-0,0120	-0,0120	-0,0026	
G12	0,0061	0,0144	0,0144	0,0129	
G13	0,0068	-0,0196	-0,0196	0,0103	
G14	0,0071	0,0115	0,0115	-0,0010	
G15	0,0073	0,0030	0,0030	0,0077	
A1	0,0147	-0,0466	-0,0466	0,0159	
A2	0,0050	-0,0058	-0,0058	-0,0321	
A3	0,0034	0,0209	0,0209	-0,0040	
A4	0,0035	0,0315	0,0315	0,0202	
		<u> </u>	<u> </u>		

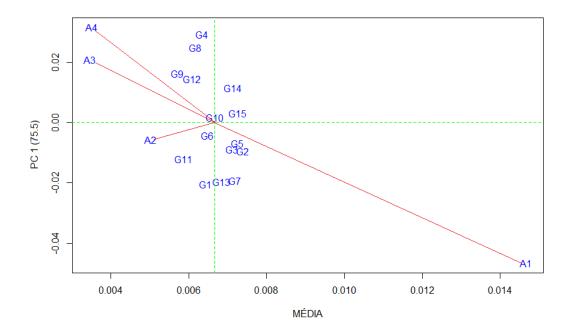


Figura 10 - Biplot AMMI1-ponderado (Médias vs PC1) para os dados de produtividade de milho ( $kg\ ha^{-1}$ ), com 15 genótipos (G) e quatro ambientes (A)

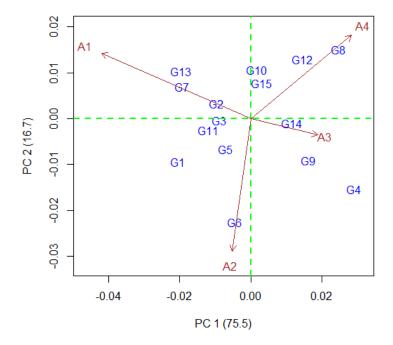


Figura 11 - Biplot AMMI2-ponderado (PC1 vs PC2) para os dados de produtividade de milho  $(kg\ ha^{-1})$ , com 15 genótipos (G) e quatro ambientes (A)

## 4.4 GGE biplot ponderado (W-GGE biplot) para análise de $1^o$ conjunto dados MET

Existem dois problemas na análise de GGE biplot para os dados multi-ambientes (MET), um dos principais problemas é quando existe heterogeneidade nos ambientes de testes, isso pode influenciar na análise utilizando o GGE biplot para alcançar os principais objetivos que são: investigação de mega-ambiente; avaliação de genótipos dentro de cada mega-ambiente; avaliação de ambiente de teste de cada mega-ambiente e compreensão as causas da GEI.

A estrutura de variância e covariância do erro para genotípica difere significativamente entre ensaios ou ambientes por causa da heterogeneidade dos ambientes, logo essa heterogeneidade deve ser levada em consideração na análise de GEI (RODRIGUES et al., 2014). Para solucionar este problema com ambientes heterogêneos, neste trabalho criou-se uma generalização do modelo GGE biplot: W-GGE (Weighted Genotype main effects + Genotype environment interaction) biplot, podendo ser chamado de GGE biplot ponderado.

Tabela 14 - Produção de dados de ambiente-centrados com os 15 genótipos (G1 a G15) testadas em quatro ambientes

Genótipo	Ambientes					
	A1	A2	A3	A4		
G1	0,0552	0,1342	0,0030	0,1558		
G2	0,0574	0,1532	0,0036	0,2823		
G3	0,0562	0,1450	0,0037	0,2466		
G4	0,0462	0,1304	0,0039	0,2511		
G5	0,0561	0,1533	0,0040	0,2503		
G6	0,0517	0,1538	0,0027	0,2077		
G7	0,0586	0,1451	0,0034	0,2487		
G8	0,0480	0,1030	0,0034	0,2795		
G9	0,0464	0,1109	0,0029	0,1905		
G10	0,0534	0,1249	0,0033	0,2633		
G11	0,0520	0,1158	0,0023	0,1568		
G12	0,0491	0,1097	0,0026	0,2676		
G13	0,0576	0,1285	0,0033	0,2136		
G14	0,0528	0,1420	0,0042	0,2874		
G15	0,0552	0,1379	0,0042	0,2854		

Tabela 15 - Valor singular, proporção explicada e relação da informação (IR) dos quatro componentes principais (PCs)

PC	Valor singular	Variação explicado (%)	IR
1	23,674	50,04	2,0016
2	18,948	32,06	1,2824
3	11,385	11,57	0,4628
4	8,42	6,33	0,2532

Este modelo W-GGE biplot é uma generalização do GGE biplot para análise de dados MET, que permite levar em consideração a heterogeneidade da variância do erro ao longo dos ambientes. Os dados da Tabela 14, são valores de rendimento centrados no ambiente de 15 genótipos de milho testados em quatro ambientes; os dados centrados no meio ambiente foram obtidos dividindo-se a variância do erro de cada ambiente a partir do valor de rendimento original de cada genótipo em cada ambiente (expressão 39). De acordo com a relação de informação (IR) dos quatro componentes, os dois primeiros PCs contêm padrões (IR > 1) (Tabela 15).

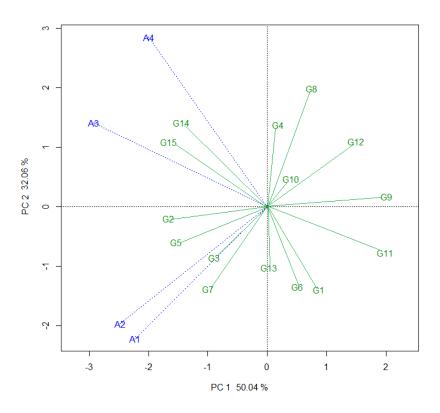


Figura 12 - O W-GGE Biplot para os dados de produtividade de milho  $(kg \ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)

Portanto, o biplot é considerado adequado para representar os padrões nos dados. O PC1 explicou 50,04% e PC2 32,06% da variação total G + GE (Tabela 15), com isso, o W-GGE biplot explicou 82,1% da variação total G + GE (Figura 12), para este mesmo conjunto de dados, o GGE biplot tinha explicado apenas 63,96% da variação total G + GE. O W-GGE biplot foi construído no ambiente centrado (centralização = 2), não-dimensionado (escala = 0) e particionamento em valor singular com foco no meio ambiente (PVS = 2).

Os genótipos G9, G11, G12 formaram ângulos obtusos com todos os ambientes (A1, A2, A3 e A4), isto significa que esses genótipos tiveram menores produtividades médias em todos os ambientes de testes (Figura 12). Essa informação pode ser confirmada pela Tabela 14. Os genótipos G11 e G9 apresentaram vetores maiores do que G12, portanto, devem ter piores produtividades em todos os ambientes. Os genótipos G1 e G6 formaram ângulos obtusos com A4 e A3, esses dois genótipos tiveram rendimento bem abaixo da média nestes dois ambientes, é o mesmo caso dos genótipos G10 e G8 nos ambientes A1 e A2 (Figura 12).

Os genótipos G2, G5, G15 e G14 formaram ângulos agudos com todos os ambientes, significando que esse grupo de genótipos produziu acima da média em todos os ambientes. Os G15 e G14 apresentaram vetores pouco maiores do que G2 e G5, mas, G2 e G5 formaram ângulos menores com A2 e A1 do que G14 e G15; é o mesmo caso, os ângulos formados pelo A3 e A4 com G14 e G15, portanto, G15 e G14 tiveram maior produtividade nos ambientes A3 e A4, G2 e G5 tiveram mais produtividade nos ambientes A2 e A1 (Figura 12). Isto também pode ser confirmado a partir da Tabela 14. Os desvios desta previsão podem ocorrer, porque os dois genótipos não têm exatamente os mesmos ângulos com os ambientes e porque a qualidade do ajuste do biplot é de 82,1%, em vez de 100%.

### 4.4.1 Análise Mega-ambiente com W-GGE biplot (1º conjunto de dados)

Na Figura 13 encontra-se o W-GGE biplot ("Which-won-where") que é uma ferramenta eficaz para análise de mega-ambiente sendo o mesmo gráfico apresentado da Figura 12, exceto para as linhas complementares. O W-GGE biplot foi construído no ambiente centrado (centralização = 2), não-dimensionado (escala = 0) e (PVS = 2). Os vértices do polígono são formados pelos genótipos: G7, G1, G11, G9, G12, G8, G14, G15, G2 e G5. os quatros ambientes foram cortados em 3 grupos pelas as linhas que saíram da origem do biplot e os grupos são formados por: i) A1 e A2; ii) A3; iii) A4 (Figura 13).

O genótipo G7 é o vértice do setor em que o ambiente A1 e A2 são colocados,

portanto, é o genótipo que teve melhor desempenho neste ambiente; o G15 é o vértice no setor em que A3 é colocado, logo, é o genótipo mais adaptado nestes ambientes e é o mesmo caso do genótipo G14 no ambiente A4 (Figura 13). Nos setores de G8, G12, G9, G11, G1 que não contém ambientes, isto significa que estes genótipos não eram produtivos em nenhum ambiente, ou seja, estes genótipos são os piores genótipos em relação à produtividade em alguns ou em todos os ambientes.

O padrão da GE cruzado sugere que os ambientes alvos podem ser divididos em três mega-ambientes diferentes. Se um GGE biplot de dimensão 2 é considerado como não adequado em exibir os padrões da GEI, uma opção é a construção de um GGE biplot baseado em um subconjunto de dados, removendo genótipos que produziram mal em todos ou na maioria dos ambientes de teste, assim, com a exclusão de genótipos de baixo rendimento levar a menor proporção de G em relação ao GE no novo biplot e, portanto, uma maior separação dos ambientes de teste (YAN, 2011).

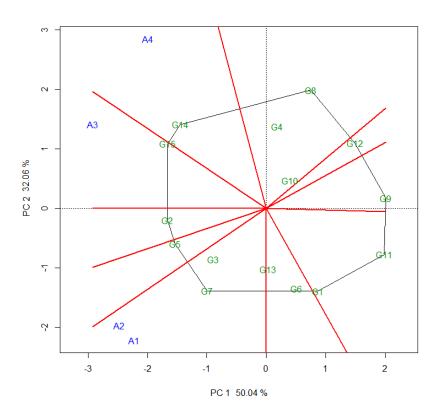


Figura 13 - O W-GGE Biplot ("Which-won-where") mostra que para os dados de produtividade de milho  $(kg\ ha^{-1})$ , quais dos genótipos apresentaram melhor desempenho em que ambientes

### 4.4.2 Avaliação de genótipos baseado em W-GGE biplot (1º conjunto de dados)

Foi utilizado o W-GGE biplot "Média versus Estabilidade" (Figura 14) para identificar um genótipo "ideal", que deve ter alto desempenho médio e alta estabilidade por meio de um mega-ambiente. O pequeno círculo representa o "ambiente-médio" e é definida pelas coordenadas médias de todos os ambientes de teste no biplot. Este biplot é baseado em (PVS = 1), ou seja, os valores singulares são totalmente compartimentados para os escores genotípicos.

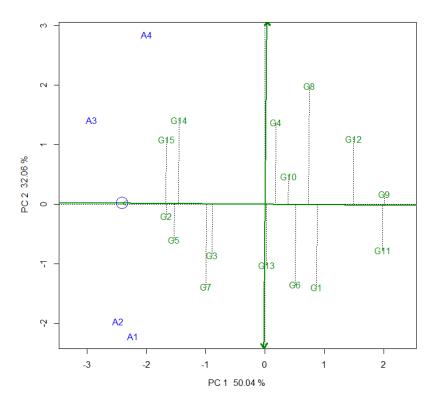


Figura 14 - O W-GGE Biplot ("Média versus Estabilidade") com eixo do ambiente-média (EAM) para mostrar o desempenho médio e estabilidade dos genótipos para os dados de produtividade de milho  $(kg\ ha^{-1})$ 

Assim, os genótipos são classificados de acordo com sua produtividade média da seguinte forma: G15 > G2 > G5 > G14 > G7 > G3 > média geral > G13 > G4 > ... > G12 > G11 > G9. Nota-se que o G8 foi altamente instável, pois teve rendimento abaixo da média em todos os ambientes de testes; os genótipos G14 e G4 foram instáveis, a diferença é o G14 teve produtividade acima da média, pois produziu bem no A4, já o G4 teve rendimento abaixo da média em todos os ambientes (Figura 14).

O G15 não foi estável, mas foi o genótipo que apresentou maior produtividade

média, principalmente no ambiente A3. O G11 foi instável e também o segundo pior em relação a produtividade. Os Genótipos G2 e G9 foram os mais estáveis em comparação com os restantes, a grande diferença entre os dois é a produtividade, o G2 foi o segundo melhor genótipo em relação a produtividade e há uma proximidade com o ambiente A2, enquanto que G9 foi o pior produtivo (Figura 14). Se o biplot explica apenas uma pequena proporção da variação total G + GE, alguns genótipos aparentemente estáveis podem não ser verdadeiramente estáveis como as suas variações não podem ser completamente explicadas neste biplot.

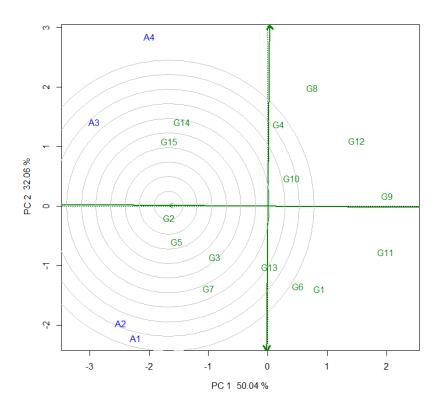


Figura 15 - O W-GGE Biplot com eixo do ambiente-média (EAM) para classificar os genótipos em relação ao genótipo ideal (no centro dos círculos concêntricos)

O W-GGE biplot (com o centro dos círculos concêntricos) sobre o eixo do ambientemédia (EAM) (Figura 15), os genótipos localizados mais próximos do centro dos círculos concêntricos são mais próximos de ser um genótipo ideal. Assim, G2 foi considerado o genótipo "ideal" (Estabilidade e produtividade) neste conjunto de dados, portanto, o G2 é um excelente genótipo para recomendação para os melhoristas; em seguida, os melhores foram os genótipos G5, G15, G3 e G14, respectivamente. No caso do G15, este teve maior produtividade média, mas não foi estável e portanto não foi um genótipo ideal. Os piores genótipos para recomendação foram G11 e G9 (Figura 15). O genótipo G9 foi altamente estável (Figura 14 e 15), mas, não indica que este genótipo teve rendimento bom e está longe de ser um genótipo ideal.

# 4.4.3 Avaliação do Ambiente de Teste baseado no W-GGE biplot (1º conjunto de dados)

O objetivo da avaliação do "teste-ambiente" é identificar ambientes de teste que podem ser utilizados para selecionar genótipos superiores de forma eficaz para um mega-ambiente. Um ambiente de teste "ideal" deve ser tanto de discriminação dos genótipos e representante do ambiente alvo. O W-GGE biplot "Discriminação contra Representatividade" (Figura 16) foi projetado para esta finalidade.

Os ambientes de teste com vetores mais longos são mais discriminantes em relação aos genótipos. Aqueles ambientes com um vetor curto são menos discriminantes, o que significa que todos os genótipos tendem a executar de forma semelhante e pouca ou nenhuma informação sobre as diferenças genotípicas podem ser reveladas em tais ambientes, portanto, não devem ser utilizados como ambientes de teste. Todos os ambientes (A1, A2, A3 e A4) apresentaram vetores longos, significando que são mais discriminantes em relação aos genótipos (Figura 16).

Uma segunda utilidade da Figura 16, é indicar ambientes de representatividade do ambiente alvo. Uma vez que os ambientes de teste têm pequenos ângulos com EAM (eixo do ambiente-média), por exemplo, A3 e A2, estes são mais representativos do ambiente alvo do que aqueles que têm ângulos maiores, por exemplo, A1 e A4. Portanto, os ambientes de teste que são discriminantes e representativos (A3 e A2) são ambientes de teste ideal para a seleção de genótipos adaptados. Os ambientes de teste discriminante mas não representativo (A1 e A4) são úteis para a seleção de genótipos adaptados especificamente se os ambientes alvos podem ser divididos em mega-ambientes; ou para seleção de genótipos instáveis se o ambiente alvo é um único mega-ambiente (Figura 16).

Dentro de um único mega-ambiente, o ambiente de teste ideal deve ser mais discriminante e também mais representativo do ambiente alvo. Um "ambiente de teste ideal" é o centro dos círculos concêntricos (Figura 17). Verifica-se que o ambiente A3 é o mais próximo a este ponto, portanto, é o ambiente de teste ideal e o ambiente A2 é o segundo melhor. O ambiente A4 foi o pior para selecionar genótipos adaptados a toda a região (Figura 17).

Para resumir, o W-GGE biplot pode efetivamente resolver os três principais objetivos da análise de dados MET: análise de mega-ambiente, avaliação de genótipos e avaliação ambiente de teste. A análise W-GGE biplot para os dados de MET pode ajudar os pesquisadores a entender melhor seu ambiente alvo, para estabelecer estratégias e testes de melhoramento de baixo custo, além de identificar genótipos superiores que são amplamente ou especificamente adaptados (YAN, 2011).

