



**INTERACCIÓN GENOTIPO – AMBIENTE:
GGE BILOT Y MODELOS AMMI**

M^a Elisa Frutos Bernal

Tutor: Purificación Galindo Villardón

UNIVERSIDAD DE SALAMANCA

Departamento de Estadística

Máster en Análisis Avanzado de Datos Multivariantes

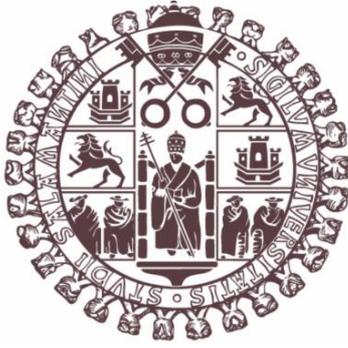
Trabajo Fin de Máster

**INTERACCIÓN GENOTIPO –
AMBIENTE:
GGE BILOT Y MODELOS AMMI**

AUTOR: Elisa Frutos Bernal

TUTOR: Purificación Galindo Villardón

Año 2011



Dpto. de Estadística

Universidad de Salamanca

PURIFICACIÓN GALINDO VILLARDÓN

Directora del Departamento de Estadística de la Universidad de Salamanca

CERTIFICA que **D.^a Elisa Frutos Bernal** ha realizado en la Universidad de Salamanca, bajo su dirección, el trabajo que para optar título de Máster en Análisis Avanzado de Datos Multivariantes presenta con el título ***Interacción Genotipo-Ambiente: GGE Biplot y Modelos AMMI***, autorizando expresamente su lectura y defensa.

Y para que conste, firma el presente certificado en Salamanca a 10 de julio de 2011.

Purificación Galindo Villardón

INDICE

RESUMEN	6
INTRODUCCIÓN	7
OBJETIVOS	12
MATERIAL Y MÉTODOS	13
CAPÍTULO I: LOS MÉTODOS BILOT	14
1.1. INTRODUCCIÓN	14
1.2. MÉTODOS BILOT CLÁSICOS	16
1.2.1 DEFINICIÓN Y FUNDAMENTACIÓN TEÓRICA.....	16
1.2.2 OBTENCIÓN DE LOS MARCADORES.....	17
1.2.2.1 GH-BILOT O CMP-BILOT (COLUMN METRIC PRESERVING).....	19
1.2.2.2 JK-BILOT O RMP-BILOT (ROW METRIC PRESERVING)	20
1.2.2.3 SQRT-BILOT O BILOT SIMÉTRICO	21
1.2.2.4 HJ-BILOT O RCMP-BILOT	21
1.2.3 PROPIEDADES DE LOS MARCADORES	22
1.2.3.1 GH-BILOT O CMP-BILOT.....	22
1.2.3.2 JK-BILOT O RMP-BILOT	23
1.2.3.3 HG-BILOT O CRMP BILOT	24
1.2.4 CALIDAD DE REPRESENTACIÓN.....	25
1.2.4.1 CALIDAD DE REPRESENTACIÓN DE LOS GH-BILOT	25
1.2.4.2 CALIDAD DE REPRESENTACIÓN DEL JK-BILOT	26
1.2.4.3 CALIDAD DE REPRESENTACIÓN DEL HJ-BILOT	26
1.2.5 REPRESENTACIONES BILOT: INTERPRETACIÓN DE LOS RESULTADOS	26
1.2.5.1 ÍNDICES DE AYUDA A LA INTERPRETACIÓN	26
1.2.5.2 INTERPRETACIÓN GEOMÉTRICA	27
1.3 PROGRAMA GGEBIPLOR	29
CAPÍTULO II: MODELOS CON EFECTO INTERACCIÓN MULTIPLICATIVO (AMMI).....	31
2.1. INTRODUCCIÓN	31
2.1. FUNDAMENTO TEÓRICO.....	33
2.2. PROGRAMA GGEBIPLOR	37
CAPÍTULO III: GGE BILOT	39
3.1. INTRODUCCIÓN	39
3.2. CONCEPTO DE GGE Y GGE BILOT.....	40
3.3. ANALISIS BILOT DE MET	43
3.3.1. COMPARACIONES SIMPLES UTILIZANDO EL GGE BILOT.....	46
3.3.1.1. RENDIMIENTO DE LOS DISTINTOS CULTIVOS EN UN AMBIENTE DADO.....	47
3.3.1.2. ADAPTACIÓN DE UN CULTIVO A LOS DISTINTOS AMBIENTES	48
3.3.1.3. COMPARACIÓN ENTRE DOS GENOTIPOS.....	49
3.3.2. INVESTIGACIÓN DE MEGA – AMBIENTES.....	51
3.3.2.1. WHICH-WON-WHERE	51
3.3.2.2. INVESTIGACIÓN DE MEGA-AMBIENTES	53
3.3.3. EVALUACIÓN DE CULTIVOS EN UN MEGA-AMBIENTE.....	53
3.3.3.1. RENDIMIENTO MEDIO Y ESTABILIDAD DE LOS GENOTIPOS	54
3.3.3.2. ORDENACIÓN DE GENOTIPOS RESPECTO AL GENOTIPO “IDEAL”	55
3.3.4. EVALUACIÓN DE AMBIENTES DE PRUEBA	57
3.3.4.1. RELACIONES ENTRE LOS AMBIENTES.....	57
3.3.4.2. PODER DISCRIMINANTE DE LOS ENTORNOS DE PRUEBA	57
3.3.4.3. REPRESENTATIVIDAD DE LOS ENTORNOS DE PRUEBA	58

3.3.4.4. AMBIENTE “IDEAL” DE PRUEBA PARA LA SELECCIÓN DE GENOTIPOS:.....	60
CAPÍTULO IV: PROGRAMA EN R PARA EL ANÁLISIS GGE.....	62
CONCLUSIONES	69
GLOSARIO.....	70
BIBLIOGRAFIA	71

RESUMEN

Los fenotipos son una mezcla del genotipo (G), el medio ambiente (E) y la interacción entre ellos (GEI). La interacción complica el proceso de selección de genotipos con rendimiento superior. Se han desarrollado numerosos métodos para el estudio de la interacción como son los modelos AMMI (GAUCH 1992) y el GGE Biplot (YAN et al. 2000).

El análisis BIPLLOT se ha convertido en una herramienta estadística muy importante en el fitomejoramiento y en la investigación científica. En este trabajo se revisan los principios básicos del análisis Biplot y su utilización en el análisis de datos de ensayos multi-ambiente (MET). Además se presentan los modelos AMMI donde se combinan las técnicas de Análisis de la Varianza y Análisis de Componentes Principales y que tienen como objetivo explicar la interacción asociada a un ANOVA bifactorial, a partir de una representación BIPLLOT. También se presenta el GGE Biplot como alternativa a los modelos AMMI. El GGE Biplot trabaja con la matriz en la que se combina el efecto principal del genotipo y la interacción genotipo ambiente (G+GE). El GGE Biplot es una herramienta de visualización de datos, que muestra gráficamente la interacción GEI en una tabla de dos vías. El GGE Biplot es eficaz para el análisis de mega-ambientes (which-won-where), la evaluación de genotipos (rendimiento medio y estabilidad) y la evaluación de medio ambientes (poder discriminante). Finalmente se presenta un software desarrollado en R que permite la utilización del GGE Biplot sobre tablas de dos vías del tipo Genotipo-Ambiente ó Genotipo-Tratamiento.

Palabras clave: Biplot, HJ Biplot, AMMI, GGE, GEI.

INTRODUCCIÓN

Comprender la relación entre el rendimiento de los cultivos y el medio ambiente ha sido un problema clave para mejoradores y genetistas. El rendimiento de los cultivos está en función del genotipo, el ambiente y la interacción genotipo-ambiente (GEI). Se dice que existe interacción genotipo ambiente cuando diferentes cultivos o genotipos responden de forma diferente en diversos ambientes. Los investigadores coinciden en que la interacción genotipo ambiente solo es importante cuando es significativa y causa un cambio significativo en la ordenación de los genotipos en los distintos ambientes. Para que la interacción pueda ser detectada a través de procedimientos estadísticos es necesario tener al menos dos genotipos distintos evaluados en al menos dos ambientes. Por lo tanto un modelo básico que incluye la interacción es el siguiente:

$$P = G + E + GE$$

Este modelo se puede escribir como: $P_{ij} = \mu + G_i + E_j + (GE)_{ij}$, donde μ es la media global. Por tanto puede desprenderse de este modelo que para un genotipo dado puede haber muchos fenotipos dependiendo del medio ambiente y de la GEI.

La interacción genotipo ambiente se puede visualizar al comparar el rendimiento relativo de los genotipos X e Y en dos entornos. Los distintos tipos de interacción se muestran en las figuras 1, 2 y 3

La primera gráfica (Figura 1) muestra que los genotipos X e Y responden de forma similar en los dos ambientes, debido a que sus respuestas son paralelas. El genotipo X se comporta mejor que en el genotipo Y en los dos entornos. Este caso representa la ausencia de interacción.

La Figura 2 muestra interacción no crossover. Los genotipos X e Y responden de manera diferente a los entornos pero su ordenación permanece sin cambios. La respuesta de los dos genotipos a los ambientes no es aditiva. La magnitud de la diferencia aumenta.

El siguiente gráfico (Figura 3) muestra interacción de tipo crossover. La dirección en la que se modifican los genotipos X e Y en los ambientes es opuesta, el rendimiento de X se incrementa mientras que el de Y se decrementa.