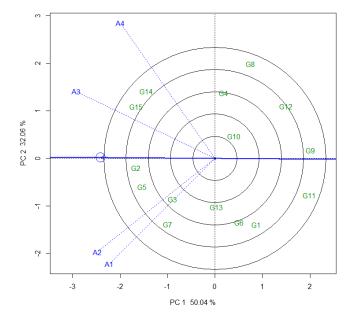


Figura 16 - O W-GGE biplot "discriminação e representatividade" para mostrar a capacidade discriminação e de representatividade dos ambientes de teste

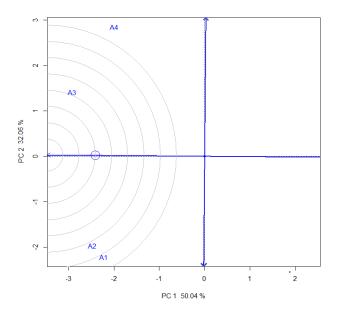


Figura 17 - O W-GGE biplot discriminação e representatividade para classificar ambientes de teste em relação a um ambiente de teste ideal (representado por centro dos círculos concêntricos)

# 4.5 Comparação das análises (AMMI vs AMMI-ponderado vs GGE biplot vs W-GGE biplot) para 1º conjunto de dados MET

#### 4.5.1 Modelo AMMI versus GGEbiplot

O modelo GGE biplot apresenta o efeito de genótipo (G) como um efeito multiplicativo da GEI. Em geral, o PC1 está correlacionado com o efeito do genótipo e é possível ter alta precisão na visualização gráfica do escore PC1 quanto ao efeito genotípico (YAN; KANG, 2003; YAN; HOLLAND, 2010). Esta proporcionalidade da resposta de genótipos é mais lógica e biologicamente plausível, pois há correlação quase perfeita entre o rendimento de grãos e o PC1 (YAN; KANG, 2003; FRITSCHE-NETO et al., 2010; SINEBO; LAKEW; FEYISSA, 2010). Por outro lado, no modelo AMMI2, tem-se observado correlação negativa entre PC1 e o rendimento (KVITSCHAL et al., 2009). Entretanto, no modelo AMMI1, há correlação quase perfeita com o rendimento de grãos, assim como na GGE biplot, pois é um modelo baseado no PC1 versus rendimento de grãos.

O modelo AMMI possui maior acurácia na visualização do padrão em experimentos, quando comparado ao GGE biplot (GAUCH, 2006). Entretanto, Yan et al. (2007) afirmaram que o modelo GGE é mais vantajoso em relação ao modelo AMMII (PC1 × Rendimento), pois é capaz de explicar uma maior porcentagem da GEI em relação ao gráfico AMMII. Nesse sentido, Gauch, Piepho e Annicchiarico (2008) estudaram que enquanto a análise GGE captou 77,3% do padrão, o modelo AMMII captou apenas 64,6%. O modelo AMMI2, normalmente, apresenta maior porcentagem de explicação do padrão, quando comparado ao GGE biplot, sendo indicado como o melhor modelo da análise AMMI (GAUCH; PIEPHO; ANNICCHIARICO, 2008; BALLESTRE et al., 2009). Entretanto, nem sempre a maior captação de porcentagem de explicação representa maior precisão na seleção de genótipos superiores, isso porque pode haver ruídos na porcentagem de explicação.

O modelo AMMI1 é menos adequado em relação ao GGE para a definição de mega-ambientes, pois, no modelo AMMI1, os ambientes são apresentados apenas na abscissa e os genótipos não são apresentados por pontos, mas por linhas retas (YAN et al., 2007). Gauch, Piepho e Annicchiarico (2008) contestaram tal comparação afirmando que o modelo AMMI1 possui geometria mais simples e informativa e que capta menos ruídos do que o GGE e GAUCH (2006) critica o modelo GGE biplot, pois esse não separa o efeito do genótipo (G) da GEI, tal como a análise AMMI. Entretanto, Yan et al. (2011) e Yan et al. (2007) relataram que, o modelo GGE biplot é capaz de reter maior explicação da G+GEI e os melhoristas têm interesse em selecionar plantas com base no G e na GEI, simultaneamente. Assim, o efeito principal do ambiente não tem

importância quando se realizam testes de cultivares, pois este efeito é irrelevante na identificação de genótipos superiores (YAN; KANG, 2003). Embora o desempenho da produtividade seja o efeito combinado do genótipo, ambiente e da GEI, apenas G e GEI são relevantes e devem ser considerados simultaneamente na avaliação da adaptabilidade e estabilidade de genótipos (YAN, 2011; SILVA; BENIN, 2012).

A única diferença entre as metodologias GGE e AMMI é que a primeira analisa diretamente o efeito de G+GEI, enquanto que o modelo AMMI separa o efeito de G da GEI, mas essa separação não é capaz de conferir superioridade à análise AMMI (YAN; KANG, 2003; YAN et al., 2007). Nesse contexto, YAN et al. (2007) argumentam que a separação de G da GEI pode acarretar em equívocos na seleção de genótipos, pois a GEI é apenas um dos componentes da adaptação específica. Portanto, um genótipo pode interagir negativamente com determinado ambiente, mas, ainda assim, apresentar alto rendimento nessa condição ambiental.

Comparando as metodologias AMMI e GGE Biplot na formação de megaambientes, CAMARGO-BUITRAGO et al. (2011) relataram 95,2% de coincidência nos resultados entre as metodologias. Assim, o modelo GGE biplot é mais indicado para a identificação de megaambientes, seleção de ambientes representativos e discriminativos e indicação de cultivares mais adaptadas e estáveis a ambientes específicos (YAN et al., 2007; GAUCH et al., 2008; ASFAW et al., 2009; YAN, 2011). Por outro lado, a análise AMMI pode ser utilizada com eficiência na identificação de condições ambientais superiores para a exploração agrícola (seleção de locais de cultivo) e genótipos de superior desempenho médio (GAUCH et al., 2008; YAN, 2010). Vários estudos recentes (YAN; HOLLAND, 2010; CAMARGO-BUITRAGO et al., 2011; YAN, 2011; GAUCH, 2013; RODRIGUES, et al., 2014) demonstraram que ambas as metodologias, quando utilizadas adequadamente, são eficientes na definição de mega-ambientes.

### 4.5.2 Comparação dos modelos para o 1º conjunto de dados MET

Para este conjunto de dados MET de produtividade de milho, provenientes da empresa Criagene SK que fica no município de Casa Branca no estado de São Paulo, o experimento conta com 15 genótipos em 4 ambientes com 2 blocos. Na Tabela 16, são apresentados os valores singulares dos quatro componentes principais das análise AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot (GGE biplot-ponterado). Neste conjunto de dados, esses valores foram fundamentais para fazer a comparação entre os modelos.

Tabela 16 - Valores singulares dos quatro componentes principais (PCs) para os modelos AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot

PC	$AMMI(V.S.)^1$	$AMMI-ponderado (V.S.)^1$	GGE biplot $(V.S.)^1$	W-GGE biplot $(V.S.)^1$
1	1561,9546	0,00097	21	23,674
2	981,3147	0,00046	16,601	18,948
3	712,2389	0,00031	15,634	11,385
4	5,91E-06	4,70E-12	12,619	8,42

(V.S.)<sup>1</sup>: Valores singulares

A Tabela 17 é uma tabela de dupla entrada ou uma matriz de correlação entre os valores singulares dos componentes principais dos quatro modelos (AMMI; AMMI-ponderado; GGE biplot e W-GGE biplot), dada da seguinte forma:

Tabela 17 - A matriz de correlação dos valores singulares de PCs dos modelos AMMI, AMMIponderado, GGE biplot e W-GGE biplot

	AMMI	AMMI-ponderado	GGE biplot	W-GGE biplot
AMMI	1,0000	0,9791	0,9834	0,9499
AMMI-ponderado	0,9791	1,0000	0,9991	0,9494
GGE biplot	0,9834	0,9991	1,0000	0,9406
W-GGE biplot	0,9499	0,9494	0,9406	1,0000

Na comparação das metodologias AMMI e GGE biplot existe uma forte correlação (r > 95) de PCs entre genótipo e a produtividade média em todos os ambientes (YAN et al., 2001; CROSSA; CORNELIUS; YAN, 2002; YAN; RAJCAN, 2002) e a correlação foi alta também entre as metodologias (AMMI vs GGE biplot): r=0.974 (CAMARGO-BUITRAGO et al., 2011). Neste trabalho, pela Tabela 17, apresentou-se a correlação entre AMMI-ponderado vs GGE biplot ( $r_1=0.9991$ ), ou seja, as duas metodologias foram altamente correlacionadas, isso mostra também a coincidência entre os resultados. A correlação AMMI vs GGE biplot:  $r_2=0.9834$ ; A correlação de AMMI vs AMMI-ponderado:  $r_3=0.9791$ ; AMMI com W-GGE biplot:  $r_4=0.9499$  é muito próxima dos valores da correlação da AMMI-ponderado vs W-GGE biplot ( $r_5=0.9494$ ) e W-GGE biplot vs GGE-biplot ( $r_6=0.9406$ ). Pelos valores da Tabela 17 (matriz de correlação), verificou-se que as quatro metodologias são muito semelhantes em relação a análise de dados MET, pois essas

metodologias (AMMI, AMMI-ponderado, GGE, W-GGE biplot) são altamente correlacionadas entre si.

Tabela 18 - Proporções explicadas e IR dos componentes principais dos modelos AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot

PC	AMMI-(%1)	IR	AMMI-p-(%1)	IR	GGE - $(\%^1)$	IR	W-GGE -(%1)	IR
1	62,4	2,49	75,5	3,02	39,35	1,57	50,04	2,00
2	24,6	1,00	16,7	0,67	24,61	1,00	32,06	1,28
3	13,0	0,52	7,8	0,31	21,82	0,87	11,57	0,46
4	0	0	0	0	14,22	0,57	6,33	0,25

(%1): Variação Explicada pelos os modelos

A Tabela 18 apresentou as variações explicadas pelos componentes principais dos modelos AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot e a relação de informação (IR). Pelo critério dos PCs que contêm os padrões da seguinte forma: (IR  $\geq$  1), o PC1 dos quatro modelos apresentaram valores maiores do que 1, o AMMI-ponderado apresentou o maior valor (IR): 3,02 e GGE biplot apresentou menor valor (IR): 1,57. Portanto, o PC1 do AMMI-ponderado explicou maior proporção da variação dos dados e o menor deles foi do GGE biplot.

O primeiro componente sempre é o mais importante para a avaliação dos genótipos e ambientes e isto pode ser mostrado pela Tabela 18 com os valores do IR. Portanto, na interpretação dos biplot dos quatro modelos, sempre é importante considerar que o PC1 que apresenta maior variação do total. O biplot na abscissa apresenta os escores de PC1 e na ordenada os escores do PC2, dos genótipos e ambientes. Portanto, o biplot é considerado adequado para representar os padrões nos dados.

Na análise AMMI, os dois primeiros componentes (PC1 e PC2) explicaram 87% da variação total da GEI, já na análise AMMI-ponderado, os dois primeiros componentes (PC1 e PC2) explicaram 92,2% da variação total da GEI. Na análise GGE biplot, os dois primeiros componentes explicaram apenas 63,96% da variação total do efeito do genótipo mais interação genótipo × ambiente (GGE) e na análise W-GGE biplot, os dois primeiros componentes explicaram 82,1% da variação total do GGE.

O GGE biplot foi o modelo que mais prejudicou com a heterogeneidade dos ambientes, portanto, foi o que menos explicou a variação total da interação. O AMMI-ponderado foi o melhor modelo para este conjunto de dados, explicou a maior variação com os dois primeiros componentes, portanto, as interpretações dos gráficos de biplot são as mais confiáveis para representar os padrões nos dados sobre genótipos e ambientes.

#### 4.6 Análise AMMI e Biplot para 2º conjunto de dados MET

#### 4.6.1 Análise de variância conjunta para dois fatores

Os dados utilizados foram obtidos pelo CIMMYT (INTERNATIONAZ MAIZE AND WHEAT IMPROVEMENT CENTER) em experimentos realizados em 7 países, caracterizando-se, assim, experimentos multi-ambientais. Foram nove genótipos de milho, cada genótipo foi avaliado em sete ambientes com 4 blocos, caracterizando assim, um delineamento aleatorizado em blocos.

A análise de variância conjunta dos dados considerando 9 genótipos em 7 ambientes com 4 blocos é apresentada na Tabela 19. Os resultados da análise de variância conjunta relativa à produtividade de milho  $(kg\ ha^{-1})$  mostraram que os efeitos de genótipos (G), ambientes (E) e GEI são significativos (p < 0,01). Esses resultados indicam que os genótipos apresentaram comportamento diferenciado nos ambientes, o que justifica estudo mais aprofundado sobre o comportamento dos genótipos para identificar suas magnitudes de interação com os ambientes. Um dos principais resultados de interesse apresentado na tabela é a soma de quadrados da interação  $(SQ_{GEI} = 54796176)$ , que representa 16,41% da variabilidade total dos dados.

Tabela 19 - Análise de variância conjunta para um conjunto de dados com 9 genótipos avaliados em 7 ambientes com 4 blocos

Fonte de Variação	$\operatorname{GL}$	SQ	QM	F	Valor-p
Blocos/Ambiente	21	13888180	661342	1,2411	0,2234284
Ambiente (E)	6	153607395	25601233	38,7110	2,716e-10 ***
Genótipo (G)	8	22163552	2770444	5,1991	8,162e-06 ***
GEI	48	54796176	1141587	2,1423	0,0002007 ***
Resíduo	168	89522977	89522977	-	-
Total	251	333978280	-	-	-

NS: Não-significativo; \*\*Significativo a 0,01 de probabilidade.

O coeficiente de variação (CV%) do experimento pela análise de variância conjunta mostrou-se de baixa magnitude (14,92%), permitindo inferir que os experimentos foram bem conduzidos. As estimativas das médias dos genótipos e dos ambientes, com relação a produtividade

<sup>\*\*\*</sup>Significativo a 0,001 de probabilidade.

em  $kg\ ha^{-1}$ , são apresentados na Tabela 20. O genótipo G6 apresentou a maior produtividade, logo seguidos dos genótipos G5, G4, G3 e G9, respectivamente. Já os genótipos G8, G1 e G2 apresentaram as piores produtividades médias. Com relação aos ambientes, têm-se que o ambiente A6 apresentou a maior produtividade média, seguido do ambiente A4 e as piores produtividades médias foram dos ambientes A1 e A2, respectivamente.

Tabela 20 - Médias dos genótipos, ambientes e posição das médias em relação a produtividade

Genótipo	Média	Posição	Ambiente	Média	Posição
$G_1$	4696,146	8	$A_1$	3612,531	7
$G_2$	4696,939	7	$A_2$	4209,444	6
$G_3$	4776.596	4	$A_3$	5104,919	3
$G_4$	5167,014	3	$A_4$	5229,997	2
$G_5$	5331,161	2	$A_5$	4955,231	4
$G_6$	5396,304	1	$A_6$	6305,306	1
$G_7$	4704,307	6	$A_7$	4837,614	5
$G_8$	4554,346	9			
$G_9$	4719,382	5			

Tabela 21 - Valores estimados da interação dupla de 9 genótipos e 7 ambientes para a produção em  $kg\ ha^{-1}$ 

Ambiente/	A1	A2	A3	A4	A5	A6	A7
Genótipo							
G1	206,45	-283,69	646,54	-466,99	-377,99	328,85	-53,16
G2	10,56	-94,16	28,39	-70,78	442,88	-72,39	-244,50
G3	-49,69	-10,34	129,09	23,26	-660,33	270,48	297,54
G4	-166,07	56,29	-835,86	526,07	443,58	99,06	-123,08
G5	-885,49	-568,18	630,02	163,42	20,74	138,81	500,68
G6	0,82	165,38	-402,82	247,18	133,39	107,92	-251,87
G7	-69,24	746,85	473,38	-698.45	-488,79	$628,\!52$	-592,27
G8	1255,48	-476,96	482,39	-449,27	-139,59	-979,83	307,79
G9	-302,81	464,80	-1151,12	725,57	626,12	-521,41	158,86

Considerando a equação (22) aplicada à matriz de médias da equação (21), obtémse a matriz de interação GE, apresentada na Tabela 21. A próxima etapa da análise AMMI corresponde ao ajuste da interação pela decomposição em valores singulares ( $GE = USV^T$ ) aplicado à matriz de interação GE. Esta matriz tem posto igual p = min(8,6) = 6, portanto a  $SQ_{GEI}$  pode ser decomposta em até 6 componentes principais.

A matriz U contêm os vetores singulares à esquerda da interação, S é uma matriz diagonal contendo os valores singulares e V é uma matriz cujas colunas contêm os vetores singulares à direita da interação e são apresentadas da seguinte forma:

$$S = \begin{pmatrix} 2571,413 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1927,729 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1531,839 & 0 & 0 & 0 \\ 0 & 0 & 0 & 819,006 & 0 & 0 \\ 0 & 0 & 0 & 0 & 583,948 & 0 \\ 0 & 0 & 0 & 0 & 0 & 111,539 \end{pmatrix}$$

$$\boldsymbol{V} = \begin{pmatrix} 0,262 & 0,621 & -0,501 & -0,113 & -0,374 & -0,003 \\ -0,171 & -0,292 & -0,577 & -0,198 & 0,608 & 0,034 \\ 0,694 & -0,078 & 0,286 & 0,284 & 0,266 & 0,369 \\ -0,487 & 0,022 & 0,272 & -0,239 & -0,219 & 0,663 \\ -0,413 & 0,172 & 0,068 & 0,764 & 0,059 & -0,257 \\ 0,116 & -0,666 & -0,053 & -0,021 & -0,564 & -0,281 \\ 0 & 0,222 & 0,504 & -0,476 & 0,224 & -0,527 \end{pmatrix}$$

$$\boldsymbol{U} = \begin{pmatrix} 0,262 & 0,621 & -0,501 & -0,113 & -0,374 & -0,003 \\ -0,171 & -0,292 & -0,577 & -0,198 & 0,607 & 0,034 \\ 0,694 & -0,078 & 0,286 & 0,284 & 0,266 & 0,369 \\ -0,413 & -0,031 & -0,054 & 0,049 & -0,512 & -0,309 \\ 0,089 & -0,211 & 0,811 & 0,155 & 0,262 & 0,148 \\ -0,487 & 0,022 & 0,272 & -0,239 & -0,219 & 0,664 \\ -0,413 & 0,172 & 0,068 & 0,764 & 0,059 & -0,257 \\ 0,116 & -0,666 & -0,053 & -0,021 & -0,564 & -0,281 \\ 0 & 0,222 & 0,504 & -0,476 & 0,224 & -0,527 \end{pmatrix}$$

Na análise AMMI, a soma de quadrados da interação  $(SQ_{GEI})$  corresponde exatamente à soma dos autovalores ou quadrados dos valores singulares que estão na diagonal da matriz  $\boldsymbol{S}$   $(\sum_{k=1}^{n} \lambda_k^2)$  (Tabela 22), que pode estar inflacionado devido à presença de ruídos (variação não explicável) na variável resposta. Em virtude disto, faz-se necessário realizar um ajuste da interação por decomposição por valores singulares, aplicada à matriz de GEI  $(\boldsymbol{GE})$ . Esta matriz tem posto p=6, ou seja, pode ser decomposta em seis componentes principais, que são a soma de quadrados parciais na análise de variância.

Os dois primeiros componentes (PC1 e PC2) foram significativos ao nível de 0,1% de probabilidade pelo teste F de Gollob (1968) na decomposição da GEI dos experimentos pelo modelo AMMI (Tabela 22). O PC1 explicou 48,3% da variação total da  $SQ_{GEI}$  e o PC2 explicou 27,1%, consideradas uma resposta padrão presente na  $SQ_{GEI}$ .

A Tabela 22 mostra que PC3 a PC6 são não significativos, portanto, são desprezíveis e contém apenas ruído, que podem diminuir a eficiência da interpretação da estabilidade dos genótipos e ambientes na análise gráfica. Assim, a interpretação gráfica, considerando a variação contida nos dois primeiros eixos das componentes principais foram suficientes para avaliar a estabilidade e adaptabilidade dos genótipos e também dos ambientes de testes.

Tabela 22 - Porcentagem do desdobramento da soma de quadrados da interação ( $G \times E$ ) captada por componente principal (PC)

PCs	$\operatorname{GL}$	Proporção	Proporção	SQ	QM	$\mathbf{F}$	Valor p
		$SQ_{PC}/SQ_{GEI}$	Acumulada(%)				
PC1	13	0,483	48,3	26448664,61	2034512,66	3,82	0,0000 ***
PC2	11	0,271	75,4	14864569,21	1351324,47	2,54	0,0054 **
PC3	9	0,171	92,5	9386117,29	1042901,92	1,96	0,0469
PC4	7	0,049	97,4	2683083,01	383297,57	0,72	0,6552
PC5	5	0,025	99,9	49763,29	16587,76	0,51	0,7684
PC6	3	0,1	100,0	49763,29	16587,76	0,03	0.9930
Total	48	1,00	-	54796176	-	-	-

NS: Não-significativo; \*\*Significativo a 0,01 de probabilidade.

Os resultados para o desdobramento da GEI, correspondente a cada membro da família de modelos AMMI, ou seja, os modelos AMMI0, sendo toda a GEI, com 48 GL; modelo AMMI1 é a soma de quadrados atribuída ao primeiro eixo do componente principal (PC1) e o

<sup>\*\*\*</sup>Significativo a 0,001 de probabilidade.

restante da interação, depois de retirados os 13 GL; modelo AMMI2 com 11 GL, a soma de quadrados atribuída ao segundo eixo do componente principal (PC2); modelo AMMI3, com 9 GL para PC3, até o AMMI6 com 3 GL para PC6 (Tabela 22).

Na avaliação da família dos modelos AMMI (AMMI0, AMMI1, AMMI2, AMMI3, AMMI4, AMMI5 e AMMI6), verificou-se pelo teste F, com os graus de liberdade ajustados pelo método de Gollob (1968), os dois primeiros componente principal da interação (PC1 e PC2) foram significativos, o que levaria à seleção dos modelos AMMI1 e AMMI2 para explicar a interação de forma significativa e apresentaram o melhor padrão de resposta para os genótipos aos diferentes ambientes (Tabela 22). Nesse caso, é possível a construção de dois gráficos biplot: i) AMMI1 - Média da produtividade dos genótipos vs PC1 (Figura 18); ii) AMM2 - PC1 vs PC2 (Figura 19).

Tabela 23 - Coordenadas dos eixos das abscissas (x) e de coordenadas (y), para os dois gráficos Biplot

Genótipos	Biplo	t AMMI1	Biplot AMMI2		
Ambientes	Média $(x)$	PC1(y)	PC1(x)	PC2(y)	
G1	4696,146	-19,181	-19,181	-3,305	
G2	4696,939	2,331	2,331	2,285	
G3	$4776,\!596$	-7,315	-7,315	-6,032	
G4	$5167,\!014$	20,918	20,918	-1,372	
G5	$5331,\!161$	-4,543	-4,543	-9,266	
G6	$5396,\!304$	9,278	9,278	-2,642	
G7	$4704,\!307$	-15,719	-15,719	-21,580	
G8	$4554,\!346$	-17,904	-17,904	$35{,}729$	
G9	4719,382	32,135	32,135	6,184	
A1	$3612,\!531$	-13,275	-13,275	27,260	
A2	$4209,\!444$	8,683	8,683	-12,838	
A3	5104,919	-35,168	-35,168	-3,412	
A4	$5229,\!997$	24,701	24,701	0,965	
A5	$4955,\!231$	20,937	20,937	7,534	
A6	$6305,\!306$	-5,874	-5,874	-29,250	
A7	4837,614	-0,003	-0,003	9,741	

Os gráficos dos biplots das Figuras 18 e 19 (AMMI1 e AMMI2) foram utilizados para analisar a dispersão da variação dos genótipos e ambientes e da interação entre estes. Como

já mencionado, o biplot AMMI1 contém a variação dos efeitos aditivos principais de genótipos e ambientes, visualizado na horizontal do gráfico (eixo das abscissas) e a variação dos efeitos multiplicativos da GEI, visualizado na vertical (eixo das ordenadas). O biplot AMMI2 é visualizado apenas os efeitos multiplicativos da GEI contidos nos dois primeiros PCs e dispersos em ambas as ordenadas do gráfico.

Os genótipos foram classificados de acordo com sua produtividade média da seguinte forma: G6 > G5 > G4 > média geral dos genótipos > G3 > G9 > G7 > G2 > G1 > G8; A classificação dos ambientes são: A6 > A4 > A3 > A5 > média geral dos ambientes > A7 > A2 > A1 (Figura 18; Tabela 23). O G6 foi o genótipo que apresentou maior produtividade e o A6 foi o melhor entre os ambientes de testes.

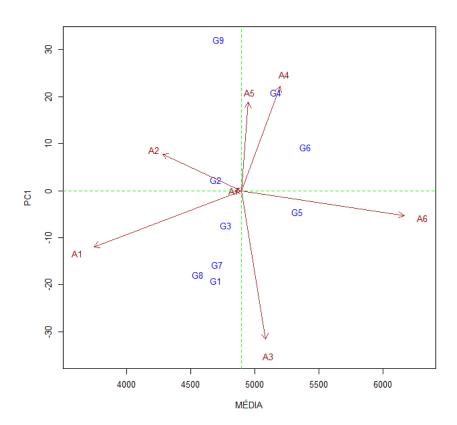


Figura 18 - Biplot AMMI1 (Médias vs PC1) para os dados de produtividade de milho  $(kg \ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)

Em relação às adaptações específicas dos genótipos em ambientes, verifica-se que os genótipos G7, G1, G3 e G5 foram mais adaptados aos ambientes A3 e A6. O G7 deve ser o genótipo que apresentou maior produtividade média em A3 e A6. O genótipo G8 tem uma adaptação específica em A1, e o G9 em A4 e A5. Observa-se que os vetores relativos a esses

genótipos apontam para uma direção comum, que pode ser orientada pelos ambientes. Além disso, constata-se falta de adaptação do genótipo G8 aos ambientes A2 e A6; do genótipo G1 aos ambientes A4 e A5; dos genótipos G9 e G4 ao ambiente A3 (Figura 19).

O ambiente A7 foi altamente estável (Figuras 18 e 19), mas não foi o "ideal" ambiente de teste, pois não apresentou bom desempenho em relação a produtividade. Os ambientes A4 e A3 foram altamente instável, mas, em relação a produtividade foram o segundo e terceiro melhores entre todos os ambientes de testes. O A6 foi estável e também o melhor ambiente em relação a produtividade, portanto, o A6 foi o "ambiente de teste ideal" (produtividade e estabilidade). Em contra partida, o pior ambiente de teste em relação a produtividade e estabilidade foi o A1 (Figuras 18 e 19).

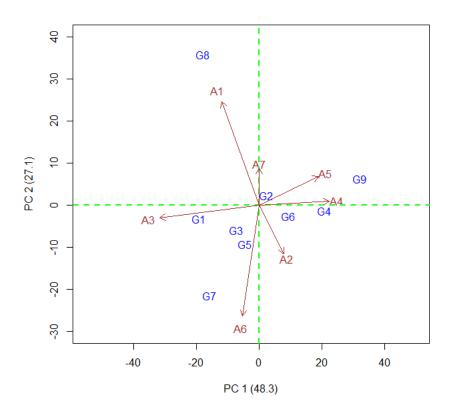


Figura 19 - Biplot AMMI2 (PC1 vs PC2) para os dados de produtividade de milho  $(kg \ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)

O genótipo G8 foi altamente instável, pois teve rendimento inferior ao esperado nas ambientes A2 e A6, enquanto produziu relativamente bem em A1 e A7 (Figura 19); O G7 e G9 não foram estáveis, mas o G7 produziu bem nos ambientes A3, A6 e G9 nos ambientes A4 e

A5. Os genótipos mais estáveis foram G2, G6, G3 e G5 (Figura 18; Tabela 23), estes podem ser cultivados em todos os locais onde foram feitos os estudos. Entre estes, apenas os genótipos G6 e G5 tiveram produtividade acima da média geral e foram o primeiro e segundo melhores genótipos em relação a produtividade, respectivamente.

O genótipo G2 foi o mais estável, mas não foi o genótipo "ideal" (Estabilidade e produtividade) por causa da sua produtividade abaixo da média. Os genótipos ideais foram G6 e G5, que podem ser recomendados para os melhoristas (Figura 18 e 19). Note-se que, se o biplot explica apenas uma pequena proporção da variação total, alguns genótipos aparentemente estáveis podem não ser verdadeiramente estáveis como as suas variações não podem ser completamente explicadas neste biplot. As Figuras 18 e 19 ilustram um conceito importante em relação à estabilidade, o termo "estabilidade elevada" só tem sentido quando associado ao desempenho médio; o genótipo estável é desejado apenas quando apresenta alta performance média (YAN; TINKER, 2006; YAN, 2011).

### 4.7 GGE biplot para análise de 2º conjunto dados MET

Utilizando os dados da produtividade média  $(kg\ ha^{-1})$  de cada genótipo em cada ambiente do CIMMYT (INTERNATIONAZ MAIZE AND WHEAT IMPROVEMENT CENTER) em experimentos realizados em 7 países utilizando a análise GGE biplot, tem-se (Tabela 24):

Tabela 24 - Médias dos nove genótipos (G1 a G9) testadas em sete ambientes (A1 a A7) em relação a produtividade média  $(kg \ ha^{-1})$ 

Genótipo	Ambientes						
	A1	A2	A3	A4	A5	A6	A7
G1	3621,55	3728,33	5554,03	4565,58	4379,8	6436,73	4587,03
G2	3426,45	3918,65	4936,68	4962,58	5201,48	6036,28	4396,48
G3	3445,85	4082,13	5117,03	5136,28	4177,93	6458,8	5018,18
G4	3719,9	4539,18	4542,5	6029,5	5672,25	6677,8	4987,98
G5	3164,63	4078,85	6172,53	5831	5413,55	6881,7	5775,88
G6	4116,08	4877,55	5204,83	5979,9	5591,35	6915,95	5088,48
G7	3354,03	4767,03	5389,03	4342,28	4277,18	6744,55	4056,08
G8	4528,78	3393,25	5248,08	4441,5	4476,4	4986,25	4806,18
G9	3135,53	4500,05	3779,6	5781,38	5407,15	5609,7	4822,28

De acordo com relação de informação (IR) dos sete componentes (Tabela 25), apenas os três primeiros PCs contêm padrões (IR  $\geq$  1). Portanto, o biplot é considerado adequado para representar os padrões (IR  $_{PC1} >$  IR  $_{PC2} >$  1). O GGE biplot com base neste conjunto de dados é apresentado na Figura 20, na abscissa do biplot apresentam-se os escores de PC1 e na ordenada os escores do PC2, dos genótipos e ambientes. O biplot explicou 64,50% da variação total G + GE e foi construído no ambiente centrado (centralização = 2), não-dimensionado (escala = 0) e particionamento em valor singular com foco no meio ambiente (PVS = 2).

Tabela 25 - Valor singular, proporção explicada e relação da informação (IR) dos quatro componentes principais (PCs)

РС	Valor singular	Variação explicado (%)	IR
1	13,606	44,08	3,086
2	9,261	20,42	1,429
3	8,397	16,79	1,175
4	6,963	11,54	0,808
5	4,965	5,87	0,411
6	2,171	1,12	0,078
7	0,877	0,18	0,013

Os nove genótipos são rotulados como G1 a G9 e os sete ambientes como A1 a A7. A Figura 20 pode ser utilizada para classificar os genótipos com base no desempenho, em qualquer ambiente e também para classificar ambientes sobre o desempenho relativo de qualquer genótipo. Por exemplo, os genótipos G2 e G8 formaram ângulos obtusos com todos os ambientes de testes (A1 a A7), isto significa que esses genótipos tiveram menor produtividades médias em todos os ambientes. Essa informação pode ser confirmada pela Tabela 24. O genótipo G8 apresentou vetor maior do que G2, portanto, devem ter produtividade mais baixa que G2 em todos os ambientes.

Como outro exemplo, os genótipos G6 e G5 formaram ângulos agudos com todos os ambientes, portanto, esses genótipos produziram mais do que a média em todos os ambientes (Figura 20). O genótipo G6 apresentou maior vetor do que G5, portanto, o G6 deve ter maior produtividade em todos os ambientes. Isto também pode ser confirmado a partir da Tabela 24. Desvios desta previsão podem ocorrer, porque os dois genótipos não têm exatamente os mesmos ângulos com os ambientes e porque a qualidade do ajuste do biplot é de 64,50% da variação total de G + GE, em vez de 100%.

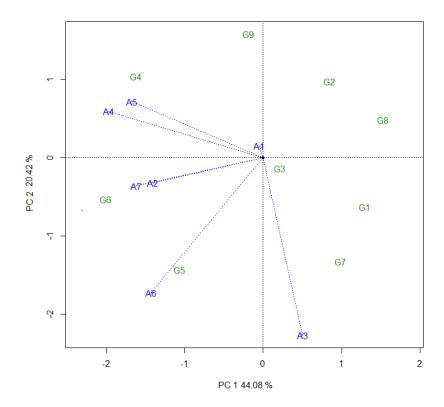


Figura 20 - O GGE Biplot para os dados de produtividade de milho  $(kg \ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)

### 4.7.1 Análise Mega-ambiente (2º conjunto de dados)

O GGE biplot de dimensão 2 é como uma aproximação suficiente dos dados (Tabela 24), o GGE biplot (Figura 21) (também é chamado como "Which-won-where") é uma ferramenta eficaz para análise de mega-ambiente. Na Figura 21 é possível ver o agrupamento visual dos ambientes de teste com base em cruzamento G × E entre os melhores genótipos. Os vértices do polígono são formados pelos genótipos: G5, G7, G1, G8, G2, G9, G4 e G6. Os sete ambientes foram cortados em 4 grupos pelas linhas que saíram da origem do biplot, formando os seguintes grupos i) A3; ii) A1, A5 e A4; iii) A2 e A7; iv) A6.

O genótipo G6 é o vértice do setor em que os ambientes A7 e A2 são colocados, portanto, é o genótipo que teve melhor desempenho nestes ambientes; o G4 é o vértice no setor em que A4, A5 e A1 são colocados, logo, é o genótipo mais adaptado nestes ambientes. Da mesma forma, verificou-se que G5 é o genótipo mais produtivo no A6 e também G7 em A3 (Figura 21).

Nos setores de G1, G8, G2 e G9 que não contém ambientes, isto significa que

estes genótipos não eram produtivos em nenhum ambiente, ou seja, estes genótipos são os piores genótipos em relação à produtividade em alguns ou em todos os ambientes. O padrão da GE cruzado sugere que os ambientes alvos podem ser divididos em quatro mega-ambientes diferentes. Este padrão parece ser repetitivo ao longo dos anos e o delineamento mega-ambiente é, portanto, significativo.

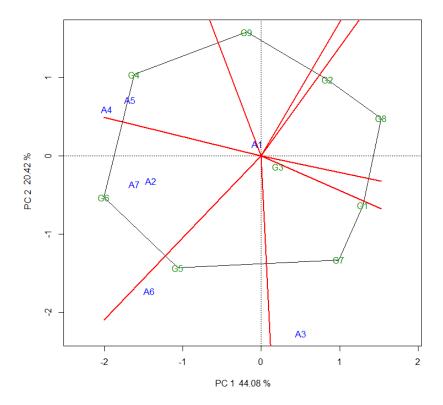


Figura 21 - O GGE Biplot ("Which-won-where") mostra que para os dados de produtividade de milho  $(kg\ ha^{-1})$ , quais dos genótipos apresentaram melhor desempenho em que ambientes

### 4.7.2 Avaliação de genótipos baseada em GGE biplot (2º conjunto de dados)

Na Figura 22, o pequeno círculo representa o "ambiente-média" e é definido pelas coordenadas médias de todos os ambientes de teste no biplot. Este biplot é baseada em ("particionamento em valor singular centrada-genótipo" ou PVS = 1), ou seja, os valores singulares são alocados inteiramente nos escores de genótipo (linha) com s = 1 (YAN, 2002; YAN, 2011).

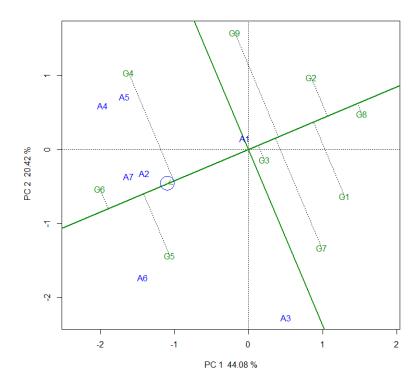


Figura 22 - O GGE Biplot ("Média versus Estabilidade") com eixo do ambiente-média (EAM) para mostrar o desempenho médio e estabilidade dos genótipos para os dados de produtividade de milho  $(kg\ ha^{-1})$ 

Assim, os genótipos são classificados de acordo com sua produtividade média da seguinte forma: G6 > G5 > G4 > média geral > G3 > G7 > G9 > G1 > G2 > G8. Os genótipos G4, G9 e G7 foram altamente instáveis; o G4 teve produtividade bem acima da média (terceiro melhor), mas teve rendimento inferior ao esperado em ambiente A3 enquanto produziu bem acima da média nos ambientes A4, A5; o G9 teve produtividade abaixo da média em todos os ambientes de teste e o G7 produziu bem apenas no ambiente A3 e abaixo da média entre restantes dos ambientes (Figura 22; Tabela 24).