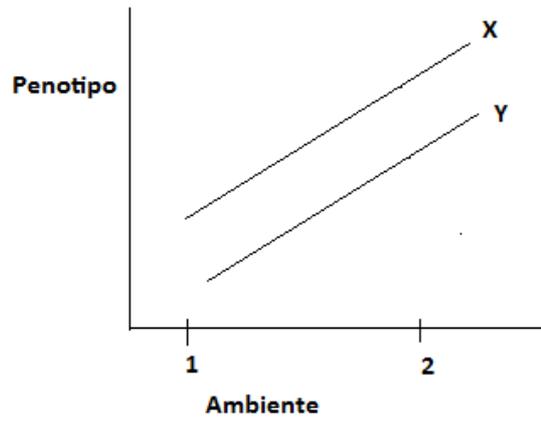


Figura 1. No interacción

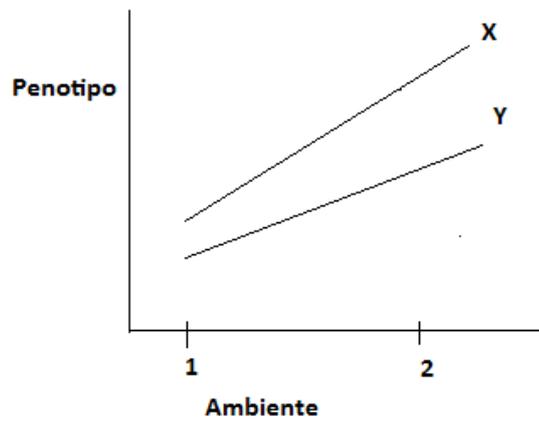


Figura 2. Interacción no crossover

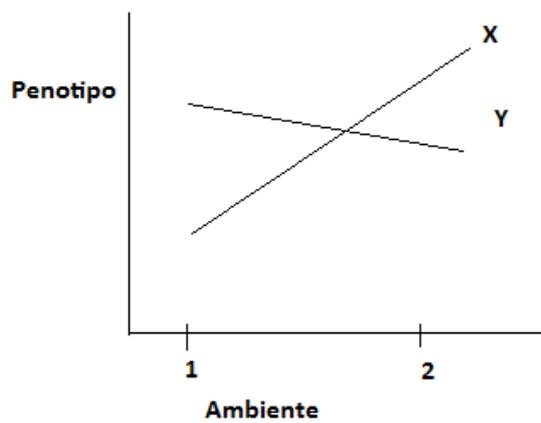


Figura 3. Interacción crossover

GAUCH y ZOBEL (1996) explican la importancia de la interacción genotipo ambiente de la siguiente forma: "Si no hubiera interacción, una sola variedad de

trigo (*Triticum aestivum* L.), de maíz (*Zea mays* L.) ó cualquier otro cultivo rendiría más en todo el mundo y además el ensayo de cultivos se podría realizar en una sola ubicación y proporcionar resultados universales. Además donde no hubiera ruido los resultados experimentales deberían ser exactos identificando la mejor variedad sin errores, y no habría necesidad de replicas. Por tanto una réplica en un lugar debería identificar la mejor variedad de trigo que florece en todo el mundo”.

Los métodos estadísticos que más atención han recibido recientemente son el análisis de patrones (DELACY et al. 1996) y el modelo de efectos principales aditivos e interacción multiplicativa (AMMI) (GAUCH y ZOBEL 1996). El modelo AMMI incorpora componentes aditivas y multiplicativas en la tabla de dos vías y permite tener en cuenta la interacción subyacente de forma eficiente. Integrando los resultados obtenidos a partir del Biplot con los de estabilidad de los genotipos se pueden agrupar los genotipos con base en la similitud de respuesta y el grado de estabilidad del rendimiento a través de los ambientes. El modelo AMMI o variaciones del mismo (por ejemplo el modelo FANOVA) se han utilizado para interpretar la interacción genotipo ambiente en diversas situaciones (GUTIERREZ et al. 1994, YAN y HUNT 1998).

El GGE Biplot se ha desarrollado recientemente y se basa en el hecho de que en la variación fenotípica total, E explica la mayor parte de la variación y G y GE suelen ser pequeñas (YAN 2002). Sin embargo sólo G y la interacción GE son relevantes en la evaluación de cultivos sobre todo cuando la interacción es repetible en diferentes años. Por lo que YAN et al (2000) los pone juntos y lo llama GGE. Siguiendo la propuesta de GABRIEL (1971), se utiliza la técnica del Biplot para visualizar el GGE de los datos MET, refiriéndonos al mismo como GGE Biplot (YAN 2001, YAN et al. 2000).

El GGE Biplot es una herramienta de visualización de datos que muestra gráficamente la interacción en una tabla de dos vías (YAN 2000). El GGE Biplot es eficaz para el análisis de mega-ambientes (por ejemplo el patrón which-won-where), la evaluación de genotipos (desde el punto de vista de la estabilidad y el rendimiento) y la evaluación de ambientes (poder discriminante de los mismos).

Para realizar un análisis GGE actualmente existe sólo un programa comercial denominado GGE Biplot Software. De ahí que dado el interés del tema presentado consideremos interesante la existencia de software libre que permita realizar los análisis anteriores.

En este trabajo se presenta el programa GGEbiplotR realizado en el lenguaje R. R es un conjunto integrado de programas para manipulación de datos, cálculo y gráficos. Entre otras características dispone de:

- Almacenamiento y manipulación efectiva de datos.
- Operadores para cálculo sobre variables indexadas (arrays), en particular matrices.
- Una amplia, coherente e integrada colección de herramientas para análisis de datos.
- Posibilidades gráficas para análisis de datos, que funcionan directamente sobre pantalla o impresora.
- Un lenguaje de programación bien desarrollado, simple y efectivo, que incluye condicionales, ciclos, funciones recursivas y posibilidad de entradas y salidas.

En el primer capítulo de este trabajo damos la definición de Biplot y las propiedades fundamentales de los diferentes tipos de representaciones. Se ofrecen los elementos necesarios que ayudan a interpretar este tipo de gráfico, haciendo énfasis en la información relacionada con los datos que podemos obtener a partir del Biplot.

El segundo capítulo lo dedicamos al análisis de los modelos con efectos interacción multiplicativo (AMMI) propuestos por GAUCH en 1988.

En el capítulo tercero se presenta el GGE Biplot. Comienza con una introducción al concepto de GGE Biplot y posteriormente se realiza una exposición detallada del análisis GGE Biplot de datos MET. Se aborda el análisis de los ambientes y los genotipos desde todas las perspectivas:

- Rendimiento de los distintos cultivos en un ambiente dado.
- Adaptación de un cultivo a los distintos ambientes.
- Comparación entre dos cultivos.
- Investigación de mega-ambientes (which-won-where).
- Evaluación de cultivos en un mega-ambiente:
 - Rendimiento medio y estabilidad de los genotipos.
 - Ordenación de los genotipos respecto al genotipo ideal.
- Evaluación de ambientes de prueba:
 - Relación entre los ambientes de prueba.
 - Poder discriminante de los entornos de prueba.

- Representatividad de los entornos de prueba.
- Ambiente ideal de prueba para la selección de genotipos.

En cada uno de los capítulos anteriores se presentan las herramientas disponibles en el programa GGEbiplotR relacionadas con el análisis presentado. Todos los gráficos que aparecen en el trabajo se han realizado utilizando dicho programa.

En el capítulo cuarto se presentan las características generales del programa GGEbiplotR.

OBJETIVOS

El objetivo general de este trabajo es presentar algunos de las técnicas estadísticas existentes para el análisis de la interacción genotipo ambiente en la evaluación de cultivos.

Como objetivos específicos tenemos los siguientes:

- Describir brevemente el Biplot: propiedades fundamentales de las distintas factorizaciones (GH-Biplot, JK-Biplot, HJ-Biplot y SQRT-Biplot) y elementos necesarios para interpretar este tipo de gráficos.
- Describir brevemente los modelos AMMI como técnica estadística utilizada en el análisis de la interacción genotipo ambiente, con el objetivo de clasificar los genotipos en estables e inestables a partir de su interacción con el ambiente.
- Describir el GGE Biplot como técnica estadística que permite el análisis de datos genotipo-ambiente. En concreto:
 - Describir su utilidad en la evaluación de genotipos
 - Describir su utilidad en la evaluación de entornos de prueba detectando la presencia o ausencia de mega-ambientes
 - Describir su utilidad en la detección de los entornos de prueba más discriminantes y/o representativos
 - Describir su utilidad en la evaluación de genotipos basada en rendimiento y estabilidad.
- Presentar el programa GGEBiplotR desarrollado en R y que permite realizar los análisis anteriores.

MATERIAL Y MÉTODOS

Como ejemplo para demostrar la utilidad del GGE Biplot y presentar el programa GGEbiplotR se utilizan los datos de los ensayos de rendimiento del trigo de invierno en Ontario 1993. El conjunto de datos consta de 18 cultivos en 9 localidades, con 4 bloques o repeticiones en cada lugar. Se utilizan los valores medios de cada combinación genotipo-localización.

CAPÍTULO I: LOS MÉTODOS BILOT

1.1. INTRODUCCIÓN

Los métodos BILOT (GABRIEL 1971) son una representación gráfica a baja dimensión de una matriz de datos X (I individuos y J variables), es decir una matriz de Datos Multivariantes.

Desde el punto de vista algebraico, el Biplot se basa en el mismo principio sobre el que se sustentan la mayoría de las técnicas factoriales de reducción de dimensionalidad, es decir, la Descomposición en Valores y Vectores Singulares (SVD) de una matriz. La diferencia fundamental es que en el Biplot se trata de reproducir el dato y se incorpora una representación conjunta de individuos y variables.

Utilizando SVD se aproxima la matriz de datos, para luego realizar un gráfico en baja dimensión cuya interpretación se basa en las propiedades geométricas del producto escalar entre vectores fila (denominados marcadores fila) y vectores columna (denominados marcadores columna), de tal forma que dicho producto reproduzca aproximadamente cada elemento de la matriz de datos X .

Las dos factorizaciones BILOT más importantes propuestas por GABRIEL (1971) fueron denominadas: GH-Biplot y JK-Biplot. El GH-Biplot consigue una alta calidad en la representación de las columnas (variables), mientras que el JK-Biplot logra una alta calidad en la representación de las filas (individuos). Por tanto, la representación Biplot no es simultánea para ninguno de estos dos tipos de Biplots porque la bondad de ajuste no es la misma para las filas y para las columnas de la matriz de datos.

GALINDO (1985, 1986) demuestra que con una conveniente elección de los marcadores es posible representar las filas y las columnas simultáneamente sobre un mismo sistema de coordenadas, con una alta calidad de representación tanto para las filas como para las columnas. GALINDO denomina a este tipo de Biplot HJ-Biplot.

En 1984, GREENACRE propone una nueva terminología para el GH y el JK Biplot, denominándolos respectivamente CMP (Column Metric Preserving) y RMP (Row Metric Preserving), aludiendo al hecho de que con el primer tipo se preserva la métrica para las columnas y con el segundo tipo la métrica para las filas.

El BIPLLOT no solamente se utiliza con fines descriptivos; puede ser aplicado también en la diagnosis de modelos (BRADU y GABRIEL 1974, 1978). En este sentido podemos ver que con una simple inspección de la posición geométrica de los marcadores se puede diagnosticar el modelo que mejor describe los datos, y por tanto la presencia o no de interacción en caso de que proceda (GOWER 1990, DIAZ-LENO 1995, BLÁZQUEZ 1998).

El BIPLLOT, ha dado lugar a nuevos métodos de análisis multivariante de datos, al ser combinado con otras técnicas clásicas; en este sentido podemos citar los modelos AMMI (GOLLOB 1968), el cual inicialmente combina las técnicas de Análisis de Varianza y Análisis de Componentes Principales, y posteriormente incorpora el BIPLLOT en lugar del A.C.P. (GABRIEL 1978, KEMPTON 1984, GAUCH 1988). Consiste en hacer un BIPLLOT a la matriz de residuales del modelo.