Os genótipos G3 e G8 foram bastante estáveis, mas apresentaram produtividade abaixo da média, principalmente o G8 foi o mais estável, mas também o pior genótipo em relação a produtividade média nos ambientes. O G5 foi o segundo melhor genótipo em relação a produtividade, mas teve rendimento bem abaixo da média no ambiente A1 (Figura 22; Tabela 24).

O Genótipo G6 foi altamente estável e também o melhor genótipo em relação a produtividade média em todos os ambientes de testes (Figura 22). Nota-se que, o biplot explicou 64,50% da variação total de G + GE, por isso, alguns genótipos aparentemente estáveis podem não ser verdadeiramente estáveis como as suas variações podem não ser completamente explicadas (100%) neste biplot.

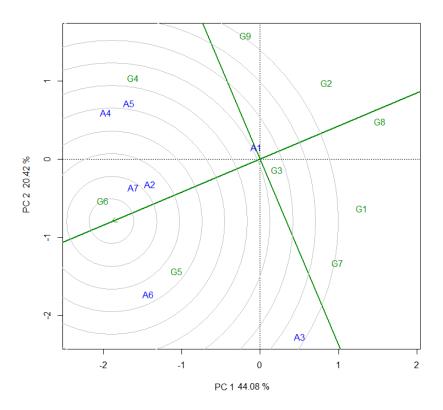


Figura 23 - O GGE Biplot com eixo do ambiente-média (EAM) para classificar os genótipos em relação ao genótipo ideal (no centro dos círculos concêntricos)

A Figura 23, mostrou-se que os genótipos localizados mais próximos do "genótipo ideal" (o centro dos círculos concêntricos) são mais desejáveis do que outros. Assim, como esperado, o G6 foi o genótipo "ideal" neste conjunto de dados e deve ser selecionado para fazer a recomendação para os melhoristas.

Em sequência, os melhores genótipos foram G5, G4, G3, G9 e G7, respectivamente. Os piores genótipos para recomendação foram G8, G2 e G1. A Figura 23 ilustra um conceito importante em relação a estabilidade, o termo "estabilidade elevada" só tem sentido quando associado ao desempenho médio; o genótipo estável é desejado apenas quando apresenta alta performance média (YAN; TINKER, 2006; YAN, 2011).

### 4.7.3 Avaliação do Ambiente de Teste (2º conjunto de dados)

Como já citado, os ambientes de teste com vetores mais longos são mais discriminantes em relação aos genótipos. Aqueles ambientes com um vetor curto são menos discriminantes, o que significa que todos os genótipos tendem a executar de forma semelhante e pouca ou nenhuma

informação sobre as diferenças genotípicas podem ser reveladas em tal ambiente, portanto, não devem ser utilizados como ambientes de teste.

Na Figura 24, pode ser visto que todos os ambientes apresentaram vetores longos em relação ao origem do biplot, exceto A1. Isto significa que estes ambientes são mais discriminantes em relação aos genótipos, principalmente os ambientes A3 e A6 que possuem os vetores maiores. O A1 apresentou um vetor muito curto a partir da origem do biplot, portanto, A1 é menos discriminante em relação aos genótipos, por isso, não deve ser utilizado como ambiente de teste.

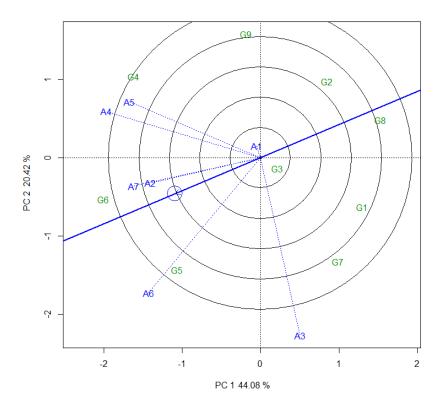


Figura 24 - O GGE biplot "discriminação e representatividade" para mostrar a capacidade discriminação e de representatividade dos ambientes de teste

Uma segunda utilidade da Figura 24 é indicar ambientes de representatividade do ambiente alvo. Os ambientes de teste que têm pequenos ângulos com EAM (eixo do ambientemédia), por exemplo, A2 e A7, são mais representativos do ambiente alvo do que aqueles que têm ângulos maiores, como por exemplo, A1 e A3. Portanto, os ambientes de teste que são discriminantes e representativos (por exemplo, A2 e A7) são ambientes de teste ideais para a seleção de genótipos adaptados.

Os ambientes de teste discriminante mas não representativos (por exemplo, A4, A5,

A6 e A3) são úteis para a seleção de genótipos adaptados especificamente se os ambientes alvos podem ser divididos em mega-ambientes; ou para seleção de genótipos instáveis se o ambiente alvo é um único mega-ambiente (Figura 24).

Ressalta-se que, dentro de um único mega-ambiente, o ambiente de teste ideal deve ser mais discriminante e também mais representativo do ambiente alvo. Um "ambiente de teste ideal" é o centro dos círculos concêntricos (Figura 25). Este é um ponto da EAM no sentido positivo, com uma distância da origem biplot igual ao vetor mais longo de todos os ambientes, sendo assim, A7 é mais próximo a este ponto, portanto, é o ambiente de teste ideal e o ambiente A2 é o segundo melhor seguidos dos ambientes A6, A4, A5, A1 e A3, respectivamente. O ambiente A3 foi o pior para selecionar genótipos adaptados a toda região (Figura 25).

Análise GGE biplot para os dados de MET pode ajudar os pesquisadores a entender melhor seu ambiente alvo, para estabelecer estratégias e testes de melhoramento de baixo custo, e identificar genótipos superiores que são amplamente ou especificamente adaptados (YAN, 2011).

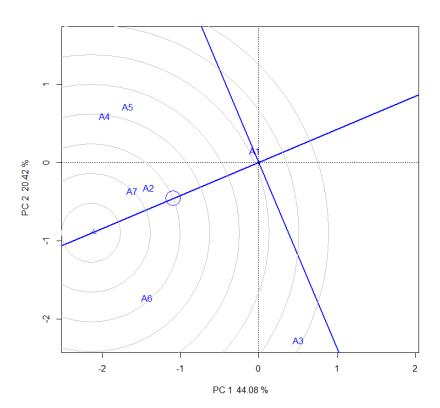


Figura 25 - O GGE biplot discriminação e representatividade para classificar ambientes de teste em relação a um ambiente de teste ideal (representado por centro dos círculos concêntricos)

## 4.8 Análise AMMI-ponderado e Biplot para $2^o$ conjunto dados multi-ambientes (MET)

Com a análise AMMI-ponderado para os dados MET, que permite levar em consideração a heterogeneidade da variância do erro ao longo dos ambientes, os dados centrados nos ambientes foram obtidos dividindo-se a variância do erro de cada ambiente a partir do valor de rendimento original de cada genótipo em cada ambiente. As estimativas das médias dos genótipos e dos ambientes centrados em relação aos ambientes são apresentados na Tabela 26. A análise de variância conjunta dos dados do CIMMYT (INTERNATIONAZ MAIZE AND WHEAT IM-PROVEMENT CENTER) em experimentos realizados em 7 países com 9 genótipos de milho em quatro ambientes com 4 blocos é a mesma utilizando AMMI ponderado e o AMMI (Tabela 19).

A variância do erro experimental de cada ambiente ou quadrados médios dos resíduos de cada ambiente ( $s_j^2$ ) pela análise da variância conjunta são (A1:  $s_1^2 = 920643,85$  ( $kg\ ha^{-1}$ )<sup>2</sup>; A2:  $s_2^2 = 301311,49\ (kg\ ha^{-1}$ )<sup>2</sup>; A3:  $s_3^3 = 1156702,04\ (kg\ ha^{-1}$ )<sup>2</sup>; A4:  $s_4^2 = 252194,9$  ( $kg\ ha^{-1}$ )<sup>2</sup>; A5:  $s_5^2 = 419953,69\ (kg\ ha^{-1}$ )<sup>2</sup>; A6:  $s_6^2 = 429636,00\ (kg\ ha^{-1}$ )<sup>2</sup> e A7:  $s_7^2 = 349600,84$  ( $kg\ ha^{-1}$ )<sup>2</sup>).

O genótipo G6 apresentou maior valor da estimativa da produtividade, seguidos dos genótipos G5, G4, G3 e G9, respectivamente. Já os genótipos G2, G8 e G1 apresentaram as piores estimativas da produtividade. Com relação aos ambientes, têm-se que o ambiente A7 apresentou a maior estimativa de produtividade média e a pior foi o ambiente A4 (Tabela 26).

Tabela 26 - Médias dos genótipos, ambientes e posição das médias em relação a produtividade (AMMI-ponderado)

Genótipo	Média-centrado	Posição	Ambiente	Média-centrado	Posição
$G_1$	0,0145	7	$A_1$	0,0174	2
$G_2$	0,0144	9	$A_2$	0,0171	3
$G_3$	0,0148	4	$A_3$	0,0115	6
$G_4$	0,0159	3	$A_4$	0,0110	7
$G_5$	0,0162	2	$A_5$	0,0131	5
$G_6$	0,0167	1	$A_6$	0,0149	4
$G_7$	0,0145	6	$A_7$	0,0207	1
$G_8$	0,0144	8			
$G_9$	0,0146	5			

Os dados da Tabela 27, se referem aos valores de rendimento centrados nos ambientes de nove genótipos de milho testados em sete ambientes. A partir desta Tabela 27 que é uma matriz de dupla entrada de genótipos e ambientes foi construída a matriz de interação  $\boldsymbol{GE}$  e também os gráficos biplot (Figura 26 e 27).

O PC1 explicou 43,5% da variação total da GEI e o PC2 explicou 26,1% (Figura 27). com isso, os dois componentes explicaram um total de 69,6% da GEI com AMMI-ponderado, enquanto a análise AMMI explicou 75,4% da GEI com o PC1 e PC2 para o mesmo conjunto de dados. Portanto, o biplot PC1 vs PC2 do AMMI-ponderado explicou menor variação de GEI em comparação com biplot AMMI, por isso, as conclusões do biplot AMMI (Figura 18 e 19) são mais confiáveis do que as do biplot AMMI (Figura 26 e 27).

Tabela 27 - Produtividade média com os valores centrados nos ambientes de testes com nove genótipos (G1 a G9) em sete ambientes (A1 a A7) pela análise AMMI-ponderado

Genótipo	Ambientes						
	A1	A2	A3	A4	A5	A6	A7
G1	0,0174	0,0152	0,0125	0,0096	0,0116	0,0153	0,0197
G2	0,0165	0,0159	0,0111	0,0105	0,0138	0,0143	0,0188
G3	0,0166	0,0166	0,0115	0,0108	0,0111	0,0153	0,0215
G4	0,0179	0,0185	0,0102	0,0127	0,0150	0,0158	0,0214
G5	0,0152	0,0166	0,0139	0,0123	0,0144	0,0163	0,0247
G6	0,0198	0,0199	0,0117	0,0126	0,0148	0,0164	0,0218
G7	0,0161	0,0194	0,0121	0,0092	0,0113	0,0160	0,0174
G8	0,0218	0,0138	0,0118	0,0094	0,0119	0,0118	0,0206
G9	0,0151	0,0183	0,0085	0,0122	0,0143	0,0133	0,0207

Analisando a Figura 26, foi possível verificar que o genótipo G6 apresentou maior valor da estimativa da produtividade, seguidos dos genótipos G5, G4 acima da média geral. Já os genótipos G9, G7, G2, G1, e G8 apresentaram os mesmos desempenhos em relação a produtividade (as piores estimativas da produtividade). Da mesma forma, a classificação dos ambientes é dada por: A7 > A1 > A2 > A6 > A5 > A3 > A4, em que A7 foi o melhor entre os ambientes de testes e A4 foi o pior (Figura 26).

Em relação às adaptações específicas dos genótipos em ambientes, observa-se que os genótipos G7, G6 foram mais adaptados aos ambientes A2 e A6; os genótipos G4 e G9 foram mais adaptados no A2, A6, A4 e A5; os genótipos G8, e G1 nos ambientes A3 e A1; os G3 e G5

no ambiente A7 (Figura 27). O genótipo G5 foi o genótipo que apresentou maior produtividade no A7; o G8 no ambiente A1; G7 no ambiente A2; G6 no ambiente A6 e G4 nos ambientes A4 e A5. Essas informações também podem ser confirmadas na Tabela 27. Além disso, constata-se falta de adaptação dos genótipos G5 e G9 ao ambiente A1; do genótipo G8 aos ambientes A2 e A6; do genótipo G7 ao ambiente A7 e G8 aos ambientes A4 e A5 (Figura 27 e Tabela 28).

Tabela 28 - Coordenadas dos eixos das abscissas (x) e de coordenadas (y), para os dois gráficos Biplot

Genótipos	Biplot AM	MI1-ponderado	Biplot AMN	iplot AMMI2-ponderado		
Ambientes	Média $(x)$	PC1 (y)	PC1(x)	PC2(y)		
G1	0,0145	-0,0198	-0,0198	0,0039		
G2	0,0144	0,0019	0,0019	0,0043		
G3	0,0148	-0,0002	-0,0002	-0,0088		
G4	0,0159	0,0171	0,0171	-0,0002		
G5	0,0162	0,0155	0,0155	-0,0516		
G6	0,0167	0,0033	0,0033	0,0154		
G7	0,0145	0,0156	0,0156	0,0528		
G8	0,0144	-0,0711	-0,0711	-0,0055		
G9	0,0146	0,0377	0,0377	-0,0104		
A1	0,0174	-0,0662	-0,0662	0,0242		
A2	0,0171	0,0403	0,0403	0,0412		
A3	0,0115	-0,0258	-0,0258	0,0005		
A4	0,0110	0,0189	0,0189	-0,0154		
A5	0,0131	0,0167	0,0167	-0,0110		
A6	0,0149	0,0186	0,0186	0,0158		
A7	0,0207	-0,0025	-0,0025	-0,0552		

O ambiente A1 foi altamente instável, mas em relação à produtividade foi o segundo melhor classificado. Os ambientes A3, A4 e A5 foram estáveis, mas não tiveram bom desempenho em produtividade. O ambiente A7 foi altamente estável pelo PC1 (Figura 26 e 27) e também foi o primeiro em relação à produtividade, portanto, o A7 foi o "ambiente de teste ideal" (produtividade e estabilidade) pela análise AMMI-ponderado.

Para avaliação de genótipos mais estáveis pelo biplot, o PC1 (43,5%) apresentou

a maior variabilidade total de GEI. O genótipo G8 foi altamente instável, pois teve rendimento inferior ao esperado aos ambientes A6, A2, A4 e A5, enquanto produziu relativamente bem em A1 e A3 (Figura 27). Os genótipos G7, G9 e G5 foram instáveis pelo falta de adaptação em alguns ambientes. Os genótipos G2, G3 e G6 foram altamente estáveis (PC1), podendo ser cultivados em todos os ambientes em que foram feitos os estudos. Entre estes, apenas o genótipo G6 teve produtividade acima da média geral e ainda foi o melhor entre todos os genótipos. Os genótipos G2 e G3 tiveram desempenho muito semelhantes em relação a produtividade e estabilidade. Os genótipos G4 e G5 apresentaram produtividade próximas e maiores que a média geral, o G4 foi estável nos ambientes de testes e G5 foi considerado instável (Figura 27 e Tabela 28).

O genótipo "ideal" (Estabilidade e produtividade) foi o G6, que pode ser considerado como o melhor entre todos os genótipos de testes, portanto, este genótipo pode ser indicado para recomendação de cultivares. Como o biplot AMMI-ponderado explicou 69,6% da variação total de GEI, logo, com AMMI-ponderado para este conjunto de dados, alguns genótipos aparentemente estáveis podem não ser verdadeiramente estáveis como as suas variações não podem ser completamente explicadas (100%) neste biplot.

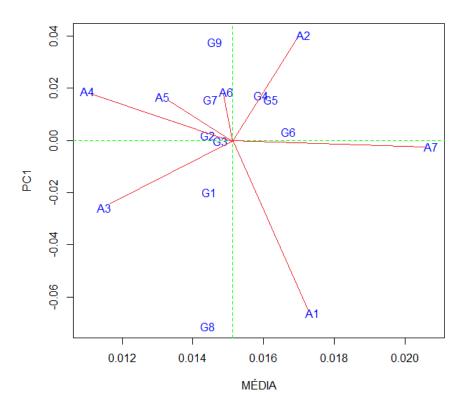


Figura 26 - Biplot AMMI1-ponderado (Médias vs PC1) para os dados de produtividade de milho ( $kg\ ha^{-1}$ ), com 15 genótipos (G) e quatro ambientes (A)

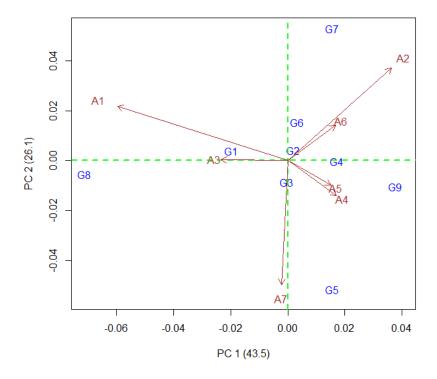


Figura 27 - Biplot AMMI2-ponderado (PC1 vs PC2) para os dados de produtividade de milho  $(kg\ ha^{-1})$ , com 15 genótipos (G) e quatro ambientes (A)

## 4.9 GGE biplot ponderado (W-GGE biplot) para análise de $2^o$ conjunto dados MET

Neste conjunto de dados do CIMMYT (INTERNATIONAZ MAIZE AND WHEAT IMPROVEMENT CENTER) em experimentos realizados em 7 países com 9 genótipos, não existe heterogeneidade entre os ambientes. Portanto, foi utilizado outro novo método para fazer a genelarização do GGE biplot (W-GGE biplot). Os dados centrados no meio ambiente para análise W-GGE biplot foram obtidos subtraindo-se a produtividade média de cada ambiente a partir do valor de rendimento original de cada genótipo em cada ambiente.

Os dados da Tabela 29, são valores de rendimento centrados no ambiente de nove genótipos de milho, testados em sete ambientes. De acordo com relação de informação (IR) dos sete componentes (Tabela 30), observou-se que apenas os três primeiros PCs contêm padrões (IR  $\geq 1$ ). Portanto, o biplot é considerado adequado para representar os padrões nos dados (IR  $_{PC1} > 1$ ). O GGE biplot com base neste conjunto de dados é apresentado na Figura 28. O PC1

explicou 44,51% e PC2 26,62% da variação total G + GEI, com isso, o W-GGE biplot explicou 71,13% da variação total G + GEI (Figura 28). Vale ressaltar que para este mesmo conjunto de dados, o GGE biplot explicou apenas 64,50% da variação total G + GEI.

Tabela 29 - Médias dos nove genótipos (G1 a G9) testadas em sete ambientes (A1 a A7) em relação a produtividade médio  $(kg\ ha^{-1})$ 

Genótipo	Ambientes						
	A1	A2	A3	A4	A5	A6	A7
G1	9,019	-481,114	449,111	-664,417	-575,431	131,424	-250,584
G2	-186,081	-290,794	-168,239	-267,417	246,249	-269,026	-441,134
G3	-166,681	-127,314	12,111	-93,717	-777,301	153,494	180,566
G4	107,369	329,736	-562,419	799,503	717,019	372,494	150,366
G5	-447,901	-130,594	1067,611	601,003	458,319	576,394	938,266
G6	503,549	668,106	99,911	749,903	636,119	610,644	250,866
G7	-258,501	557,586	284,111	-887,717	-678,051	439,244	-781,534
G8	916,249	-816,194	143,161	-788,497	-478,831	-1319,06	-31,434
G9	-477,001	290,606	-1325,32	551,383	451,919	-695,606	-15,334

Tabela 30 - Valor singular, proporção explicada e relação da informação (IR) dos quatro componentes principais (PCs)

РС	Valor singular	Variação explicado (%)	IR
1	13,672	44,51	3.116
2	10,573	26,62	1,863
3	8,561	17,45	1,222
4	5,513	$7,\!24$	0,507
5	2,867	1,96	0,137
6	2,591	1,59	0,111
7	1,629	0,63	0,044

Os genótipos G9 e G2 formaram ângulos obtusos com todos os ambientes (A1 a A7), isto significa que esses genótipos tiveram menores produtividades médias em todos os ambientes de

testes (Figura 28). Como genótipo G2 apresentou vetor maior do que G9, logo, ficou mais distante em relação todos os ambientes do que o G9, portanto, o G2 foi considerado o pior genótipo de produtividade em todos os ambientes. Por outro lado, o genótipo G5 formou ângulos agudos com os ambientes A5, A7, A4 e A6, portanto, o G5 produziu acima da média em todos os ambientes, exceto A1, A2 e A3. Da mesma forma, os genótipos G6, G7 e G4 formaram ângulos agudos com todos os ambientes, esse grupo de genótipos produziu acima da média em todos os ambientes. O G6 apresentou vetor maior do que G4 e G7, logo, G6 foi o genótipo que produziu mais em todos os ambiente (Figura 28). Essas informações podem ser confirmadas pela Tabela 29.

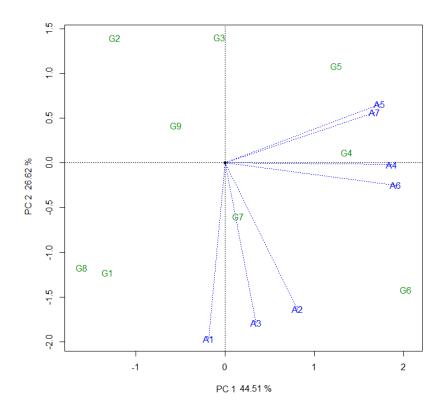


Figura 28 - O W-GGE Biplot para os dados de produtividade de milho  $(kg \ ha^{-1})$ , com nove genótipos (G) e vinte ambientes (E)

### 4.9.1 Análise Mega-ambiente com W-GGE biplot (2º conjunto de dados)

Na Figura 29 é apresentado o W-GGE biplot ("Which-won-where") que é uma ferramenta eficaz para análise de mega-ambiente sendo o mesmo formato apresentado na Figura

28, exceto para as linhas complementares. O W-GGE biplot foi construído no ambiente centrado (centralização = 2), não-dimensionado (escala = 0) e (PVS = 2). Os vértices do polígono são formados pelos genótipos: G8, G1, G6, G5, G3 e G2. Os sete ambientes foram cortados em 3 grupos pelas linhas que saíram da origem do biplot, a saber: i) A1; ii) A3, A2, A6 e A4; iii) A7 e A5 (Figura 29).

É possível notar que o genótipo G6 é o vértice do setor em que grande parte dos ambientes são colocados (A3, A2, A6 e A4), portanto, é o genótipo que teve melhor desempenho nestes ambientes (melhor produtividade média); o G5 é o vértice no setor em que A5 e A7 são colocados, logo, é o genótipo mais adaptado nestes ambientes. Nos setores de G3, G9, G2, G8 e G1 que não contém ambientes, significa que estes genótipos não eram produtivos em nenhum ambiente, ou seja, estes genótipos são os piores genótipos em relação à produtividade em alguns ou em todos os ambientes (Figura 29).

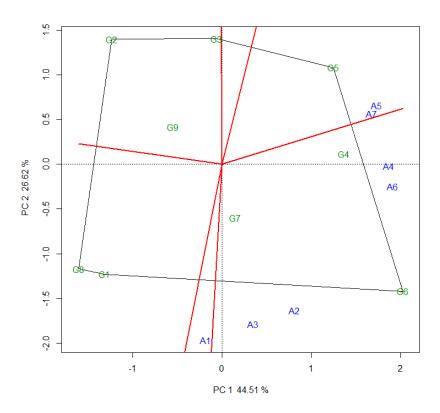


Figura 29 - O W-GGE Biplot ("Which-won-where") mostra que para os dados de produtividade de milho  $(kg\ ha^{-1})$ , quais dos genótipos apresentaram melhor desempenho em que ambientes

### 4.9.2 Avaliação de genótipos baseado em W-GGE biplot (2º conjunto de dados)

O W-GGE biplot "Média versus Estabilidade" (Figura 30) foi utilizado para identificar um genótipo "ideal", que deve ter alto desempenho médio e alta estabilidade por meio de um mega-ambiente. O pequeno círculo visualizado na Figura 30 representa o "ambiente-médio" e é definido pelas coordenadas médias de todos os ambientes de teste no biplot. Este biplot é baseada em (PVS = 1), ou seja, os valores singulares são totalmente compartimentado para as escores genotípicas.

Os genótipos são classificados de acordo com sua produtividade média da seguinte forma: G6 > G4 > G5 > G7 > média geral > G1 > G9 > G3 > G8 > G2. O G6 teve seu desempenho em produtividade nos ambientes de testes bem acima dos demais e foi altamente estável; o G9 também foi altamente estável, mas deixou a desejar em relação a produtividade, ficando abaixo da média geral. Os genótipos G5, G1, G8 e G3 foram instáveis nos ambientes de testes, e apenas o G5 teve produtividade acima da média (terceiro melhor). Os genótipos restantes ficaram abaixo da média sendo o G2 instável e também o pior genótipo de desempenho em relação à produtividade nos ambientes de testes (Figura 30).

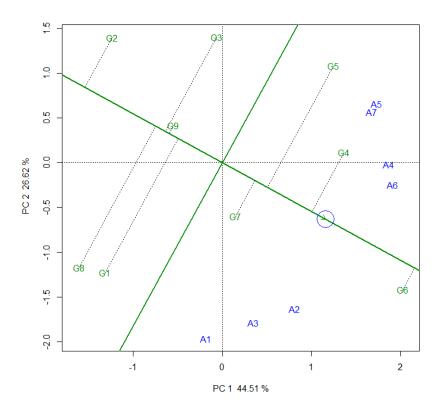


Figura 30 - O W-GGE Biplot ("Média versus Estabilidade") com eixo do ambiente-média (EAM) para mostrar o desempenho médio e estabilidade dos genótipos para os dados de produtividade de milho ( $kg\ ha^{-1}$ )

Ainda na Figura 30, os genótipos G4 e G7 apresentaram produtividade acima da média e as diferenças entre eles se dá pelo fato de G7 ser mais estável do que G5, e G4 com desempenho da produtividade média melhor do que G7.

O W-GGE biplot (Figura 31) com o centro dos círculos concêntricos sobre o EAM, os genótipos localizados mais próximo do centro dos círculos concêntricos mais próximo de ser um genótipo ideal. Assim, o G6 foi o genótipo "ideal" (Estabilidade e produtividade) neste conjunto de dados (Figura 30 e 31), portanto, o G6 é um excelente genótipo para recomendação para os melhoristas; em seguida, os melhores genótipos foram G4, G7, G9 e G3, G1, G8 e G2, respectivamente. Já os genótipos G4, G7 e G5 mostraram um bom desempenho em relação a estabilidade e produtividade e podem ser recomendados para cultivares. Os piores genótipos neste conjunto de dados foram G2, G8, G1 e G3, respectivamente, sendo os piores em relação a estabilidade e produtividade (Figura 31).

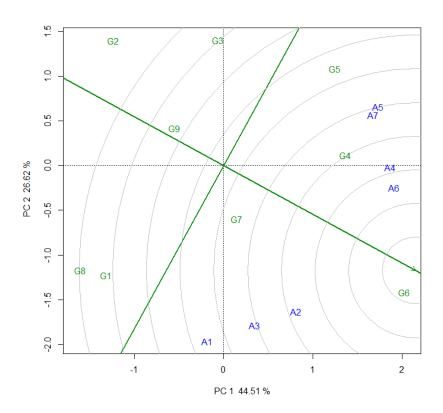


Figura 31 - O W-GGE Biplot com eixo do ambiente-média (EAM) para classificar os genótipos em relação ao genótipo ideal (no centro dos círculos concêntricos)

# 4.9.3 Avaliação do Ambiente de Teste baseado no W-GGE biplot (2º conjunto de dados)

Como já citado, o objetivo da avaliação do "teste-ambiente" é identificar ambientes de teste que podem ser utilizados para selecionar genótipos superiores de forma eficaz para um mega-ambiente. O W-GGE biplot "Discriminação contra Representatividade" (Figura 32) foi projetado para esta finalidade, dessa forma, todos os ambientes (A1 a A7) apresentaram vetores longos, significando que são mais discriminantes em relação aos genótipos (Figura 32).

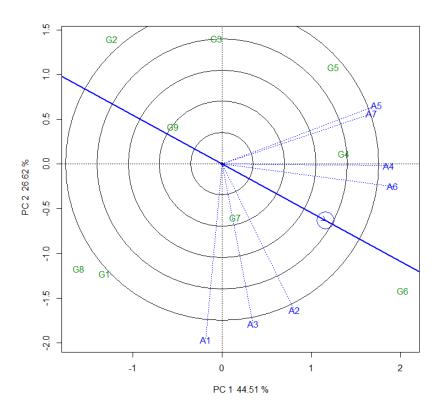


Figura 32 - O W-GGE biplot "discriminação e representatividade" para mostrar a capacidade discriminação e de representatividade dos ambientes de teste

Uma segunda utilidade da Figura 32, é indicar ambientes de representatividade do ambiente alvo. Uma vez que os ambientes de teste possuem pequenos ângulos com EAM (eixo do ambiente-média), por exemplo, A6, A4 e A2, são mais representativos do ambiente alvo do que aqueles que têm ângulos maiores, por exemplo, A7, A5, A1 e A3. Portanto, os ambientes de teste que são discriminantes e representativos (A6, A4 e A2) são ambientes de teste ideais

para a seleção de genótipos adaptados. Por outro lado, os ambientes de teste discriminantes mas não representativos (A1, A5, A3, A2 e A7) são úteis para a seleção de genótipos adaptados especificamente se os ambientes alvos podem ser divididos em mega-ambientes; ou para seleção de genótipos instáveis se ambiente alvo é um único mega-ambiente.

Sabe-se que dentro de um único mega-ambiente, o ambiente de teste ideal deve ser mais discriminante e também mais representativo do ambiente alvo, sendo assim, o ambiente A6 é o mais próximo do centro dos círculos concêntricos, portanto, é o "ambiente de teste ideal" neste conjunto de dados (discriminante e representatividade). Em sequência, os melhores são A4, A2, A7, A5, A3 e A1, respectivamente. O ambiente A1 foi o pior para selecionar genótipos adaptados a toda a região (Figura 33).

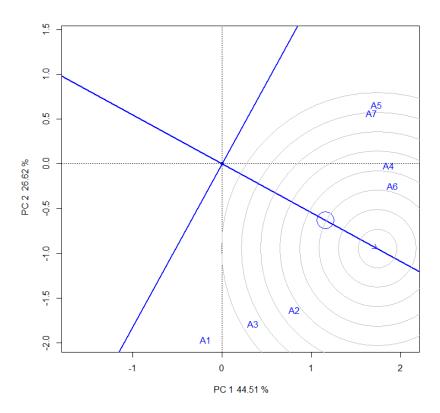


Figura 33 - O W-GGE biplot discriminação e representatividade para classificar ambientes de teste em relação a um ambiente de teste ideal (representado por centro dos círculos concêntricos)

# 4.10 Comparação das análises (AMMI vs AMMI-ponderado vs GGE biplot vs W-GGE biplot) para 2º conjunto de dados MET

Para este conjunto de dados multi-ambientes (MET) de produtividade de milho do CIMMYT (INTERNATIONAZ MAIZE AND WHEAT IMPROVEMENT CENTER). A Tabela 31, apresenta os valores singulares das sete componentes principais das análises AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot (GGE biplot-ponterado), sendo esses valores fundamentais para se fazer a comparação entre os modelos.

Tabela 31 - Valores singulares das sete componentes principais (PCs) para os modelos AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot

PC	$AMMI(V.S.)^1$	$AMMI-ponderado(V.S.)^1$	GGE biplot $(V.S.)^1$	W-GGE biplot(V.S.) <sup>1</sup>
1	909,132	0,00271	13,606	13,672
2	681,556	0,00209	9,261	10,573
3	541,587	0,00199	8,397	8,561
4	289,562	0,00089	6,963	5,513
5	206,457	0,00054	4,965	2,867
6	39,435	0,00022	2,171	2,591
7	5,53E-06	2,13E-11	0,877	1,629

(V.S.)<sup>1</sup>: Valores singulares

A Tabela 32 é uma tabela de dupla entrada ou uma matriz de correlação entre os valores singulares dos componentes principais (PCs) dos quatro modelos (AMMI; AMMI-ponderado; GGE biplot e W-GGE biplot), é dada da seguinte forma:

Tabela 32 - A matriz de correlação dos valores singulares de PCs dos modelos AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot

	AMMI	AMMI-ponderado	GGE biplot	W-GGE biplot
AMMI	1,0000	0.9906	0,9783	0,9910
AMMI-ponderado	0,9906	1,0000	0,9632	0,9869
GGE biplot	0,9783	0,9632	1,0000	0,9639
W-GGE biplot	0,9910	0,9869	0,9639	1,0000

Na Tabela 32, observou-se que a correlação entre AMMI vs W-GGE biplot foi muito forte ( $r_1$ =0,9910), as duas metodologias foram altamente correlacionadas, isso mostra também a coincidência entre os resultados. A correlação AMMI vs AMMI-ponderado:  $r_2$  = 0,9906; a correlação entre AMMI-ponderado vs W-GGE biplot:  $r_3$ = 0,9869; AMMI com GGE biplot:  $r_4$ =0,9783; a correlação GGE biplot vs W-GGE biplot:  $r_5$ =0,9639 é muito próxima da correlação entre GGE biplot vs AMMI-ponderado ( $r_5$ =0,9632). Todos os valores da correlação são > 0,95, portanto, mostraram que as quatro metodologias (AMMI, AMMI-ponderado, GGE, W-GGE biplot) são muito semelhantes em relação à análise de dados MET, pois essas metodologias são altamente correlacionadas entre si.

Na Tabela 33 é possível ver as variações explicadas pelos componentes principais dos modelos AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot e a relação de informação (IR). Pelo critério dos PCs que contêm os padrões da seguinte forma: (IR  $\geq$  1), o PC1 dos quatro modelos apresentaram valores maiores do que 1, o AMMI apresentou o maior valor (IR): 3,38 e AMMI-ponderado apresentou menor valor (IR): 3,05. Portanto, o PC1 do AMMI explicou maior proporção da variação dos dados e o menor deles foi do AMMI-ponderado.

Tabela 33 - Proporções explicada e IR dos componentes principais dos modelos AMMI, AMMI-ponderado, GGE biplot e W-GGE biplot

PC	AMMI-(%1)	IR	AMMI-p-(%1)	IR	GGE -(% <sup>1</sup> )	IR	W-GGE -(%1)	IR
1	48,3	3,38	43,5	3,05	44,08	3,086	44,51	3,116
2	27,1	1,90	26,1	1,83	20,42	1,429	26,62	1,863
3	17,1	1,19	23,6	1,65	16,79	1,175	17,45	1,222
4	4,9	0,34	4,8	0,34	11,54	0,808	7,24	0,507
5	2,5	0,18	1,7	0,12	5,87	0,411	1,96	0,137
6	0,1	0,007	0,3	0,021	1,12	0,078	1,59	0,111
7	0	0	0	0	14,22	0,18	0,63	0,044

(%<sup>1</sup>): Variação Explicada pelos os modelos

O primeiro componente sempre é o mais importante para a avaliação dos genótipos e ambientes, isto pode ser reforçado pela Tabela 33 com os valores do IR. Portanto, na interpretação dos biplot dos quatro modelos, sempre é importante considerar o PC1 que apresentar a maior variação do total.

Na análise AMMI, os dois primeiros componentes (PC1 e PC2) explicaram 75,4% da variação total da interação genótipo × ambiente (GEI) no biplot, e na análise AMMI-ponderado, foram de 69,6% da variação total da GEI. Na análise GGE biplot, os dois primeiros componentes

explicaram apenas 64,5% da variação total do efeito do genótipo mais interação genótipo  $\times$  ambiente (GGE) e na análise W-GGE biplot, os dois primeiros componentes explicaram 71,13% da variação total do GGE.

Para este conjunto de dados MET, o GGE biplot foi o modelo que menos explicou a variação total da interação. O modelo AMMI foi o melhor modelo para explicação do GEI, pois explicou a maior variação com os dois primeiros componentes, portanto, as interpretações dos gráficos biplot são as mais confiáveis para representar os padrões nos dados sobre genótipos e ambientes.