Otra técnica que utiliza el BIPLLOT es el GGE Biplot (YAN et al. 2000, YAN 2001). El GGE Biplot utiliza el modelo SREG, que incluye G+GE en el término bilineal. Con el GGE Biplot podemos identificar los genotipos de mayor producción en cada ambiente y además caracterizar los genotipos que tienen mayor estabilidad. Por otra parte el gráfico permite visualizar el agrupamiento de genotipos y ambientes (mega-ambientes) con patrones similares de respuesta, así como identificar los ambientes más representativos y los más discriminatorios (YAN y RAJCAN 2002).

Ambos métodos han sido aplicados en problemas agrícolas, más específicamente en el Análisis de la interacción Genotipo-Ambiente.

Actualmente continúan las investigaciones relacionadas con el BIPLLOT. Las dos tendencias iniciales de investigación siguen desarrollándose (descripción y diagnosis). Sobre todo se trata de combinar el BIPLLOT con los métodos clásicos del análisis de datos; surgiendo nuevas técnicas de análisis cuya información queda resumida en un BIPLLOT.

1.2. MÉTODOS BILOT CLÁSICOS

1.2.1 Definición y fundamentación teórica.

Los métodos Biplot como toda técnica del Análisis Multivariante, tiene como objetivo representar la información contenida en una matriz de datos en un espacio de menor dimensión, generalmente el plano, con la menor pérdida de información.

Por tanto los **datos de partida** para la construcción de un BILOT, están contenidos en una matriz que denominaremos X de orden $(I \times J)$ y de rango r , en general las I filas corresponden a los individuos y las J columnas a las variables.

La **fundamentación teórica** de los Biplots Clásicos, se basa en la aproximación de la matriz de datos X por una de menor rango q , siendo $q < r$, a través de la Descomposición en Valores Singulares de X . Se realiza luego, una factorización en matrices de marcadores filas y de marcadores columnas de manera que el producto escalar entre los marcadores aproxime "lo mejor posible" los valores de X .

Si la matriz X es de rango dos, es posible lograr una representación exacta sobre el plano. Cuando el rango de X es mayor a dos, entonces la factorización Biplot garantiza la representación gráfica aproximada de dicha matriz.

Definición: Un Biplot para una matriz de datos X $(I \times J)$ es una representación gráfica mediante marcadores (vectores): a_1, a_2, \dots, a_I para las filas y b_1, b_2, \dots, b_J para las columnas de X , de forma tal que el producto interno aproxime el elemento x_{ij} de la matriz de partida lo mejor posible.

Si consideramos los marcadores a_1, a_2, \dots, a_I como filas de una matriz A y los marcadores b_1, b_2, \dots, b_J como filas de una matriz B , entonces podemos escribir:

$$X \cong AB'$$

Tanto los marcadores a_i para las filas, como los marcadores b_j para las columnas estarán representados en un espacio de dimensión $q \leq r$, siendo q el número de ejes retenidos y r el rango de X .

La factorización Biplot garantiza la representación gráfica aproximada de la matriz, ya que cada x_{ij} puede reconstruirse de la siguiente forma:

$$x_{ij} = a_i' b_j \quad (i = 1 \dots I; j = 1 \dots J)$$

Considerando la aproximación Biplot de los elementos de la matriz X y aplicando la definición de producto escalar tenemos que:

$$x_{ij} = a_i' b_j = \| \text{Proy}(a_i/b_j) \| (\text{signo}) \| b_j \|$$

Siendo:

$\| \text{Proy}(a_i/b_j) \|$ = longitud de la proyección de a_i sobre b_j

$\| b_j \|$ = módulo del segmento que une el origen con el extremo del vector b_j .

Por tanto el problema ahora radica en obtener los marcadores para las filas y columnas de X, es decir obtener las matrices de marcadores A y B.

1.2.2 Obtención de los marcadores

El objetivo por tanto es encontrar una matriz $X_{(q)} = A_{(q)} B_{(q)}$ de rango q que aproxime lo mejor posible a X, en el sentido de los mínimos cuadráticos, es decir $X_{(q)}$ debe minimizar la siguiente expresión:

$$\sum_i \sum_j (x_{ij} - x_{(q)ij})^2 = \text{traza} [(X - X_{(q)})(X - X_{(q)})']$$

Para todas las matrices $X_{(q)}$ de rango q o menor.

El método más conocido para aproximar una matriz a bajo rango es el propuesto por ECKART y YOUNG (1936, 1939) que puede encontrarse también en YOUNG y HOUSEHOLDER (1938), GABRIEL (1971), GREENACRE (1984), entre

otros autores. Se basa en la descomposición en valores y vectores singulares de la matriz que deseamos aproximar:

$$X = UDV'$$

Siendo:

X (IxJ): la matriz de datos.

r: es el rango de X ($r \leq \min(I,J)$)

U (Ixr): la matriz cuyas columnas contienen los vectores propios de XX'

V (Jxr): la matriz cuyas columnas contienen los vectores propios de $X'X$

D(rxr): matriz diagonal que contiene los valores propios de X.

Debe cumplirse que $U'U = V'V = I$, es decir, las columnas de U y V son ortonormales para garantizar así la unicidad de la factorización.

Por lo tanto, la mejor aproximación de X en rango q ($X_{(q)}$) es:

$$X_{(q)IxJ} = U_{(q)Ixq} D_{(q)qxq} V'_{(q)qxJ} = \sum_{k=1}^q \lambda_k u_k v_k'$$

Donde $U_{(q)}$ y $V_{(q)}$ son las matrices construidas con las q primeras columnas de U y V respectivamente, mientras que $D_{(q)}$ es la matriz diagonal que contiene los q mayores valores singulares distintos de cero de X (λ_k).

Un algoritmo para el cálculo puede verse en GOLUB y REINSCH (1971).

Por lo tanto:

$$X = AB' = UDV'$$

La selección de distintas factorizaciones conduce a distintos marcadores y por ende a distintos Biplots. En el caso de la métrica identidad, se pueden elegir los marcadores de acuerdo a los distintos valores de γ en la siguiente descomposición:

$$A = UD^\gamma \quad B = VD^{1-\gamma} \quad (1.1)$$

Dependiendo del valor seleccionado para γ ($\gamma = 0, 1, \frac{1}{2}$) tenemos los Biplots Clásicos de GABRIEL (1971): el GH-Biplot, el JK-Biplot y el SQRT-Biplot respectivamente.

GREENACRE en 1984, introduce una nueva terminología para el GH y el JK Biplot, denominándolos respectivamente CMP (**C**olumn **M**etric **P**reserving) y RMP (**R**ow **M**etric **P**reserving), aludiendo al hecho de que con el primer tipo se preserva la métrica para las columnas y con el segundo tipo la métrica para las filas. Esta terminología ha sido aceptada por GABRIEL, razón por la cual en las referencias más recientes se opta por la misma.

Definiremos a continuación los diferentes tipos de Biplot Clásicos (GABRIEL 1971) para después desarrollar el Biplot introducido por GALINDO (1986).

1.2.2.1 GH-Biplot o CMP-Biplot (Column Metric Preserving)

Cuando en la expresión (1.1) $\gamma = 0$ tenemos el GH-Biplot o CMP (Column Metric Preserving), luego tendremos lo siguiente:

$$X = AB' = (UD^\gamma)(D^{1-\gamma}V') = (U)(DV')$$

Luego la factorización en el CMP-Biplot corresponde a la elección de marcadores siguiente:

$$A = U \quad B = VD$$

Como trabajamos con la métrica identidad en el espacio de las filas, A debe verificar que: $\mathbf{A}'\mathbf{A}=\mathbf{I}$.

$$\text{Luego: } X'X = (AB')'AB' = BA'AB' = BB'.$$

Es decir:

$$X'X = BB'$$

Por lo tanto este Biplot preserva la métrica euclídea usual entre las columnas pero no entre las filas, por lo que se obtiene una alta calidad de representación para las columnas.

1.2.2.2 JK-Biplot o RMP-Biplot (Row Metric Preserving)

Cuando en la expresión (1.1) $\gamma = 1$ se define el JK-Biplot o RMP-Biplot (Row Metric Preserving), luego tenemos lo siguiente:

$$X = AB' = (UD)(V')$$

La elección de los marcadores es:

$$\mathbf{A} = \mathbf{UD} \quad \mathbf{B} = \mathbf{V}$$

En este tipo de Biplot se impone la métrica $\mathbf{B}'\mathbf{B}=\mathbf{I}$ en el espacio de las filas de la matriz $X_{(I \times J)}$.

Eligiendo los marcadores según la expresión:

$$XX' = AB'(AB')' = AB'BA' = AA'$$

Luego:

$$XX' = AA'$$

De ahí el nombre de este tipo de Biplot (RMP-Biplot), el cual preserva la métrica usual entre las filas pero no entre las columnas, obteniéndose alta calidad de representación para las filas.

A este Biplot GABRIEL lo llamó JK-Biplot porque utilizó J para denotar la matriz de marcadores fila y K la matriz de marcadores columna.

1.2.2.3 SQRT-Biplot o Biplot Simétrico

Si en la expresión (1.1) se elige $\gamma = \frac{1}{2}$ se obtiene lo que GABRIEL denominó SQRT-Biplot o Biplot-Simétrico, donde:

$$X = AB' = (UD^{1/2})(D^{1/2}V')$$

Este Biplot asigna papeles simétricos tanto a las filas como a las columnas de X.

La elección de marcadores es:

$$A = UD^{1/2} \quad B = VD^{1/2}$$

1.2.2.4 HJ-Biplot o RCMP-Biplot

El HJ-Biplot propuesto por GALINDO (1985, 1986) es una representación gráfica multivariante de marcadores fila y columna, elegidos de forma que puedan superponerse en el mismo sistema de referencia con máxima calidad de representación.

El HJ-Biplot, tiene como objetivo principal la interpretación simultánea de las relaciones entre individuos (filas) y variables (columnas) de la matriz de datos X, no siendo, a diferencia del GH-Biplot y del JK-Biplot, la búsqueda de aproximación de elementos de la matriz de datos X.

Los elementos de la matriz X están centrados por filas y columnas, por lo que la métrica introducida en el espacio de las filas es equivalente a la inversa de la matriz de covarianzas entre variables, mientras que en el espacio de las columnas la métrica es equivalente a la inversa de la matriz de dispersión entre los individuos.