### 5 CONCLUSÕES

A utilização conjunta das metodologias AMMI, GGE Biplot, AMMI-ponderado e W-GGE biplot tornam mais eficazes a exploração da interação genótipo × ambiente e efetivamente resolveram os quarto principais objetivos para os dados multi-ambientes (MET): 1) investigação mega-ambiente para a compreensão do ambiente de destino; 2) avaliação de cultivares dentro de cada mega-ambiente; 3) avaliação do ambiente de teste dentro de cada mega-ambiente; e 4) compreensão das causas da interação genótipo × ambiente (GE).

A Análise com as quatro metodologias para os dados de MET podem ajudar os pesquisadores a entender melhor seu ambiente alvo, com o objetivo de estabelecer estratégias e testes de melhoramento de baixo custo, além de identificar genótipos superiores que são amplamente ou especificamente adaptados.

Na comparação das metodologias para os dados multi-ambientes provenientes da empresa Criagene SK, os valores da correlação dos quatro modelos foram ≥ 0,95, portanto, mostraram que as metodologias (AMMI, AMMI-ponderado, GGE, W-GGE biplot) são muito semelhantes em relação a análise de dados MET, pois são altamente correlacionados entre si. O GGE biplot foi o modelo que foi mais prejudicado com a heterogeneidade dos ambientes, portanto, foi aquele que menos explicou a variação total da interação G+GE (63,96%). O AMMI-ponderado foi o melhor modelo para este conjunto de dados, já que explicou a maior variação com os dois primeiros componentes (92,2%), portanto, as interpretações dos gráficos biplot são as mais confiáveis para representar os padrões nos dados sobre genótipos e ambientes. Os genótipos "ideais" (Estabilidade e produtividade) pelo modelo AMMI-ponderado foram G5, G3 e G2, que podem ser considerados como os melhores entre todos os genótipos de testes. O ambiente A2 (município de Casa Branca/SP 680m) foi o "ambiente de teste ideal".

Para o conjunto de dados MET do CIMMYT (CENTRO INTERNACIONAL DE MEJORAMIENTO DE MAIZ Y TRIGO), os valores da correlação dos quatro modelos também foram > 0,95, com isso, mostraram que os metodologias (AMMI, AMMI-ponderado, GGE, W-GGE biplot) são muito semelhantes em relação a análise de dados MET, já que as correlações foram muito altas entre elas.

Para este conjunto de dados MET, o GGE biplot foi o que menos explicou a variação total da interação G+GE (64,5%). O modelo AMMI foi o melhor modelo para explicação do GEI, pois explicou a maior variação com os dois primeiros componentes (75,4%), mostrando que as interpretações dos gráficos biplot são também as mais confiáveis para representar os padrões nos dados sobre genótipos e ambientes. Pela análise AMMI: os "ideais" genótipos (produtividade

e estabilidade) foram G6 e G5, que podem ser considerados como os melhores entre todos os genótipos de testes e o ambiente A6 foi o "ambiente de teste ideal". O pior ambiente de teste em relação a produtividade e estabilidade foi o A1.

#### 5.1 Pesquisas futuras

Dando continuidade ao trabalho desenvolvimento durante o doutorado, os seguintes temas podem ser abordados futuramente:

- i Simular conjuntos de dados de interação genótipos × ambientes e utilizar os quatro métodos de análises (AMMI; GGE biplot; AMMI-ponderado; W-GGE biplot) apresentados neste trabalho para fazer as comparações.
- ii Criar um teste estatístico para comparação dos quatro análises apresentados neste trabalho para os dados simulados de interação genótipos × ambientes.
- iii Criar outros métodos de fazer as ponderações para análise AMMI-ponderado e W-GGE biplot.
- iv Fazer a comparação com os métodos (AMMI; GGE biplot; AMMI-ponderado; W-GGE biplot) para estudo da interação QTL  $\times$  E.
- v Utilizar os quatro métodos de análises (AMMI; GGE biplot; AMMI-ponderado; W-GGE biplot) para estudo de conjuntos de dados simulados de interação QTL × ambientes.

### REFERÊNCIAS

ANDERSON, T.W. An introduction to multivariate statistical analysis. 3.ed. New Jersey. 2003. 374 p.

ANNICCHIARICO, P.; Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. **Journal of Genetics and Breeding**, New Jersey, v. 46, p. 269-278, 1992.

ALLARD, R.W.; BRADSHAW, A.D. Implications of genotype environmental interactions in applied plant breeding. **Crop Science**, Madison, v.4, p.503-508, 1964.

ARCINIEGAS-ALARCÓN, S.; DIAS, C.T.S. Análise AMMI com dados imputados em experimentos de interação genótipo x ambiente de algodão. **Pesquisa Agropecuária Brasileira**, Brasília, v.44, n.11, p.1391-1397, 2009.

ASFAW, A.; ALEMAYEHU, F.; GURUM, F.; ATNAF, M.; AMMI and SREG GGE biplot anlysis for matching varieties onto soybean production environments in Ethiopia. **Scientific Research and Essay**, Lagos, v.4, n.11, p.1322-1330, 2009. Disponível em: <a href="http://academicjournals.org/article/article1380626982\_Asfaw%20et%20al.pdf">http://academicjournals.org/article/article1380626982\_Asfaw%20et%20al.pdf</a>. Acesso em: 20 mar. 2014.

BALESTRE, M.; SOUZA, J.C.; PINHO, R.G.V.; OLIVEIRA, R.L.; PAES, J.M.V. Yield stability and adaptability of maize hybrids based on GGE biplot analysis characteristics. **Crop breeding and applied biotechnology**, Viçosa, v.9, n3, p.219-228, 2009. Disponível em: <a href="http://www.sbmp.org.br/cbab/siscbab/uploads/c8eb9791-1c9f-e792.pdf">http://www.sbmp.org.br/cbab/siscbab/uploads/c8eb9791-1c9f-e792.pdf</a>>. Acesso em: 29 abr. 2014.

BORG, I.; GROENEN, P. J. F. Modern Multidimensional Scaling: Theory and Applications. New York, Springer Series in Statistics, 2005. 471p.

BRADU, D.; GABRIEL K. R.; The biplot as a diagnostic tool for model of two-way tables. **Technometrics**, Alexandria, v. 20, p. 47–68, 1978.

CAMARGO-BUITRAGO, I.; INTIRE, E.Q.M.; GORDÓN-MENDOZA, R. Identificación de mega-ambientes para potenciar el uso de genótipos superiores de arroz en Panamá. **Pesquisa Agropecuária Brasileira**, Brasília, v.46, n.9, p.1601-1069, 2011. Disponível em: <a href="http://www.scielo.br/pdf/pab/v46n9/46n09a13.pdf">http://www.scielo.br/pdf/pab/v46n9/46n09a13.pdf</a>>. Acesso em: 16 abr. 2014.

CÁRDENAS, O.; GALINDO, P.; VICENTE-VILLARDÓN, J. L.; Los Métodos Biplot: evolución y aplicaciones. **Revista Venezolana de Análisis de Coyuntura**, Caracas, n. 001, v. 13, p. 279–303, 2007.

CARVALHO, C.G.P.; ARIAS, C.A.A.; TOLEDO, J.F.F.; ALMEIDA, L.A.; KIIHL, R.A.S.; OLIVEIRA, M.F. Interação genótipo × ambiente no desempenho produtivo da soja no Paraná. **Pesquisa Agropecuária Brasileira**, Brasília, v.37, n.7, p.989-1000, 2002.

CHAVES, J.L. Interação de cultivares com ambientes. In: NASS,L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES, M.C. Recursos genéticos e melhoramento - plantas. Rondonópolis: Fundação MT, 2001. p.673-713.

CHAVES, L.J.; VENCOVSKY, L.; GERALDI, I.O. Modelo não linear aplicado ao estudo da interação genótipo x ambiente em milho. **Pesquisa Agropecuária Brasileira**, Brasília, v.24, n.2, p.259-269, 1989.

- COCKERHAM, C.C. Estimation of genetics variance. ln: HANSON, W.D.; ROBINSON, H.F. (Ed) **Statistical genetics and plant breeding**. Madison: National Academy of Sciences, 1963. chap.2, p.53-94.
- COX, T.F.; COX, M. A. A. Multidimensional Scaling. Chapman and Hall/CRC Monographs on Statistics and Applied Probability. Boca Raton, 2001. 328p.
- COX, C.; GABRIEL, K. R.; Some comparisons of biplot and pencil-and-paper E.D.A. methods. In: LAUNER, R. L.; SIEGEL A. F. (Ed.). **Modern data analysis**. London: Academic Press, 1982. p. 45-82.
- CROSSA, J.; CORNELIUS, P.L. Sites regression and shifted multiplicative models clustering of cultivar trials sites under heterogeneity of error variance. **Crop Science**, Madison, v.37, p.406–415, 1997.
- CROSSA, J.; CORNELIUS, P.L.; YAN, W. Biplot of linear-bilinear models for studying crossover genotype x environment interaction. **Crop Science**, Madison, v.42, p.619-633, 2002.
- CROSSA, J.; YANG, R. C.; CORNELIUS, P.L. Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. **Journal of Agricultural**, **Biological**, and **Environmental Statistics**, Alexandria, v.9 p.362–380, 2004.
- CORNELIUS, P. L. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. **Crop Science**, Madison, v.33, p.1186-1193, 1993.
- CORNELIUS, P.L.; CROSSA, J.; SEYEDSADR, M.S. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. **CRC Press**, Boca Raton, v.8, p.199–234, 1996.
- CROSSA, J.; CORNELIUS, P.L. Sites regression and shifted multiplicative models clustering of cultivar trials sites under heterogeneity of error variance. **Crop Science**, Madison, v.37, p.406–415, 1997.
- CRUZ, C.D.; REGAZZI, A.J. Modelos biométricos aplicados ao melhoramento genético.  $2^a$  ed. Viçosa: UFV, 1997. p.390.
- DENIS, J.B.; GOWER, J.C. Asymptotic covariances for parameters of biadditive models. **Utilitas Mathematica**, Canada, v.46, p.193–205, 1994.
- \_\_\_\_\_. Asymptotic confidence regions for biadditive models: Interpreting genotype-environment interactions. **Applied Statistics**, London, v.45, p.479–493, 1996.
- DENIS, J.B. PAZMAN, A. Bias of LS estimators in non-linear regression models with constraints. Part II: Biadditive models. **Applied Statistics**, London, v.44, p.375–403, 1999.
- DIAS, C.T.S. Método para a escolha de componentes em modelo de efeito principal aditivo e interação multiplicativa. 2005. 80 p. Tese (Livre-Docência)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba. 2005.
- DIAS, C.T.d.S.; KRZANOWSKI, W.J. Model Selection and Cross Validation in Additive Main Effect and Multiplicative Interaction Models. **Crop Science**, Madison, v.43, p.865-873, 2003.

\_\_\_\_\_. Choosing components in the additive main effect and multiplicative interaction (AMMI) models. **Scientia Agricola**, Piracicaba, v.63, p.169-175, 2006.

DUARTE, J.B.; VENCOVSKY, R. **Interação genótipo** × **ambiente:** uma introdução à análise "AMMI". Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60p. (Série Monografias, 9).

EBERHART, S.A.; RUSSELL, W.A. Stability parameters for comparing varieties. Crop Science, Madison, v.6, p.36-40, 1966.

FALCONER, D.S. Introduction to quantitative genetics. Harlow: Longman, 1989. 438 p.

FALCONER, D.S.; MACKAY, T.F.C. Introduction to quantitative genetics. Edinburgh: Longman, 1996. 480 p.

FERREIRA, P.V. Estatística experimental aplicada à Agronomia. Maceió: ADUFAL, 2000. 422p.

FERREIRA, D.F. Estatística Multivariada. Lavras: UFLA, 2011. 675p.

FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, Victoria, v.14, p.742-754, 1963.

FRITSCHE-NETO, R.; MIRANDA, G. V.; DELIMA, R. O., SOUZA, H. N. Factor analysis and SREG GGE biplot for the genotype × environment interaction stratification in maize. **Ciência Rural**, Santa Maria, v.40, p.1043-1048, 2010. Disponível em: <a href="http://www.scielo.br/pdf/cr/v40n5/a576cr2396.pdf">http://www.scielo.br/pdf/cr/v40n5/a576cr2396.pdf</a>>. Acesso em: 25 abr. 2014.

GABRIEL, K.R. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, Cambridge, v.58, p.453-467, 1971.

\_\_\_\_\_. Biplot display of multivariate matrices for inspection of data and diagnosis. In: BARNETT, V. (Ed.). **Interpreting Multivariate Data**, London: Wiley, 1981. p. 147-173.

GABRIEL, K. R.; ZAMIR, S.; Lower rank approximation of matrices by least squares with any choice of weights. **Technometrics**, Alexandria, v. 21, n. 04, p. 489–498, 1979.

GALINDO, M. P. Una alternativa de representación simultánea: HJ-Biplot. **Questíio**, Catalunya, n. 01, p. 13–23, 1986.

GAUCH, H. G. Model selection and validation for yield trials with interaction. **Biometrics**, Washington, v.44, p.705-715, 1988.

\_\_\_\_\_. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Amsterdam, Elsevier, 1992. 278p.

\_\_\_\_\_. Statistical Analysis of Yield Trials by AMMI and GGE. **Crop Science**, Madison, v.46, p.1488-1500, 2006.

\_\_\_\_\_. A Simple Protocol for AMMI Analysis of Yield Trials. **Crop Science**, Madison, v.53, p.1860-1869, 2013.

GAUCH, H.G.; ZOBEL, R.W. AMMI analysis of yield trails. In: KANG, M.S.; GAUCH, H.G. (Ed) **Genotype by environment interaction**. New York: CRC Press, 1996. p.416-428.

\_\_\_\_\_. Identifying mega-environments and targeting genotypes. **Crop Science**, Madison, v.37, p.311-326, 1997.

GAUCH, H.G.; PIEPHO, H.P.; ANNICCHIARICO, P. Statistical analysis of yield trials by AMMI and GGE: Further considerations. **Crop Science**, Madison, v.48, p.866–889, 2008.

GGEBiplotGUI. GGEBiplotGUI: Interactive GGE Biplots in R. São Paulo, 2013.

GOLLOB, H.F. A statistical model which combines feature of factor analytic and analysis of variance techniques. **Psychometrika**, New York, v.33, p.73-115, 1968.

GOWER, J. C.; HARDING S.; Nonlinear biplots. **Biometrika**, Great Britain, n. 75, p. 445–455, 1988.

GOWER, J. C. Generalized Biplots. Biometrika, Great Britain, n. 79, p. 475–493, 1992.

\_\_\_\_\_. Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. **Biometrika**, Great Britain, v.53, p.325-338, 1966.

GUSMÃO, L. 1985. An Adequate Design for Regression-Analysis of Yield Trials. **Theoretical and Applied Genetics**, New York, v.71, p.314-319, 1985.

HONGYU, K.; GARCIA-PEÑA, M.; ARAÚJO, L.B.; DIAS, C.T.S. Statistical analysis of yield trials by AMMI analysis of genotype x environment interaction. **Biometrical Letters**, Poznan, v.51, p.89-102, 2014.

HOTELLING, H. Analysis of a Complex of Statistical Variables into Principal Components. **Journal of Educational Psychology**, Columbia, v.24, p.417-441, 1933.

\_\_\_\_\_. The Most Predictable Criterion. **Journal of Educational Psychology**, Columbia, v.26, p.139-142, 1935.

\_\_\_\_\_. Relations Between Two Sets of Variables. **Biometrika**, Great Britain, v.28, p.321-377, 1936.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Madison: Prentice Hall International, 1998. 816p.

JOLLIFFE, I. T. Principal Component Analysis. Springer Series in Statistics, 2nd ed. **Springer-Verlag**, New York, 2002. 488p.

KEMPTON, R.A. The use of biplots in interpreting variety by environment interactions. **Journal of Agricultural Science**, Islamabad, v. 103 p. 123–135, 1984.

KVITSCHAL, M. V.; FILHO, P. S. V.; SCAPIM, C. A.; GONÇALVES-VIDIGAL, M. C.; SAGRILO, E. PEQUENO, M. G.; RIMOLDI, F. Comparison of methods for phenotypic stability analysis of Cassava (Manihot esculenta Crantz) genotypes for yield and storage root dry matter content. **Brazilian Archives of Biology and Technology**, Curitiba, v.52, n.1, p.163-175, 2009. Disponível em: <a href="http://www.scielo.br/pdf/">http://www.scielo.br/pdf/</a> babt/v52n1/a22v52n1.pdf>. Acesso em: 28 abr, 2014.

KRUSKAL, J. B. Multidimensional Scaling by Optimizing Goodness-of-Fit to a Nonmetric Hypothesis. **Psychometrika**, Baltimore, v.29, p.1-27, 1964a.

\_\_\_\_\_. Nonmetric Multidimensional Scaling: A Numerical Method. **Psychometrika**, Baltimore, v.29, p.115-129, 1964b.

KRZANOWSKI, W. J. **Principles of Multivariate Analysis**: A User's Perspective. Oxford Statistical Science Series. Oxford University Press, New York, 2000. 608p.

LIPKOVICH, I.; SMITH, E.P. BiPlot. Blacksburg, 2002a.

\_\_\_\_\_. Biplot and Singular Value Decomposition Macros for Excel. **Journal of Statistical Software**. Blacksburg, v.7, p.1-15, 2002b. Disponível em: <a href="http://www.jstatsoft.org/v07/i05">http://www.jstatsoft.org/v07/i05</a>. Acesso em: 02 abr. 2014.

MANDEL, J. Non-additivity in two-way analysis of variance. **Journal of the American Statistical Association**, New York, v.56, p.878-888, 1961.

\_\_\_\_\_. The partitioning of interactions in analysis os variance. **Journal of Research of the National Bureau of Standards, Series B**, Washington, v.73, p.309-328, 1969.

\_\_\_\_\_. A new analysis os variance model for non-adittive data. **Technometrics**, Alexandria, v.13, n.1, p.1-18, 1971.

\_\_\_\_\_. Use of the singular value decomposition in regression analysis. **The American Statistician**, Washington, v.36, n.1, p.15-24, 1982.

MARTÍN-RODRIGUEZ, J.; GALINDO, P.; VICENTE-VILLARDÓN, J. L.; Comparison and integration of subspaces from a biplot perspective. **Journal of Statistical Planning and Inference**, Maryland Heights, v. 2, n. 102, p. 411-423, 2002.

MAZER, S.J.; SCHICK, C.T. Constancy of population parameters for life history and floral traits in Raphanus sativus L. I. Norms of reaction and the nature of genotype by environment interactions. **Heredity**, Oxford, v.67, n.2, p.143-156, 1991.

MINITAB, Institute Inc. **Minitab 9.0**, State College, 2007. Disponível em: <a href="http://www.minitab.com/">http://www.minitab.com/</a>>. Acesso em: 22 mar. 2014.

PEARSON, K. Lines and Planes of Closest Fit to Systems of Points in Space. **Philosophical Magazine**, London, v.2, p.559-572, 1901.

PIEPHO, H.P. Robustness of statistical test for multiplicative terms in additive main effects and multiplicative interaction model for cultivar trial. **Theoretical and Applied Genetics**, New York, v. 90, p. 438-443, 1995.

RAMALHO, M.A.P.; SANTOS, J.B.; ZIMMERMANN, M.J.O. Genética quantitativa em plantas autógamas: aplicações ao melhoramento do feijoeiro. Goiânia: UFG, 1993. 271P.

RAMÍREZ, G.; VASQUEZ, M.; CAMARDIEL, A.; PÉREZ, B.; GALINDO, P.; Detección grafica de la multicoliearidad mediante el h-plot de la inversa de la matriz de correlaciones. **Revista Colombiana de Estadística**, Bogotá, v.28, n.2, p.207-219, 2005.

R, DEVELOPMENT, C. T. R: Language and Environment for Statistical Computing. **R** Foundation for Statistical Computing, Vienna, 2009. Disponível em: <a href="http://www.R-project.org/">http://www.R-project.org/</a>>. Acesso em: 24 mar. 2014.

ROBERTSON, A. Experimental design on the mesurement of herdabilities and genetic correlations. **Biometrical genetics**. New York: Pergamon, 1959. 186p.

RODRIGUES, P.C. New strategies to detect and understand genotype-by-environment interactions and QTL-by-environment interactions. Lisbon: Faculdade de Ciências e Tecnologia, 2012. 145p.

RODRIGUES, P.C.; MALOSETTI, M.; GAUCH, H. G.; VAN EEUWIJK F.A. A Weighted AMMI Algorithm to Study Genotype-by-Environment Interaction and QTL-by-Environment Interaction. **Crop Science**, Madison, v.54, p.1555-1570, 2014.

SAMMON, J. W. A Nonlinear Mapping for Data Structure Analysis. **IEEE Transactions on Computers**, Washington, v.18, p.401-409, 1969.

SAS, Institute Inc. **SAS 9.2**, Cary, 2009. Disponível em: <a href="http://www.sas.com/">http://www.sas.com/</a>>. Acesso em: 15 mar. 2014.

SILVA, R. R.; BENIN, G. Análises Biplot: conceitos, interpretações e aplicações. **Ciência Rural**, Santa Maria, v.42, n.8, p.1404-1412, ago. 2012.

SINEBO, W.; LAKEW, B.; FEYISSA, A. Biplot analysis of grain yield barley grown under differing management levels in years of contrasting season-end drought. **Journal of Plant Breeding and Crop Science**, Ottawa, v.2, n.6, p.152-164, 2010. Disponível em: <a href="http://www.academicjournals.org/jpbcs/PDF/pdf%202010/June/Sinebo%20et%20al.pdf">http://www.academicjournals.org/jpbcs/PDF/pdf%202010/June/Sinebo%20et%20al.pdf</a>>. Acesso em: 25 abr. 2014.

SOUZA, É. C. Os Métodos Biplot e Escalonamento Multidimensional nos delineamentos experimentais. 2010. 135p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2010.

SPSS, Institute inc. **SPSS 17.0**, Chicago, 2008. Disponível em: <a href="http://www.spss.com/">http://www.spss.com/>. Acesso em: 10 mar. 2014.

SQUILASSI, M.G. Interação de genótipos com ambientes. Embrapa Tabuleiros Costeiros, Aracaju, 2003. 47p.

STATACORP, Institute inc. **StataCorp 10.0**, College Station, 2007. Disponível em: <a href="http://www.stata.com/">http://www.stata.com/</a>. Acesso em: 22 mar. 2014.

SHELBOURNE, C. Genotype environment interaction: its implications in forest tree improvement. **Proceedings**, Tokyo, 1972, 28p.

TIERNEY, L. Lisp-Stat: An Object-Oriented Environment for Statistical and Dynamic Graphics. Wiley Series in Probability and Mathematical Statistics. John Wiley Sons, 1990. 416p.

TORGERSON, W.S. Multidimensional Scaling: 1. Theory and Method. **Psychometrika**, Baltimore, v.17, p.401-419, 1952.

TRETHOWAN, R.M.; CROSSA, J; GINKEL, M.V.; RAJARAM, S. Relationships among bread wheat international yield testing locations in dry areas. **Crop Science**, Madison, v.41, n.5, p.1461-1469, 2001.

TUKEY, J.W. One degree of freedom for non-additivity. **Biometrics**, Washington, v.5, p.232-242, 1949.

UDINA, F. Interactive Biplot Construction. **Journal of Statistical Software**, Alexandria, v.13, n.5, p.1-16, 2005a. Disponível em: <a href="http://www.jstatsoft.org/v13/i05/">http://www.jstatsoft.org/v13/i05/</a>>. Acesso em: 27 mar. 2014.

\_\_\_\_\_. Universitat Pompeu Fabra, Barcelona, 2005b. Disponível em: <a href="http://tukey.upf.es/xls-biplot/index.html">http://tukey.upf.es/xls-biplot/index.html</a>. Acesso em: 22 mar. 2014.

VARELA, M. Los métodos Biplot como herramienta de análisis de interacción de ordem superior en un modelo lineal/bilineal, 2002. 155p. Tese-Universidade de Salamanca, Salamanca, 2002.

VICENTE-VILLARDÓN, J. L.; GALINDO-VILLARDÓN, M. P.; BLÁSQUEZ-ZABALLOS, A.; Logistic Biplots. In: GREENACRE, M.; BLASIUS, J. (Ed.). **Multiple Correspondence Analysis and related methods**, London: Chapman Hall, 2006. p. 491-509.

WARZECHA, T.; ADAMSKI, T.; KACZMAREK, Z.; SURMA, M.; CHELKOWSKI, J.; WIŚNIEWSKA, H.; KRYSTKOWIAK, K.; KUCZYŃSKA, A. Genotype-by-environment interaction of barley DH lines infected with Fusarium culmorum (W.G.Sm.)Sacc. **Field Crops Research**, Amsterdam, v.120, p.21-30, 2011.

YAN, W.; GGE Biplot. A windows application for graphical analysis of multienvironment trial data and other types two-way data. **Agronomy Journal**, Madison, v. 93, p. 1111-1118, 2001.

\_\_\_\_\_. Singular-value partition for biplot analysis of multi-environment trial data. **Agronomy Journal**, Madison, V. 94, P.990-996, 2002.

\_\_\_\_\_. GGE Biplot vs. AMMI Graphs for Genotype-by-Environment Data Analysis. **Journal of the Indian Society of Agricultural Statistics**, New Delhi, v.65, n.2, p.181-193, 2011. Disponível em: <a href="http://isas.org.in/jisas/jsp/volume/vol65/issue2/06-Weikai%20Yan.pdf">http://isas.org.in/jisas/jsp/volume/vol65/issue2/06-Weikai%20Yan.pdf</a>. Acesso em: 20 abr. 2014.

YAN, W.; CORNELIUS, P.L.; CROSSA,J.; HUNT, L.A. Two types of GGE biplot for analyzing multi-environment trial data. **Crop Science**, Madison, v.41, p.656-663, 2001.

YAN, W; HOLLAND, J. B. A heritability-adjusted GGE biplot for test environment evaluation. **Euphytica**, Wageningen, v. 171 p. 355–369, 2010.

YAN, W.; KANG, M.S. **GGE Biplot Analysis:** A Graphical Tool for Breeders, Geneticists, and Agronomists. Boca Raton, 2003. 271 p.

\_\_\_\_\_. **GGEbiplot**, Version 5, 2006. Disponível em: <a href="http://www.ggebiplot.com/">http://www.ggebiplot.com/</a>>. Acesso em: 22 mar. 2014.

YAN, W.; KANG, M.S.; MA, B.; WOODS, S.; CORNELIUS, P.L. GGE biplot vs. AMMI analysis of genotype-by-environment data. **Crop Science**, Madison, v.47, p.643-655, 2007.

YAN, W.; RAJCAN, I. Biplot analysis of test sites and trait relations of soybean in Ontario. **Crop Science**, Madison, v.42, p.11-20, 2002.

YAN, W; TINKER, A. Biplot analysis of multi environment trial data: principles and applications. Canadian Journal of Plant Science, Ottawa, v.86, n.3, p.623-645, 2006.

YANG, R.C.; CROSSA, J.; CORNELIUS, P.L. BURGUE $\tilde{N}$ O, J. Biplot analysis of genotype  $\times$  environment interaction: Proceed with caution. **Crop Science**, Madison, v.49, p.1564–1576, 2009.

YATES, F.; COCHRAN, W.G. The analysis of groups of experiments. **Journal of Agricultural Science**, Cambridge, v.28, p.556-580, 1938.

### **APÊNDICES**

### APÊNDICE A

```
Programação do modelo AMMI no SAS
a) *Análise de variância conjunta;
proc anova data = AMMIMED;
class AMB GENOT;
model YIELD = AMB GENOT AMB*GENOT;
test H = AMB GENOT E = AMB*GENOT;
run;
*Matriz de Médias;
proc sort data=AMMIMED;BY GENOT AMB;run;
proc means data=AMMIMED mean;
var YIELD;
by GENOT AMB;
OUTPUT OUT=MEDIAS_1;
run;
PROC PRINT DATA=MEDIAS_1; RUN;
DATA AMMI;
SET MEDIAS_1;
IF _STAT_ = ''MEAN";
DROP OBS _TYPE_
                 _FREQ_ _STAT_;
PROC PRINT; RUN;
/*
b) Ajuste dos efeitos principais e obtenção da matriz de interações GxE:
title ''ESTUDO DE INTERAÇÃO GxE { MÉTODO AMMI";
title2 ''Análise com dados de Média (QM'originais/r)";
proc glm data=AMMI;
  class GENOT AMB;
  model YIELD=GENOT AMB;
  run; title2 'Médias de Ambientes e Genótipos';
  Lsmeans GENOT AMB;
   Output out=GEAM r=GE;
run;
proc sort data=AMMI;
by GENOT;
proc transpose data=AMMI out=MEDIAS (drop=_name_) prefix=A;
 var YIELD;
  by GENOT;
proc print data=MEDIAS (drop=GENOT);
  title2 "'A matriz X de médias (gxa)";
run;
proc sort data=GEAM;
by GENOT;
proc transpose data=GEAM out=GXE (drop=_name_) prefix=A;
  var GE;
  by GENOT;
run;
proc print data=GXE (drop=GENOT);
  title2 ''Matriz de interações GxE (gxa)";
```

```
144
run;
/*
II) A decomposição da integração GxE
a) DVS da matriz de interações GxE:
*/
proc iml;
GLERR= 540; QMERR= 755888;
use work.MEDIAS; setin work.MEDIAS;
read all var {A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13 A14 A15 A16
A17 A18 A19 A20} into X;
ng=nrow(X); na=ncol(X);
GM=X[,:]; AM=X[:,]; GG=X[:,:];
CORR=(ng*na)*(GG##2); SQTR=ssq(X)-CORR;
SQGEN=na*ssq(GM)-CORR;
SQAMB=ng*ssq(AM)-CORR;
SQINT=SQTR-SQGEN-SQAMB;
GLMOD=(ng-1)+(na-1)+(ng-1)*(na-1);
XG=GM*J(1,na,1); XA=J(ng,1,1)*AM;
MG=J(ng,1,1)*GG*J(1,na,1);
GE=X-XG-XA+MG;
EIGV=EIGVAL(GE*GE');
*print EIGV;
*QUIT;
NNUL1=EIGV<-1E-8; NNUL2=EIGV>1E-8; NNUL=NNUL1+NNUL2;
r=(NNUL^{=0})[+,];
title2 'Decomposição por Valores Singulares (A=USV')da matriz GE';
SQGE=ssq(GE);
print 'SQ dos elementos da matriz GE = SQ(GxE):' SQGE;
print 'O posto da matriz GE:' r;
CALL SVD(eG, VS, eA, GE); S=DIAG(VS[1:r,]); U=eG[,1:r]; VT=eA[,1:r]';
print'Matriz de valores singulares(S):' S[format=12.4]/
  'Matriz de autovalores (U) associados a genótipos (AA'): U[format=12.6]/
  'Matriz de autovetores (V') associados a ambientes (A'A): VT[format=12.6]/;
/*
b) Desdobramento da SQ(GxE) por AMMI (ACP ou DVS):
*/
Li=VS##2; AUTV=Li[1:r,];
SAUTV=SUM(Li);
PRP=INV(SAUTV)*Li;PROP=PRP[1:r,];
RESD1=SAUTV-Li[1:1,];
PACUM=CUSUM(PROP)*100;
LINH=(1:r)';
RESUMO=LINH||AUTV||PROP||PACUM;
colu={'CP/Eixo''Autovalor''Prop./CP''% Acumulada'};
title2 'Resultados do desdobramento da SQ(GxE) por AMMI (DVS ou ACP)';
```

```
print 'Notas:',
$'______',
'Autovalores da ACP<=>Desdobramento da SQ(GxE) por AMMI',
'Proporção da SQ(GxE) atribuída a cada CP (ou eixo singular de interação)',
'Percentagem da SQ(GxE) acumulada até o n-ésimo CP(ou eixo singular)'
'______';$
print RESUMO [colname=colu format=12.4];
print, 'Soma dos Autovalores<=>SQ(GxE):' SAUTV/;
title2;
/*
c) Escores de genótipos e ambientes e aproximações sucessivas de GE e X:
*/
n=0;
 GXEI=j(ng,na,0); GLACU=0;
 SQGEI=ssq(GXEI);
 PROPI=INV(SAUTV)*SQGEI;
 PSQTR=inv(SQTR)*(SQGEN+SQAMB);
 XPO=XG+XA-MG;
 print
           'Resultados da avaliação do modelo: AMMI' n;
          'SQmod/SQtrat:'PSQTR'SQge(mod)/SQge:'PROPI,
 print
           '_____, /;
 print 'Matriz de respostas preditas (gxa): Modelo AMMI' n;
 print XPO[format=12.2]/;
do n=1 to r by 1; i=n;
 VG=eG[,i]; ES=S[i,i]; VA=eA[,i];
 IN=VG*ES*VA';
 G=VG*(sqrt(ES));ESGE=G';
 H=sqrt(ES)*VA;ESAM=H';
 if i=1 then ECP1=G//H;
 if i=2 then ECP2=G//H;
 if i=3 then ECP3=G//H;
 if i=4 then ECP4=G//H;
 if i=5 then ECP5=G//H;
 if i=6 then ECP6=G//H;
 if i=7 then ECP7=G//H;
 if i=8 then ECP8=G//H;
 GEI=G*H'; ZE=IN-GEI;
 SQGEI=SQGEI+ssq(GEI);
 PROPI=INV(SAUTV)*SQGEI;
 GLAMI=ng+na-1-(2*i); QMAMI=inv(GLAMI)*(ssq(GEI));
 GLACU=GLACU+GLAMI;
 GLRAMI = (ng-1)*(na-1)-GLACU;
 if i<r then QMRAMI=inv(GLRAMI)*(SAUTV-SQGEI); else QMRAMI=0;
 FMODI=inv(QMERR)*QMAMI; FRESI=inv(QMERR)*QMRAMI;
 GXEI=GXEI+GEI;
 XPI=XG+XA-MG+GXEI;
 PSQTR=inv(SQTR)*(SQGEN+SQAMB+SQGEI);
 prob1=1-probf(FMODI,GLAMI,GLERR);
```

```
if i<r then prob2=1-probf(FRESI,GLRAMI,GLERR); else prob2={.};</pre>
  FV1=GLAMI||QMAMI||FMODI||PROB1;FV2=GLRAMI||QMRAMI||FRESI||PROB2;
  FV3=GLERR||QMERR||{.}||{.};
  FV=FV1//FV2//FV3;
  lin={'AMMIn(GxE)' 'Res.GxE/AMMIn' 'Erro médio'};
  col={'GL''QM''teste F''Pr>F'};
  print 'Resultados da avaliação do modelo: AMMI' n;
  print FV[rowname=lin colname=col format=12.4];
  print 'SQmod/SQtrat:' PSQTR 'SQge(mod)/SQge:' PROPI/;
  print 'Escores para genótipos e ambientes / Modelo ajustado: AMMI' n,
  'Escores de genótipos' ESGE [format=12.4],
  'Escores de ambientes' ESAM [format=12.4]/;
  print 'Aproximação DVS para a Matriz de interações (gxa): Modelo AMMI' n;
  print GXEI[format=12.4]/;
  print 'Matriz de respostas preditas (gxa): Modelo AMMI' n;
  print XPI[format=12.2]/;
end;
/*
d)Resultados para a representação gráfica (Biplot):
*/
GEN=1:ng;AMB=1:na;
OBJ=(GEN | AMB);;
PROD=GM//AM';
RES=OBJ||PROD||ECP1||ECP2||ECP3||ECP4||ECP5||ECP6||ECP7||ECP8;
RESGEN=RES[1:9,1:4];
RESAMB=RES[10:29,1:4];
VAR1={'GENOTIPOS' 'MEDIAS' 'IPCA1' 'IPCA2' 'IPCA3'};
VAR2={'AMBIENTES' 'MEDIAS' 'IPCA1' 'IPCA2' 'IPCA3'};
create RESULT1 from RESGEN [colname=VAR1];
append from RESGEN;
close RESULT1;
create RESULT2 from RESAMB [colname=VAR2];
append from RESAMB;
close RESULT2;
quit;
proc print data=RESULT1;
title2 "Resultados úteis para a representação gráfica em Biplot";
title3 ''Médias e escores (3 primeiros IPCA's) para genótipos";
proc print data=RESULT2;
title2 "Resultados úteis para a representação gráfica em Biplot";
title3 ''Médias e escores (3 primeiros IPCA's) para ambientes";
run;
/* e) Representação gráfica (Biplot): */
data BIPLOT1GEN(keep=xsys ysys x y color function position size text style);
length text \$ 10;
set RESULT1;
```

```
text=GENOTIPOS;
style = 'ZAPFB';
xsys='2'; ysys='2'; color='blue'; position='5'; function='label';
size=1.2;
x=IPCA1;
y=IPCA2;
data BIPLOT1AMB(keep=xsys ysys x y color function position size text style);
length text \$ 10;
set RESULT2;
text=AMBIENTES;
style = 'SWISSB';
xsys='2'; ysys='2'; color='red'; position='5'; function='label';
size=1.4;
x=IPCA1;
y=IPCA2;
data GRAFICO1;
set BIPLOT1GEN BIPLOT1AMB;
data BIPLOT2GEN(keep=xsys ysys x y color function position size text style);
length text \$ 10;
set RESULT1;
text=GENOTIPOS;
style = 'ZAPFB';
xsys='2'; ysys='2'; color='blue'; position='5'; function='label';
size=1.2;
x=MEDIAS;
y=IPCA1;
data BIPLOT2AMB(keep=xsys ysys x y color function position size text style);
length text \$ 10;
set RESULT2;
text=AMBIENTES;
style = 'SWISSB';
xsys='2'; ysys='2'; color='red'; position='5'; function='label';
size=1.4;
x=MEDIAS;
y=IPCA1;
data GRAFICO2;
set BIPLOT2GEN BIPLOT2AMB;
proc gplot data=GRAFICO1;
symbol1 v=none i=none color=white;
plot y*x=1 y*x=1/anno=GRAFICO1 overlay vref=0 href=0;
title1 'Biplot IPCA2(y) vs IPCA1(x)';
title2 'Obs.: N genótipos maiores e ambientes menores';
run;
proc gplot data=GRAFICO2;
symbol1 v=none i=none color=white;
plot y*x=1 y*x=1/anno=GRAFICO2 overlay vref=0 href=0;
title1 'Biplot IPCA1(y) vs MEDIAS(x)';
title2 'Obs.: N genótipos maiores e ambientes menores';
run;
```