Si la DVS de X es:

$$X = UDV'$$

Se elige la siguiente factorización:

$$A = UD \quad B = VD$$

Dado que en el HJ-Biplot se puede hacer una representación simultánea de filas y columnas se lo denomina también RCMP-Biplot (Row-Column-Metric-Preserving).

GALINDO (1996) lo denominó HJ-Biplot, ya que siguiendo la nomenclatura de GABRIEL adoptó la siguiente notación para las matrices de marcadores fila y columna:

$$J = UD \quad H = VD$$

En el HJ-Biplot, la representación de filas y columnas de la tabla de datos es simétrica por lo que se obtiene la misma calidad de representación tanto para los individuos (filas) como para las variables (columnas) (GALINDO y CUADRAS, 1986).

1.2.3 Propiedades de los marcadores

Las propiedades de los marcadores filas y columnas dependen del Biplot elegido. El conocimiento de estas propiedades es de fundamental importancia a la hora de la elección del tipo de Biplot a utilizar así como para la interpretación de los resultados obtenidos.

1.2.3.1 GH-Biplot o CMP-Biplot

Recordemos que en este tipo de Biplot la métrica utilizada es $A'A=I$.

Propiedades de los marcadores columna ($B=VD$):

1. El producto escalar de las columnas de X , coincide con el producto escalar de los marcadores columna, o sea: $x_j'x_{j'} = b_j'b_{j'}$, o bien en forma matricial $X'X = B'B$
2. Si los elementos de la matriz X , han sido centrados por columnas, se obtiene el Biplot de Componentes Principales. En este caso la longitud al cuadrado de los vectores que representan a los marcadores columna, aproximan la covarianza entre las variables correspondientes:

$$b_j'b_{j'} = COV(y_j, y_{j'})$$

De esta propiedad se desprenden las tres propiedades siguientes:

- 2.1. La longitud al cuadrado de los marcadores columna aproxima la varianza de la variable X , y por lo tanto la longitud, aproxima la desviación estándar.

$$\|b_j\| = \|x_j\| = \sqrt{Var(x_j)}$$

- 2.2. El coseno del ángulo que forman dos marcadores columna, aproxima la correlación entre las variables correspondientes.
- 2.3. La distancia euclídea entre dos vectores columna, coincide con la distancia entre los respectivos marcadores columna.

Propiedades de los marcadores fila ($A=U$):

La distancia de Mahalanobis entre dos filas (i y s) de X se aproxima mediante la distancia euclídea entre dos marcadores fila (i y s).

1.2.3.2 JK-Biplot o RMP-Biplot

En este tipo de Biplot la métrica utilizada es $B'B=I$.

Propiedades de los marcadores filas (A=UD):

1. El producto escalar entre marcadores filas coincide con el producto escalar, con la métrica identidad, de las filas de la matriz X:

$$XX' = AA'$$

2. La distancia entre marcadores fila, coincide con la distancia euclídea entre filas de la matriz X.
3. Los marcadores fila coinciden con las coordenadas de las filas en el espacio de las componentes principales. Esta propiedad implica que podemos estudiar similitudes entre individuos con pérdida mínima de información.

Propiedades de los marcadores columnas (B=V):

1. Las coordenadas para las columnas coinciden con las proyecciones de los ejes originales (base canónica) en el espacio de las componentes principales.
2. El producto interno entre marcadores columna aproxima el producto interno entre columnas de X con la métrica (XX'), o sea:

$$x_j'(XX')x_{j'} = b_j'b_{j'}$$

3. La similitud entre columnas se aproxima utilizando la inversa de la matriz de dispersión entre individuos, no siendo posible interpretar los ángulos en términos de correlación.

1.2.3.3 HG-Biplot o CRMP Biplot

En este biplot, como los marcadores para las columnas coinciden con los marcadores del GH-Biplot y los marcadores para las filas coinciden con los del JK-Biplot, entonces las propiedades coinciden con las que hemos descrito anteriormente para dichos marcadores. Sólo se agregan aquí las relacionadas con las relacionadas con las representaciones simétricas de filas y columnas.

1. Los marcadores fila y columna se pueden representar en el mismo sistema de referencia.

1.2.4 Calidad de representación

Como dijimos al principio de este capítulo el objetivo de los métodos Biplot es encontrar una matriz $X_{(q)} = A_{(q)}B'_{(q)}$ de rango q , que aproxime lo mejor posible a la matriz de datos X de rango r , siendo $q < r$. Si la matriz X es de rango dos, es posible lograr una representación exacta sobre el plano, en otro caso se harán solo aproximaciones de X , y en estos casos, que es lo que sucede en la mayoría de las ocasiones es importante evaluar la calidad de la representación.

1.2.4.1 Calidad de representación de los GH-Biplot

La calidad de representación global se calcula como el cociente entre la suma de cuadrados de $X_{(q)}$ y la suma de cuadrados de los elementos de X .

Dado que la suma de cuadrados de X es igual a la traza de $(X'X)$, y como:

$$tr(X'X) = \sum_{k=1}^r \lambda_k = \sum_{k=1}^r \alpha_k^2$$

Donde α_k , son los q mayores valores singulares asociados a los λ_k valores propios de la matriz $X'X$ ($\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_q$).

Por lo que la **calidad de representación global** será igual a:

$$CRG = \left(\frac{\sum_{k=1}^q \alpha_k^2}{\sum_{k=1}^r \alpha_k^2} \right) \times 100$$

Si trabajamos con la matriz de varianzas y covarianzas:

$$S = \left[\frac{1}{(I-1)} \right] X'X, \text{ sustituyendo } X=UDV'$$

Entonces: $S = VDU'UDV$, y considerando la métrica utilizada: $A'A=U'U=I$

Se tiene que $S = VD^2V'$

De donde la **calidad de representación de las columnas** será:

$$CRC = \left(\frac{\sum_{k=1}^q \alpha_k^4}{\sum_{k=1}^r \alpha_k^4} \right) \times 100$$

Por lo tanto, la calidad de representación de las columnas de la matriz X es óptima.

La calidad de representación de las filas es muy baja.

1.2.4.2 Calidad de representación del JK-Biplot

La calidad de representación para las filas es óptima, en cambio para las columnas es baja.

1.2.4.3 Calidad de representación del HJ-Biplot

Respecto a la calidad de representación este Biplot posee una importante ventaja respecto a los anteriores ya que las filas y columnas tienen ambas una calidad de representación óptima.

1.2.5 Representaciones Biplot: Interpretación de los resultados

Las Métodos Biplot, como todo método del Análisis Multivariante, presentan los resultados del análisis en un subespacio que, generalmente, es de dimensión dos, aunque la configuración inicial sea de dimensión mayor. Por ello en primer lugar es importante elegir correctamente el número de ejes a retener para obtener una representación adecuada en dimensión reducida.

Suponiendo que hemos seleccionado correctamente el número de ejes para poder explicar en forma adecuada los datos originales es importante realizar una buena interpretación de los resultados. Para ello es importante tener en cuenta las propiedades de los marcadores, la interpretación geométrica y ciertos índices o medidas de bondad de aproximación.

1.2.5.1 Índices de ayuda a la interpretación

La calidad global de representación de los Biplot que se definió en el apartado 1.2.5 no resulta suficiente para evaluar los ajustes individuales de cada fila y cada columna, por lo que es importante a efectos de la interpretación, considerar las contribuciones del factor al elemento fila y al elemento columna (calidades de representación), que permiten determinar la parte de variabilidad de cada elemento explicada por el factor.

La interpretación de esas contribuciones se hace en forma similar al Análisis de Correspondencias, en términos de inercia explicada o varianza absorbida (GALINDO y CUADRAS, 1996; GREENACRE, 1984).

El índice o medida que mide la relación entre los ejes de la representación Biplot y cada una de las variables observadas es lo que se denomina Contribución Relativa del Factor al Elemento (variable) y se lo simboliza CRFE.

La CRFE mide la parte de variabilidad de cada una de las variables explicada por el factor.

Si todos los elementos tienen peso (masa) unitario, la inercia de una variable (elemento columna) con respecto al centro de gravedad, es igual al cuadrado de la distancia de dicho punto al centro de gravedad.

Entonces, la Contribución del Factor al Elemento columna (variable) es:

$$CFE_{Col} = \frac{b_{jk}^2}{\sum_{k=1}^q b_{jk}^2}$$

La CRFE permite conocer las variables que son las responsables de la ordenación de los individuos sobre las proyecciones en cada uno de los ejes Biplot.

Si los datos están centrados, el centro de gravedad θ coincide con el origen, luego:

$$d^2(b_j, \theta) = \sum_{k=1}^q b_{jk}^2$$

En este caso, la CFE coincide con el cuadrado del coeficiente de correlación, entre la variable j-ésima y el eje k-ésimo.

La contribución del Factor al Elemento fila o Calidad de Representación (CLR) es:

$$CFE_{fila} = \frac{a_{ik}^2}{\sum_{k=1}^q a_{ik}^2}$$

1.2.5.2 Interpretación geométrica

En las representaciones Biplot, las filas de la matriz de marcadores filas y las columnas de la matriz de marcadores columna, son coordenadas de puntos en un espacio euclídeo referido a los mismos ejes ortogonales. Por lo tanto la interpretación geométrica está basada en conceptos geométricos simples, más específicamente en la proyección sobre una representación plana, de las características principales de las filas y columnas de una cierta matriz X ($I \times J$).

Los marcadores columnas, b_j se representan como vectores. Los marcadores a_i se representan como puntos. Esta forma de representar ambos marcadores facilita la proyección de los marcadores a_i (filas) sobre el marcador columna b_j .

La dirección de los vectores b_j representa la dirección en la que aumentan los valores de la variable correspondiente, y las proyecciones de todos los puntos fila a_i sobre un vector columna en particular, reproducen aproximadamente los elementos de la columna j -ésima en la matriz original, permitiendo al mismo tiempo una ordenación aproximada de los individuos (filas) respecto a esa variable en particular.

Definida entonces la forma de representación de marcadores filas y columnas, interpretaremos:

1. La distancia entre individuos como disimilaridades entre los mismos, esto es a una menor distancia entre marcadores fila sobre el Biplot le corresponderá una menor disimilaridad entre los individuos; especialmente si los individuos están bien representados.
2. En el GH-Biplot interpretaremos la variabilidad y correlación de las variables, a través de las longitudes y ángulos de los vectores que las representan. En el JK-Biplot no podemos hacer este tipo de interpretación para las variables, aunque las coordenadas nos darán una idea aproximada de cuál es la relación con los ejes.
3. Las relaciones entre individuos y variables en términos de producto escalar, o sea a través de las proyecciones de los puntos individuos sobre los vectores que representan a las variables.
4. La ordenación de los individuos respecto a una variable en particular, a través del orden encontrado en las proyecciones de los marcadores fila sobre el eje determinado por el marcador columna correspondiente a esa variable.

1.3 PROGRAMA GGEbiplotR

El programa permite seleccionar desde la pantalla de selección de modelo qué Biplot se quiere realizar (Figura 1.1). Por defecto la opción seleccionada es GH-Biplot.

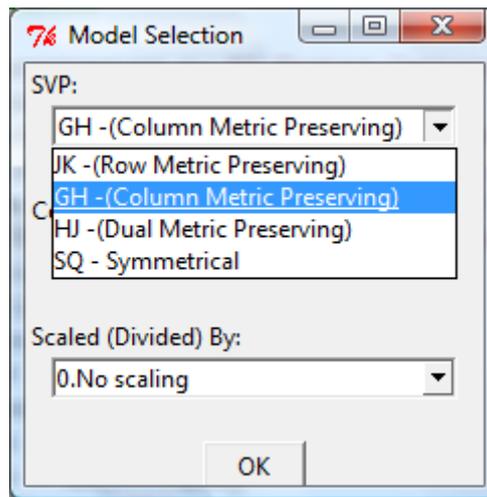


Figura 1.1. Pantalla de selección de modelo

Posteriormente se puede modificar la opción seleccionada por medio del menú (Figura 1.2).

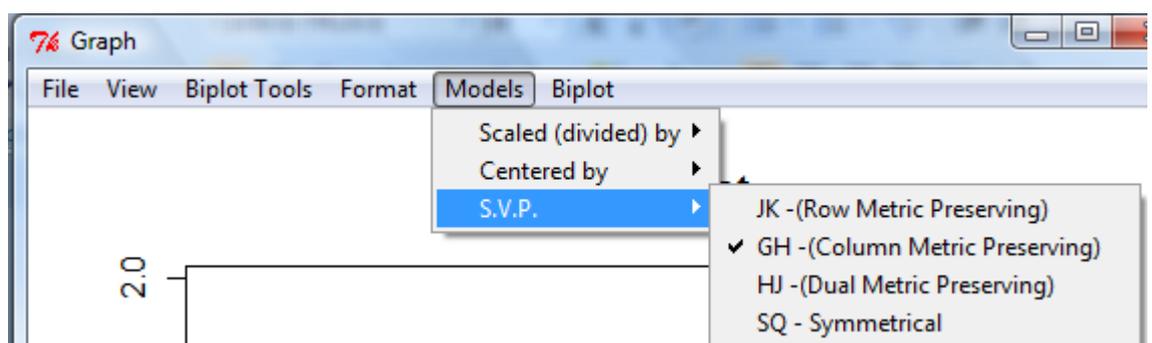


Figura 1.2. Menú de selección de modelo

Por medio de la opción de menú File/ Open log file se accede al fichero de log del Biplot seleccionado (Figura 1.3).

```
R Information
GGE BIPLLOT
Centered by: 2.Tester-Centered G+GE
Scaled (Divided) by: 0.No scaling
SVP: GH -(Column Metric Preserving)

Eigenvalues and variance explained"x"
"1"      5,011
"2"      2,857
"3"      2,063
"4"      1,307
"5"      1,102
"6"      0,982
"7"      0,778
"8"      0,728
"9"      0,283

Row coordinates:
"AXIS1" "AXIS2"
"ann"   0,153  0,648
"ari"   -0,185  0,324
"aug"   -0,21   0,603
"cas"   -0,477 -0,462
"del"   -0,359  0,342
"dia"   -0,35   0,109
"ena"   0,664  0,744
```

Figura 1.3. Fichero de log

CAPÍTULO II: MODELOS CON EFECTO INTERACCIÓN MULTIPLICATIVO (AMMI).

2.1. INTRODUCCIÓN

Los modelos con término multiplicativo han sido muy utilizados para describir la interacción en tablas de dos vías ya que tienen la ventaja de permitir una representación Biplot (simultanea) de las filas y columnas de la tabla y por tanto permiten identificar qué combinaciones de niveles causan la interacción. Estos modelos se clasifican en internos y externos (VAN EEUWIJK y KROONENBERG 1998): internos cuando estiman la interacción haciendo uso solamente de la información contenida en la tabla inicial de datos; externos cuando además utilizan información proveniente de variables externas, ya sean medidas sobre filas, columnas o ambos factores de variación.

Algunos de los modelos con término multiplicativo que aparecen en la literatura son los siguientes: **Modelo concurrente de Tukey** (TUKEY 1949), **Modelos de Regresión sobre la media** (YATES y COCHRAN 1938, MANDEL 1961, FINLAY y WILKINSON 1963, EBERHART y RUSSELL 1966). Todos ellos se caracterizan por tratar de explicar la interacción a partir de un solo término multiplicativo, lo cual en muchos casos resulta insuficiente.

En los **Modelos AMMI** (GAUCH 1988) se mantiene la descomposición en términos multiplicativos de la interacción y esta no es forzada a tener una característica específica: se incluyen tantos términos como sean necesarios para explicar la variabilidad asociada a la interacción. Estos modelos combinan las técnicas de Análisis de la Varianza y Análisis de Componentes Principales (GAUCH y ZOBEL (1989)) y tienen como objetivo explicar la interacción asociada a un ANOVA bifactorial, a partir de una representación BILOT.

Los modelos AMMI han sido aplicados fundamentalmente en experimentos de campo, más específicamente en el análisis de la interacción Genotipo-Ambiente, con el objetivo de clasificar genotipos en estables e inestables a partir de su interacción con el ambiente (VAN EEUWIJK 1995 a y b, KANG y GAUCH 1996, ROMAGOSA et al 1996).

Una variedad o genotipo se introduce en la producción cuando además de tener altos rendimientos, presenta un grado de estabilidad aceptable, es decir, reacciona favorablemente en diferentes condiciones ambientales.

Una vez obtenidos los datos experimentales, el análisis estadístico más utilizado comúnmente es el Análisis de la Varianza para un arreglo bifactorial en el que los factores considerados son el genotipo y el ambiente. A partir del análisis de la varianza podemos detectar la presencia o no de interacción mediante la F de Snedecor correspondiente a esta fuente de variación. Sin embargo, una vez detectada, no la interpretamos, nos limitamos a seleccionar los genotipos con mayores valores medios, sin tener en cuenta su grado de estabilidad.

Los modelos AMMI al permitir una representación BIPLLOT de filas (genotipos) y columnas (ambientes) dan la posibilidad de estudiar el grado de estabilidad de los genotipos al ser probados en diferentes ambientes.

Existen otras formas de hacer referencia a este tipo de modelos con término multiplicativo en la interacción y efectos principales aditivos (modelos AMMI); así por ejemplo, GABRIEL (1978) y DENIS (1991) lo denominan modelos bilineales; por otra parte, DENIS y GOWER (1992, 1994) los llaman modelos biaditivos.

2.2. FUNDAMENTO TEÓRICO

El modelo lineal al que se ajustan los datos experimentales en un arreglo bifactorial bajo un Diseño Completamente Aleatorizado es el siguiente:

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad i = 1..I \quad j = 1..J \quad k = 1..K \quad (2.1)$$

Siendo:

I: n° de niveles del primer factor.

J: n° de niveles del segundo factor.

K: n° de observaciones por tratamiento o combinaciones de niveles.

y_{ijk} : K ésima observación correspondiente a la combinación de niveles ij.

α_i, β_j : Efectos principales para filas y columnas respectivamente.

$(\alpha\beta)_{ij}$: Efecto interacción.

El estimador mínimo cuadrático correspondiente a la interacción se calcula de la forma:

$$(\widehat{\alpha\beta})_{ij} = y_{ij.} - y_{i..} - y_{.j.} + y_{...} \quad (2.2)$$

En la literatura se presentan modelos en los que se trata de explicar la interacción a partir de términos multiplicativos. En dichos términos aparecen como factores los efectos principales asociados a filas, columnas o ambas fuentes de variación. Sin embargo, en muchos casos resulta imposible modelar la interacción a partir de estos términos por la complejidad de la misma. Estos modelos han sido desarrollados para el caso de una observación por celda, es decir no tenemos repeticiones que permitan controlar la variabilidad dentro de la celda, y por tanto no es posible estimar el error experimental.

TUKEY (1949) fue el primero en proponer un modelo para el análisis de la interacción en experimentos de dos vías para diseños no replicados.

El modelo propuesto es de la forma:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j + e_{ij} \quad (2.3)$$

Donde $\alpha_i\beta_j$ es el producto de los efectos principales y λ un coeficiente de regresión. Contrastar la hipótesis $\lambda = 0$ será equivalente a un test de hipótesis para contrastar que el producto de los efectos no contribuye a la predicción de Y_{ij} . Por tanto su finalidad es distinguir entre el modelo aditivos que sólo incluye los efectos principales y el modelo que contempla un término de interacción.

Otro tipo de modelos con término multiplicativo muy utilizado para describir la interacción en tablas de dos vías son los Modelos de regresión sobre la media (YATES y COCHRAN 1938, MANDEL 1961), los cuales se definen de la siguiente forma:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \lambda\gamma_i\beta_j \quad (2.4)$$

$$Y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\gamma_j \quad (2.5)$$

En (2.4): $(\widehat{\alpha_i\beta_j})_{ij} = \lambda\gamma_i\beta_j$ y el modelo se denomina Regresión por filas.

En (2.5): $(\widehat{\alpha_i\beta_j})_{ij} = \lambda\alpha_i\gamma_j$ y el modelo se denomina Regresión por columnas.

Como su propio nombre indica se estiman los valores de γ_i a partir de las regresiones de los efectos principales en la variable dependiente.

En el contexto del Análisis de Interacción Genotipo-Ambiente, los modelos de Mandel son también conocidos como **Modelos de Finlay y Wilkinson** (FINLAY y WILKINSON 1963).

En todos los casos anteriores se fuerza a que la interacción tenga características muy específicas, por tanto puede darse el caso de que la descomposición en términos multiplicativos no sea suficiente para explicar la variabilidad asociada a la interacción.

En los modelos AMMI (GOLLOB 1968) se combinan las técnicas de Análisis de la Varianza y Análisis de Componentes Principales. Se introducen en el modelo tantos

términos multiplicativos como sean necesarios para explicar la variabilidad de la interacción.

El método consiste en hacer la descomposición en valores singulares de la matriz Z de orden I x J formada por los estimadores de las interacciones:

$$Z = (z_{ij}) = (\widehat{\alpha\beta})_{ij} = y_{ij} - y_{i..} - y_{.j.} + y_{...} \quad (2.6)$$

Al hacer la descomposición en valores singulares de Z tenemos:

$$z_{ij} = (\widehat{\alpha\beta})_{ij} = \sum_{m=1}^M \lambda_m u_{mi} v_{mj} \quad (2.7)$$

Siendo M el rango de Z.