### APÊNDICE B

```
Programação dos Modelos AMMI, AMMI-ponderado, GGE biplot e GGE biplot-ponderado no R
dados<-read.csv2("dados.csv",head=TRUE);dados
library(agricolae)
library(klaR)
Ambiente<- dados[, 2]
Genotipo <- dados[, 1]
Bloco<- dados[, 3]
Prod<- dados[, 4]</pre>
ENV=Ambiente
GEN=Genotipo
REP=Bloco
Y=Prod
MSE = 0
number = TRUE
graph = "biplot"
model<- AMMI(Ambiente, Genotipo, Bloco, Prod, graph="biplot"); model</pre>
model<- AMMI(Ambiente,Genotipo,Bloco, Prod, graph="biplot",number=FALSE)</pre>
names(model)
names(model$biplot)
dados1<-read.csv2("dados2.csv",head=TRUE);dados1
attach(dados1)
REP <- 4
MSerror <- 823706
model <- AMMI (ENV, GEN, REP, MÉDIA, MSerror, graph="biplot"); model
model <- AMMI(ENV, GEN, REP, MÉDIA, MSerror, graph="biplot",number=FALSE)
xlim=c(-80,150), graph="biplot"
pc<- princomp(model$genXenv, cor = FALSE)</pre>
рс
pc$loadings
summary(pc)
model$biplot
bplot<-model$biplot[,1:4]</pre>
bplot
x11()
model<-AMMI(ENV, GEN, REP, MÉDIA, MSerror, ylim=c(-30,40), xlim=c(-55,50),
graph="biplot");model
model<- AMMI(ENV, GEN, REP, MÉDIA, MSerror, ylim=c(-30,40), xlim=c(-55,50),
graph="biplot",number=FALSE)
rm(REP)
bplot<-model$biplot[,1:4];bplot</pre>
attach(bplot)
#plot(PC1,PROD)
######### PC1 vs PC2 ###########
par(cex=0.8)
plot(PC1,PC2,cex=0.0,text(PC1,PC2,labels=row.names(bplot),col=c("black","black",
"black", "black", "black", "black", "black", "black", "black", "black", "black", "black",
```

```
"black", "black", "black", "black", "black", "black", "black", "black", "black", "black",
"black", "black", "black", "black", "black", "black", "black")), frame=TRUE)
abline(h=0,v=0,lty=2,col="black")
par(cex=0.8)
plot(MÉDIA, PC1, cex=0.0, text(MÉDIA, PC1, labels=row.names(bplot), col=c("black",
"black", "black", "black", "black", "black", "black", "black", "black",
"black", "black", "black", "black", "black", "black", "black", "black",
"black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "black", "bl
"black")),frame=TRUE)
abline(h=0,v= MEANS,lty=2,col="black")
par(cex=0.8)
plot(MÉDIA, PC1, cex=0.0, text(MÉDIA, PC1, labels=row.names(bplot), col=c("blue",
"blue", "blue", "blue", "blue", "blue", "blue", "blue", "blue", "red4", "red4",
"red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4"
"red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4", "red4")), frame=TRUE)
MEANS<-mean(MÉDIA)
#plot(YLD,PC1)
abline(h=0,v= MEANS,lty=2,col="green")
abline(h=25, lty=2, col="black")
abline(h=-25, lty=2,col="black")
amb<-subset(bplot,type=="ENV")</pre>
amb
detach(bplot)
attach(amb)
s <- seq(length(MÉDIA))
arrows(MEANS, 0, 0.9*(MÉDIA[s]-MEANS)+MEANS, 0.9*PC1[s], col= "brown",
lwd=1.8,length=0.1)
cova<-cov(model$genXenv)</pre>
cova
values <- eigen (cova)
values
total <- sum (values $ values )
total
round(values$values*100/total,2)
library(agricolae)
library(klaR)
dados<-read.csv2("dados.csv",head=TRUE);dados
ENV=AMB
GEN=GENOT
REP=bloco
Y=PROD
MSE=0
number = TRUE
graph = "biplot"
AMMI1=function (ENV, GEN, REP, Y, MSE = 0, number = TRUE, graph = "biplot",
{
```

```
name.y <- paste(deparse(substitute(Y)))</pre>
cat("\nANALYSIS AMMI: ", name.y, "\nClass level information\n")
ENV <- as.factor(ENV)</pre>
GEN <- as.factor(GEN)</pre>
nenv <- length(unique(ENV))</pre>
ngen <- length(unique(GEN))</pre>
cat("\nENV: ", unique(as.character(ENV)))
cat("\nGEN: ", unique(as.character(GEN)))
minimo <- min(ngen, nenv)
if (length(REP) > 1) {
    REP <- as.factor(REP)</pre>
    nrep <- length(unique(REP))</pre>
    cat("\nREP: ", unique(REP))
    cat("\n\nNumber of observations: ", length(na.omit(Y)),
         "\n\n")
    modelo <- aov(Y ~ ENV + REP %in% ENV + GEN + ENV:GEN)
    cat("model Y:", name.y, " ~ ENV + REP%in%ENV + GEN + ENV:GEN\n")
    cat("Random effect REP%in%ENV\n\n")
    mm <- anova(modelo)
    nn \leftarrow mm[2,]
    mm[2, ] <- mm[3, ]
    mm[3, ] <- nn
    row.names(mm)[2] <- "REP(ENV)"</pre>
    row.names(mm)[3] <- "GEN
    mm[1, 4] \leftarrow mm[1, 3]/mm[2, 3]
    mm[1, 5] \leftarrow 1 - pf(mm[1, 4], mm[1, 1], mm[2, 1])
    print(mm)
    DFE <- df.residual(modelo)</pre>
    MSE <- deviance(modelo)/DFE
    medy <- mean(Y, na.rm = TRUE)</pre>
    cat("\nCoeff var", "\tMean", name.y, "\n")
    cat(sqrt(MSE) * 100/medy, "\t", medy, "\n")
}
else {
    DFE <- nenv * (ngen - 1) * (REP - 1)
    DFEa <- nenv * (REP - 1)
    nrep <- REP
    modelo <- aov(Y ~ ENV + GEN)</pre>
    xx <- as.matrix(anova(modelo))</pre>
    xx \leftarrow rbind(xx[1,], xx[1,], xx[2:3,], xx[3,])
    row.names(xx)[4] <- "ENV:GEN"
    row.names(xx)[5] <- "Residuals"</pre>
    xx[2, 1] \leftarrow DFEa
    xx[2, 2:5] \leftarrow NA
    xx[, 2] \leftarrow xx[, 2] * nrep
    xx[, 3] <- xx[, 3] * nrep
    xx[5, 1] \leftarrow DFE
    xx[5, 3] \leftarrow MSE
    xx[5, 2] \leftarrow MSE * DFE
    xx[1, 4] \leftarrow NA
    xx[3, 4] < -xx[3, 3]/MSE
    xx[4, 4] <- xx[4, 3]/MSE
```

```
xx[1, 5] \leftarrow NA
    xx[3, 5] \leftarrow 1 - pf(xx[3, 4], xx[3, 1], DFE)
    xx[4, 5] \leftarrow 1 - pf(xx[4, 4], xx[4, 1], DFE)
    row.names(xx)[1] <- "ENV</pre>
    row.names(xx)[2] <- "REP(ENV)"</pre>
    cat("\nREP: ", REP)
    cat("\n\n \n \n), "\n")
    cat("\nDependent Variable:", name.y, "\n\nAnalysis of variance\n")
    print(xx, na.print = "")
    medy <- mean(Y, na.rm = TRUE)</pre>
    cat("\nCoeff var", "\tMean", name.y, "\n")
    cat(sqrt(MSE) * 100/medy, "\t", medy, "\n")
}
raw <- data.frame(ENV, GEN, Y)</pre>
MEDIAS \leftarrow tapply(raw[, 3], raw[, c(1, 2)], mean)
xx <- rownames(MEDIAS)</pre>
yy <- colnames(MEDIAS)</pre>
fila <- length(xx)
col <- length(yy)</pre>
total <- fila * col
x <- character(length = total)</pre>
y <- character(length = total)</pre>
z <- numeric(length = total)</pre>
k <- 0
for (i in 1:fila) {
    for (j in 1:col) {
        k < - k + 1
        x[k] \leftarrow xx[i]
        y[k] <- yy[j]
        z[k] <- MEDIAS[i, j]</pre>
    }
}
MEDIAS <- data.frame(ENV = x, GEN = y, Y = z)
x <- MEDIAS[, 1]
y \leftarrow MEDIAS[, 2]
z \leftarrow MEDIAS[, 3]
modelo2 \leftarrow lm(z \sim x + y)
for (i in 1:length(z)) {
    if (is.na(z[i]))
         z[i] <- predict(modelo2, data.frame(x = MEDIAS[i,</pre>
             1], y = MEDIAS[i, 2])
}
V=tapply(z,x,var)
V1=tapply(z,x,length)
V2=rep(V,V1)
z1=z/V2
MEDIAS <- data.frame(ENV = x, GEN = y, Y = z1)
modelo1 <- lm(Y ~ ENV + GEN, data = MEDIAS)</pre>
residual <- modelo1$residuals</pre>
MEDIAS <- data.frame(MEDIAS, RESIDUAL = residual)</pre>
```

```
mlabel <- names(MEDIAS)</pre>
 names(MEDIAS) <- c(mlabel[1:2], name.y, mlabel[4])</pre>
 OUTRES <- MEDIAS[order(MEDIAS[, 1], MEDIAS[, 2]), ]
 OUTRES2 <- by(OUTRES[, 4], OUTRES[, c(2, 1)], function(x) sum(x,
     na.rm = TRUE))
 OUTMED <- by(OUTRES[, 3], OUTRES[, c(2, 1)], function(x) sum(x,
     na.rm = TRUE))
 s <- svd(OUTRES2)
 U <- s$u
 L <- s$d
 V <- s$v
 L <- L[1:minimo]
 SS \leftarrow (L^2) * nrep
 SUMA <- sum(SS)
 percent <- round(((1/SUMA) * SS) * 100, 1)
 minimo <- min(ngen, nenv)
 DFAMMI <- rep(0, minimo)
 acum <- DFAMMI
 MSAMMI <- DFAMMI
 F.AMMI <- DFAMMI
 PROBF <- DFAMMI
 acumula <- 0
 for (i in 1:minimo) {
     DF \leftarrow (ngen - 1) + (nenv - 1) - (2 * i - 1)
     if (DF \le 0)
         break
     DFAMMI[i] <- DF
     acumula <- acumula + percent[i]</pre>
     acum[i] <- acum[i] + acumula</pre>
     MSAMMI[i] <- SS[i]/DFAMMI[i]
     F.AMMI[i] <- round(MSAMMI[i]/MSE, 2)</pre>
     PROBF[i] <- round(1 - pf(F.AMMI[i], DFAMMI[i], DFE),</pre>
         4)
}
 SS <- round(SS, 6)
 MSAMMI <- round(MSAMMI, 6)
 SSAMMI <- data.frame(percent, acum, Df = DFAMMI, 'Sum Sq' = SS,
     'Mean Sq' = MSAMMI, 'F value' = F.AMMI, Pr.F = PROBF)
 nssammi <- nrow(SSAMMI)</pre>
 SSAMMI <- SSAMMI[SSAMMI$Df > 0, ]
 nss <- nrow(SSAMMI)</pre>
 row.names(SSAMMI) <- paste("PC", 1:nss, sep = "")</pre>
 cat("\nAnalysis\n")
 print(SSAMMI)
 LL <- sqrt(diag(L))</pre>
 SCOREG <- U %*% LL
 SCOREE <- V %*% LL
 SCORES <- rbind(SCOREG, SCOREE)</pre>
 colnames(SCORES) <- paste("PC", 1:nssammi, sep = "")</pre>
 MSCORES <- SCORES[1:ngen, ]
 NSCORES <- SCORES[(ngen + 1):(ngen + nenv), ]</pre>
 MGEN <- data.frame(type = "GEN", Y = apply(OUTMED, 1, mean),
```

```
MSCORES)
MENV <- data.frame(type = "ENV", Y = apply(OUTMED, 2, mean),</pre>
    NSCORES)
bplot <- rbind(MGEN, MENV)</pre>
bplot <- bplot[, 1:(nss + 2)]</pre>
mlabel <- names(bplot)</pre>
names(bplot) \leftarrow c(mlabel[1], name.y, mlabel[c(-1, -2)])
maxy <- max(bplot[, 4])</pre>
miny <- min(bplot[, 4])
maxx <- max(bplot[, 3])</pre>
minx <- min(bplot[, 3])</pre>
row.names(bplot) <- c(row.names(MGEN), row.names(MENV))</pre>
cp.name <- rownames(SSAMMI)[1:3]</pre>
cp.per <- SSAMMI[1:3, 1]</pre>
cp1 <- paste("PC 1 (", cp.per[1], ")", sep = "")</pre>
cp2 <- paste("PC 2 (", cp.per[2], ")", sep = "")</pre>
cp3 <- paste("PC 3 (", cp.per[3], ")", sep = "")</pre>
if (graph == "biplot") {
    plot(bplot[, 3], bplot[, 4], cex = 0, xlab = cp1, ylab = cp2,
         frame = TRUE, ...)
    if (number == TRUE) {
         text(MGEN[, 3], MGEN[, 4], cex = 0, text(MGEN[, 3],
             MGEN[, 4], labels = as.character(1:nrow(MGEN)),
             col = "blue"))
    }
    if (number == FALSE) {
         text(MGEN[, 3], MGEN[, 4], cex = 0, text(MGEN[, 3],
             MGEN[, 4], labels = row.names(MGEN), col = "blue"))
    }
    points(MENV[, 3], MENV[, 4], cex = 0, text(MENV[, 3],
        MENV[, 4], labels = row.names(MENV), col = "brown"))
    abline(h = 0, v = 0, lty = 2.5, col = "green", lwd = 2)
    s <- seq(length(MENV[, 3]))
    arrows(0, 0, 0.9 * MENV[, 3][s], 0.9 * MENV[, 4][s],
         col = "brown", lwd = 1.8, length = 0.1, code = 2)
}
if (graph == "triplot") {
    y <- bplot
    lugar \leftarrow y[y[, 1] == "ENV", 3:5]
    clones <- y[y[, 1] == "GEN", 3:5]
    maxcp \leftarrow max(y[, 3:5])
    mincp \leftarrow min(y[, 3:5])
    rango <- maxcp - mincp
    clones <- (clones - mincp)/rango</pre>
    nclon <- nrow(clones)</pre>
    lugar <- (lugar - mincp)/rango</pre>
    nlugar <- nrow(lugar)</pre>
    point1 <- cbind(clones[, 1], clones[, 2], clones[, 3])</pre>
    point2 <- cbind(lugar[, 1], lugar[, 2], lugar[, 3])</pre>
    point3 <- cbind(c(0.6, 0.6, 0), c(0, 0.8, 0.8), c(0.6,
```

```
0, 0.6))
       suppressWarnings(warning(triplot(point1, cex = 0, grid = TRUE,
          label = "", center = TRUE, frame = TRUE)))
       if (number == TRUE)
          suppressWarnings(warning(text(tritrafo(point1), as.character(1:nclon),
              adj = c(0.5, 0), col = "blue", cex = 0.8)))
       if (number == FALSE)
          suppressWarnings(warning(text(tritrafo(point1), rownames(clones),
              adj = c(0.5, 0), col = "blue", cex = 0.8)))
       suppressWarnings(warning(text(tritrafo(point2), rownames(lugar),
          adj = c(0.5, 0), col = "red", cex = 0.8)))
       suppressWarnings(warning(text(tritrafo(point3), c(cp1,
          cp2, cp3), adj = c(0.5, 0), cex = 1)))
       trilines(centerlines(3), lty = 2.5, col = "green", lwd = 2)
       for (i in 1:nlugar) {
          suppressWarnings(warning(trilines(c(point2[i, 1],
              1/3), c(point2[i, 2], 1/3), c(point2[i, 3], 1/3),
              col = "red", lty = 1)))
       }
   }
   return(list(genXenv = OUTRES2, analysis = SSAMMI, means = MEDIAS,
       biplot = bplot))
}
dados<-read.csv2("dados.csv",head=TRUE);dados
model<- AMMI1(ENV,GEN,REP, Y, MSE,graph="biplot");model</pre>
model<- AMMI1(ENV,GEN,REP, Y, graph="biplot",number=FALSE)</pre>
library(GGEBiplotGUI)
dados<-read.csv2("DadosGGE2.csv",head=TRUE);dados
GGEBiplot(Data = dados)
####### GGE biplot ponderado ##############
library(GGEBiplotGUI)
dados2<-read.csv2("DadosGGEponderado2.csv",head=TRUE);dados2
GGEBiplot(Data = dados2)
```