Llamaremos modelo AMMI de orden M a la expresión:

$$AMMI_M: E(y_{ijk}) = \mu + \alpha_i + \beta_j + \sum_{m=1}^M \lambda_m u_{mi} v_{mj} \quad (2.8)$$

Donde:

λ_m corresponde al valor singular de orden m de $Z'Z$

u_{mi} : coordenada i-ésima del vector singular de ZZ' asociado a λ_m

v_{mj} : coordenada j-ésima del vector singular de $Z'Z$ asociado a λ_m

De esta forma podemos representar las filas (genotipos) y columnas (ambientes) en un subespacio de dimensión M en el que las proximidades entre genotipos van a indicar que interactúan de manera similar con el ambiente. Así los genotipos que se ubican cerca del origen de coordenadas serán los más estables, es decir, los que interactúan poco con el ambiente. Por otra parte aquellos genotipos que se alejan del origen serán los más inestables; tendrán altos rendimientos solamente en aquellos ambientes próximos a ellos en la representación.

Al estar en presencia de una técnica en la que se realiza una descomposición en valores singulares, se plantea el problema de cuántos ejes elegir.

Si consideramos un solo factor el modelo será:

$$AMMI_1: E(y_{ijk}) = \mu + \alpha_i + \beta_j + \lambda_1 \mu_{1i} v_{1j} \quad (2.9)$$

Con dos factores:

$$AMMI_2: E(y_{ijk}) = \mu + \alpha_i + \beta_j + \lambda_1 \mu_{1i} v_{1j} + \lambda_2 \mu_{2i} v_{2j} \quad (2.10)$$

Y así sucesivamente. (GAUCH y ZOBEL 1989, MILLIKEN y JOHNSON 1989).

Para saber qué modelo es el más adecuado y por tanto cuántos términos multiplicativos deben ser incluidos en el modelo se realiza la descomposición de la suma de cuadrados de la interacción asociada al análisis de varianza para un modelo bifactorial.

Para conocer si un modelo con G términos multiplicativos es válido, se realiza el ANOVA, utilizando como variabilidad total la correspondiente a la interacción, se determina el residual de la interacción asociada a la parte de la variabilidad total (interacción) que no es explicada por los G términos multiplicativos, y se calcula la F de Snedecor correspondiente (Tabla 2.1.).

F.V	G.L	S.C	C.M	F
INT GxE	(I-1)(J-1)	$K \sum \lambda_i^2$		
AMMI ₁	(I-1)+(J-1)-1	$K \lambda_1^2$		
AMMI ₂	(I-1)+(J-1)-3	$K \lambda_2^2$		
		
AMMI _L	(I-1)+(J-1)-2L-1	$K \lambda_L^2$	S.C/G.L	C.M. AMMI _L /C.M.Res
Residual	Por diferencia	Por Dif	C.M.Res	

Tabla 2.1. ANOVA para la selección del número de términos multiplicativos

Si la F de Snedecor es significativa, se incluye el término multiplicativo asociado a λ_L , y se pasa a analizar el término L+1. En caso contrario se elimina y no se continua el análisis debido a la relación de orden que existe entre los valores singulares.

2.3. PROGRAMA GGEBiplotR

Para realizar un análisis AMMI utilizando el programa GGEBiplotR, basta con tomar como opción de centrado Double-Centered GE en la pantalla de selección de modelo (Figura 2.1) (por defecto aparece Tester-Centered G+GE), o bien posteriormente a través de la opción de menú correspondiente (Figura 2.2).

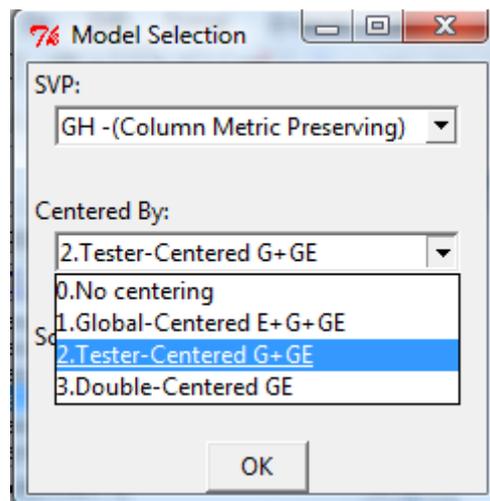


Figura 2.1. Pantalla de selección de modelo

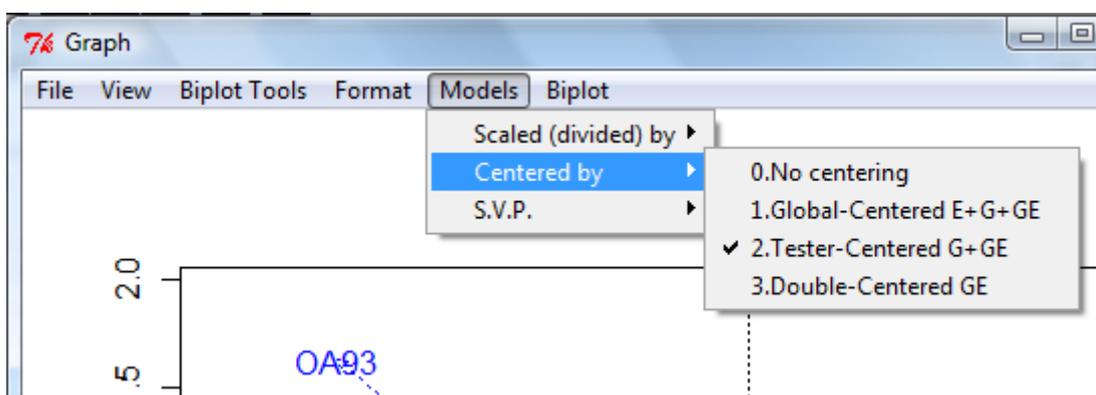


Figura 2.2. Menu de selección de modelo

CAPÍTULO III: GGE BILOT

3.1.INTRODUCCIÓN

En una tabla de dos vías genotipo-ambiente Y, el valor de cada celda se puede descomponer como un efecto mixto obtenido a partir de la media global (μ) modificado por el efecto principal genotipo (fila) (α_i), el efecto principal ambiente (columna) (β_j), la interacción genotipo ambiente (θ_{ij}) más un error aleatorio (ε_{ij}):

$$y_{ij} = \mu + \alpha_i + \beta_j + \theta_{ij} + \varepsilon_{ij} \quad (3.1)$$

La matriz P de la que se va a realizar la descomposición en valores singulares puede ser cualquier parte de Y, resultando por lo tanto diferentes modelos:

$$P_{ij} = y_{ij} = \mu + \alpha_i + \beta_j + \theta_{ij} \quad (3.2)$$

$$P_{ij} = y_{ij} - \mu = \alpha_i + \beta_j + \theta_{ij} \quad (3.3)$$

$$P_{ij} = y_{ij} - \mu - \alpha_i = \beta_j + \theta_{ij} \quad (3.4)$$

$$P_{ij} = y_{ij} - \mu - \beta_j = \alpha_i + \theta_{ij} \quad (3.5)$$

$$P_{ij} = y_{ij} - \mu - \alpha_i - \beta_j = \theta_{ij} \quad (3.6)$$

Obviamente los biplots basados en los diferentes modelos (Ecuaciones 3.2 - 3.6) tienen diferentes interpretaciones. Todos los modelos son útiles y se utilizarán unos u otros dependiendo de las preguntas a las que se quiera dar respuesta. Si sólo se está interesado en GE se debería elegir la ecuación (3.6). Este modelo también es apropiado para el estudio visual de datos de microarrays (PITTELKOW y WILSON 2003 WOUTERS et al. 2003). Si lo que nos interesa es la evaluación de genotipos la Ecuación (3.5) es la más apropiada. Los Biplots basados en esta ecuación se llaman GGE Biplots (YAN et al. 2000). Si se está interesado en el estudio de los datos completos, la Ecuación (3.2) debería ser la elección.

3.2. CONCEPTO DE GGE Y GGE BILOT

En biología la variación en el fenotipo (P) consiste en la variación del ambiente (E), del genotipo (G) y de la interacción genotipo-ambiente (GEI). Podemos escribir por tanto:

$$P = E + G + GE \quad (3.7)$$

E además puede ser particionado en año (Y), localización (L) e interacción año-localización, y GE puede descomponerse en interacción genotipo-año (GY), interacción genotipo-localización (GL) e interacción genotipo-localización-año (GLY) (COMSTOCK y MOLL 1963). En ensayos multi-ambiente para un solo año (MET) no tendrán que estimarse GY, LY y GLY y por tanto E está compuesto solo de L y GE está compuesto solo de GL.

Independientemente de si los datos son de un solo año o de varios años un fenómeno universal en todos los ensayos de rendimiento regional es que E es siempre la principal fuente de variación del rendimiento y G y GE son relativamente pequeños (GAUCH y ZOBEL 1996).

Sin embargo el efecto principal del ambiente no es relevante en la evaluación de cultivos, sólo G y GE son relevantes. Por lo tanto para la evaluación de cultivos es esencial eliminar E de los datos y centrarnos en G y GE. Esta idea la describen GAUCH y ZOBEL (1996), ellos comparan la evaluación de cultivos basada en datos MET con el hecho de sacar una fotografía de un gato sentado en una ventana muy grande. Si nuestro propósito no es mostrar lo pequeño que es el gato en relación a la ventana, es lógico hacer zoom en el gato en lugar de incluir el gato sentado en la ventana. Otro punto importante en la evaluación de cultivos es que G y GE deben considerarse simultáneamente para poder tomar decisiones.

Resumiendo:

- E es grande pero irrelevante en la evaluación de cultivos, por lo tanto debe de ser sustraído de los datos.
- Sólo G y GE son significativos en la evaluación de cultivos, y se deben considerar simultáneamente para poder tomar decisiones.

Esto se puede resumir en la siguiente fórmula:

$$P - E = G + GE \quad (3.8)$$

El término GGE es la contracción de G+GE. Un Biplot que represente el GGE de un conjunto de datos MET se llama GGE Biplot.

Las ecuaciones anteriores se pueden escribir como:

$$\hat{Y}_{ij} = \mu + \alpha_i + \beta_j + \theta_{ij} \quad (3.9)$$

o como:

$$\hat{Y}_{ij} - \mu - \beta_j = \alpha_i + \theta_{ij} \quad (3.10)$$

donde:

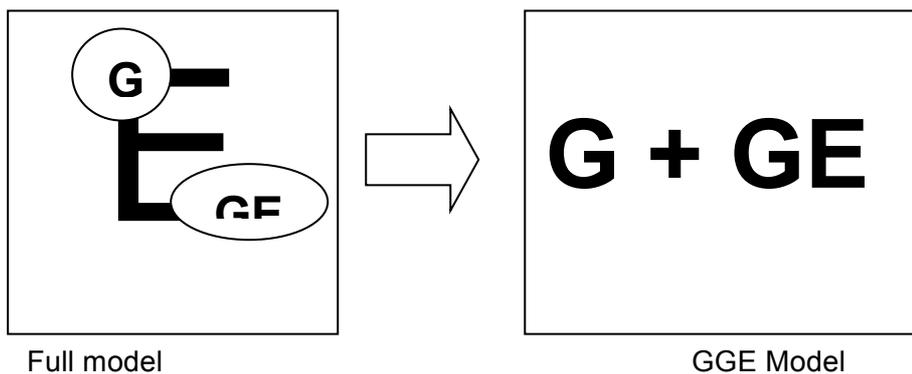
\hat{Y}_{ij} = rendimiento esperado del genotipo i en el ambiente j.

μ = media global de todas las observaciones.

α_i = efecto principal del genotipo i.

β_j = efecto principal del ambiente j.

θ_{ij} = interacción entre el genotipo i y el ambiente j.



En lugar de intentar separar G y GE, el modelo del GGE biplot mantiene G y GE juntos y descompone la mezcla GGE en dos términos multiplicativos:

$$\hat{Y}_{ij} - \mu - \beta_j = g_{i1}e_{1j} + g_{i2}e_{2j} + \varepsilon_{ij} \quad (3.11)$$

donde g_{i1} y e_{1j} son las coordenadas principales para el genotipo i y el ambiente j , respectivamente; g_{i2} y e_{2j} las coordenadas secundarias para el genotipo i y el ambiente j , respectivamente; y ε_{ij} son los residuos no explicados por los efectos primarios y secundarios. Las coordenadas anteriores se pueden obtener a partir de la descomposición en valores singulares (SVD) de la matriz GGE o a haciendo regresión de la matriz GGE frente a los efectos principales.

La forma más habitual de obtener la ecuación 3.11 es mediante la descomposición en valores singulares (SVD) de la matriz GGE:

$$\hat{Y}_{ij} - \mu - \beta_j = \lambda_1 \xi_{i1} \eta_{1j} + \lambda_2 \xi_{i2} \eta_{2j} + \varepsilon_{ij} \quad (3.12)$$

Donde λ_1 y λ_2 son los valores propios de la primera y segunda componente principal, PC1 y PC2 respectivamente; ξ_{i1} y ξ_{i2} son los vectores propios del genotipo i para PC1 y PC2 respectivamente; y η_{1j} y η_{2j} son los vectores propios del ambiente j para PC1 y PC2 respectivamente.

Para obtener la representación BILOT tenemos que repartir los valores propios anteriores entre los vectores propios del genotipo y del ambiente. Para ello utilizaremos la siguiente fórmula:

$$g_{il} = \lambda_l^{f_l} \xi_{il} \text{ y } e_{ij} = \lambda_l^{1-f_l} \eta_{lj} \quad (3.13)$$

con l igual a 1 ó 2 y f el factor de partición. Teóricamente el factor f_l puede ser cualquier valor entre 0 y 1 dando lugar a las diferentes factorizaciones Biplot.

3.3.ANALISIS BILOT DE MET

La interacción genotipo ambiente (GEI) es una cuestión que no se puede evitar en el análisis de datos MET. La interacción no es el único factor que se debe tener en cuenta en la evaluación de cultivos, el otro factor que puede llegar a ser más importante aún es el genotipo. El diagrama de flujo (Figura 3.1) modificado por YAN y HUNT (1998) resume las tareas que se deben realizar en un análisis MET. La primera pregunta es: “¿Hay interacción?”. Si no hay interacción cualquier ambiente debería ser suficiente para seleccionar las mejores variedades. Si hay interacción la siguiente pregunta es: “¿Existe interacción de cruce?” El concepto de interacción de cruce (BAKER 1990) es importante porque implica que G y GEI se deben de considerar conjuntamente en la evaluación de cultivos. Este concepto ha conducido a estudios para identificar grupos homogéneos de ambientes con interacción de cruce insignificante (CROSSA y CORNELIUS, 1997).

Si no existe interacción de cruce, es decir, un genotipo gana en todos los ambientes, cualquier ambiente debería ser suficiente para seleccionar la mejor variedad, sin embargo existirá un ambiente que será el mejor. Si la interacción de cruce es significativa nos debemos hacer una tercera pregunta: ¿Se repite el patrón “which-won-where” a lo largo de los años? Esta es una cuestión importante para determinar si el ambiente se debe dividir en diferentes mega-ambientes (YAN y HUNT 1998). Si la respuesta es no, estamos ante un complejo mega ambiente con interacción de cruce que no se repite (YAN y RAJCAN 2002). Si este es el caso, la evaluación de cultivos se debe basar en el rendimiento promedio y la estabilidad (KANG 1993,1998). Si el patrón “wich-won-where” se repite en los distintos años, surgen mega-ambientes y la evaluación de cultivos se debe realizar en cada uno de ellos.

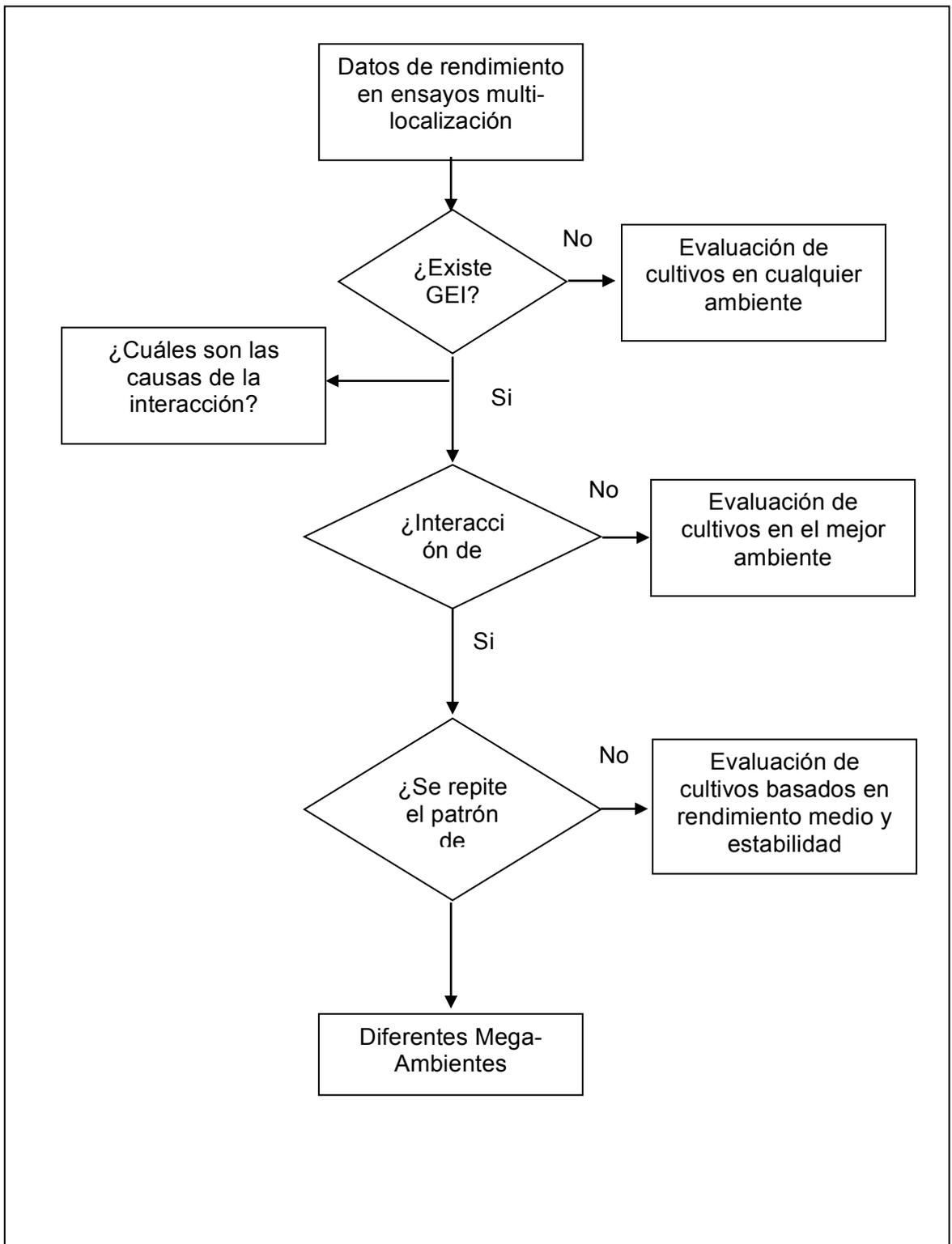


Figura 3.1. Análisis de datos MET

Como ejemplo para demostrar la utilidad del GGE Biplot se utilizan los datos de los ensayos de rendimiento del trigo de invierno en Ontario 1993. El conjunto de datos consta de 18 cultivos en 9 localidades, con 4 bloques o repeticiones en cada lugar. La tabla 3.1 muestra los valores medios de cada combinación genotipo-localización.

En la figura 3.2 se presenta el GGE Biplot de los datos anteriores obtenido con el programa GGEbiplotR donde los genotipos aparecen en verde y los ambientes en azul. Para la construcción del GGE Biplot se debe seleccionar la opción Tester-Centered G+GE en la pantalla de selección de modelo o en el menú correspondiente. El modelo utilizado explica un 78% de la variabilidad total (PC1 = 58,9% y PC2 = 19,15%).

Names	BH93	EA93	HW93	ID93	KE93	NN93	OA93	RN93	WP93
Ann	4,46	4,15	2,849	3,084	5,94	4,45	4,351	4,039	2,672
Ari	4,417	4,771	2,912	3,506	5,699	5,152	4,956	4,386	2,938
Aug	4,669	4,578	3,098	3,46	6,07	5,025	4,73	3,9	2,621
Cas	4,732	4,745	3,375	3,904	6,224	5,34	4,226	4,893	3,451
Del	4,39	4,603	3,511	3,848	5,773	5,421	5,147	4,098	2,832
Dia	5,178	4,475	2,99	3,774	6,583	5,045	3,985	4,271	2,776
Ena	3,375	4,175	2,741	3,157	5,342	4,267	4,162	4,063	2,032
Fun	4,852	4,664	4,425	3,952	5,536	5,832	4,168	5,06	3,574
Ham	5,038	4,741	3,508	3,437	5,96	4,859	4,977	4,514	2,859
Har	5,195	4,662	3,596	3,759	5,937	5,345	3,895	4,45	3,3
Kar	4,293	4,53	2,76	3,422	6,142	5,25	4,856	4,137	3,149
Kat	3,151	3,04	2,388	2,35	4,229	4,257	3,384	4,071	2,103
Luc	4,104	3,878	2,302	3,718	4,555	5,149	2,596	4,956	2,886
m12	3,34	3,854	2,419	2,783	4,629	5,09	3,281	3,918	2,561
Reb	4,375	4,701	3,655	3,592	6,189	5,141	3,933	4,208	2,925
Ron	4,94	4,698	2,95	3,898	6,063	5,326	4,302	4,299	3,031
Rub	3,786	4,969	3,379	3,353	4,774	5,304	4,322	4,858	3,382
Zav	4,238	4,654	3,607	3,914	6,641	4,83	5,014	4,363	3,111

Tabla 3.1. Rendimiento de 18 cultivos testados en 9 ambientes

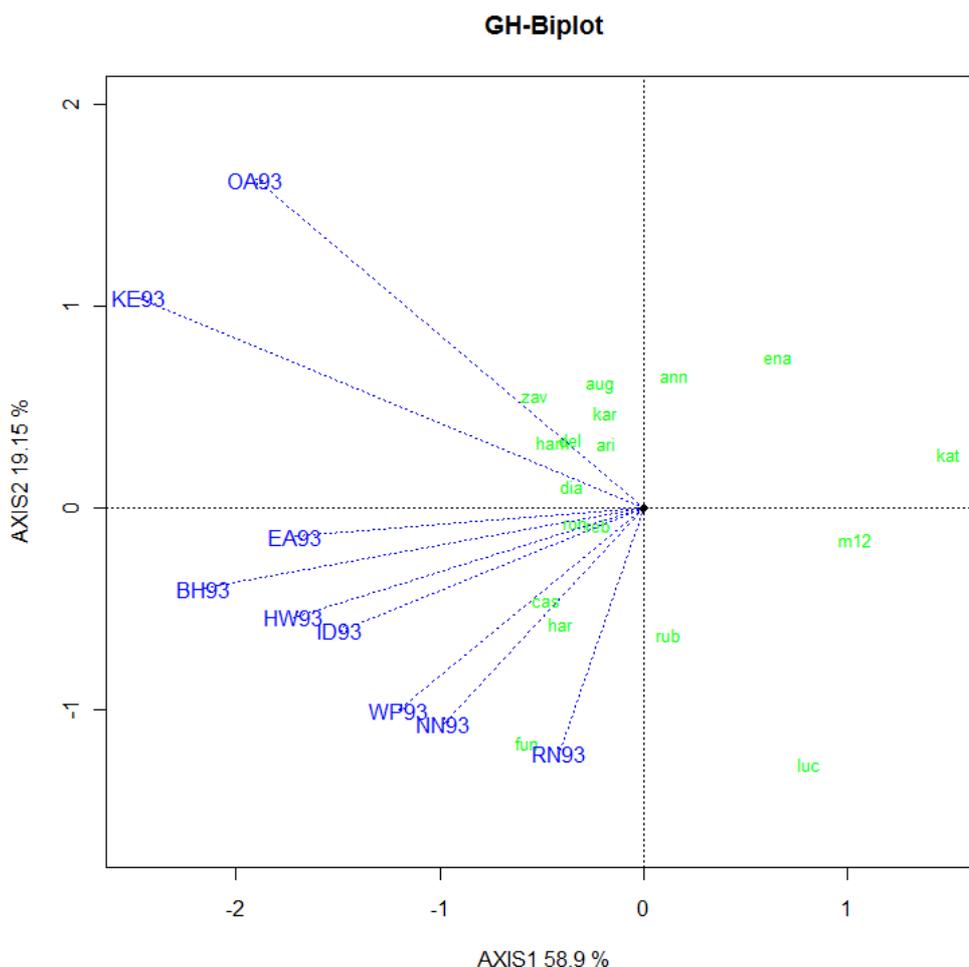


Figura 3.2. GGE Biplot basado en los datos de rendimiento del trigo de invierno (Ontario, 1993)

3.3.1. Comparaciones simples utilizando el GGE Biplot

En la figura 3.2 aparecen representados los genotipos y los ambientes, por lo que es posible estudiar las posibles interacciones entre un genotipo y un ambiente (es decir, el rendimiento de cada genotipo en cada ambiente). La regla que se utiliza es la siguiente: el rendimiento de un genotipo en un ambiente es mayor que la media si el ángulo entre el vector genotipo y el vector ambiente es menor de 90° , es peor que la media si el ángulo es mayor de 90° y es próximo a la media si el ángulo es aproximadamente de 90° . Por ejemplo el rendimiento de "kat" está por debajo de la media en todos los ambientes (ángulos obtusos), el rendimiento de "fun" está por encima de la media en todos los ambientes (ángulos agudos) salvo en OA93. Por lo tanto el ángulo determina la dirección de la interacción (es decir, si el rendimiento está

por debajo o por encima de la media). La magnitud de la interacción está determinada por el coseno del ángulo y la longitud del vector (por la propiedad del producto interno del Biplot). Por lo tanto podemos utilizar esta información para ordenar los genotipos en un ambiente basándonos en su rendimiento y ordenar los ambientes en función del rendimiento de un genotipo.

3.3.1.1. Rendimiento de los distintos cultivos en un ambiente dado

Si fuéramos productores de cultivos estaríamos interesados en conocer qué variedades son las que más se adaptan a nuestro ambiente. Podríamos obtener esta información sin más que analizar los datos, pero puede ser más fácil obtener esta información a través del GGE Biplot. Por ejemplo para analizar el rendimiento de los diferentes cultivos en un ambiente determinado, por ejemplo BH93, se traza una recta que pasa por el origen del biplot y por el marcador de BH93, que denominaremos eje BH93. Los genotipos se pueden clasificar de acuerdo a sus proyecciones sobre el eje BH93 en la dirección indicada por la flecha (Figura 3.3). Así en BH93 el cultivo de mayor rendimiento es “fun” y el de menor rendimiento es “kat”. La recta que pasa por el origen del biplot y es perpendicular al eje BH93 divide a los genotipos en dos grupos: lo que tienen rendimiento por encima de la media desde “fun” hasta “rub” y aquellos cuyo rendimiento está por debajo de la media desde “ann” hasta “kat”.

La longitud del vector que une el origen del biplot con el marcador ambiente es una medida de la capacidad del medio ambiente para discriminar entre cultivos. Un vector corto, en relación con el tamaño biplot, implica que todas las variedades tienden a tener similar rendimiento en el medio ambiente asociado. Por lo tanto, las diferencias de cultivos con base en proyecciones sobre los vectores ambiente puede no ser fiable, sino que puede reflejar sólo el ruido.

Este análisis se obtiene mediante la opción de menú Biplot Tools / Examine an Environment del programa GGEBiplotR.

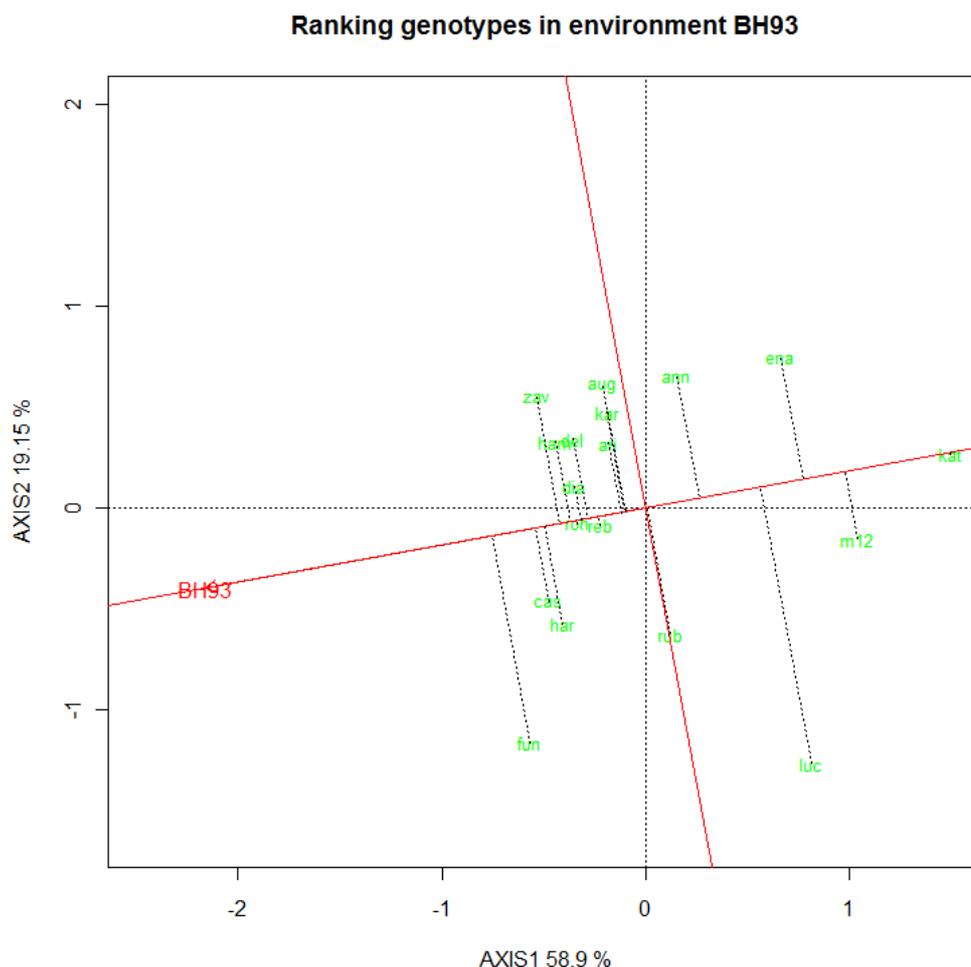


Figura 3.3. Comparación del rendimiento de los diferentes cultivos en un ambiente (BH93).

3.3.1.2. Adaptación de un cultivo a los distintos ambientes

Los productores de cultivos también pueden estar interesados en conocer qué entorno es el más adecuado para un determinado cultivo. La figura 3.4 muestra como se adapta el cultivo “fun” en los diferentes ambientes.

Se dibuja una línea que pasa por el origen del gráfico y por el marcador de “fun” que se denomina eje de “fun”. Los ambientes se ordenan a lo largo del eje “fun” en la dirección indicada por la flecha. Por tanto el rendimiento de “fun” en los diferentes ambientes fue $WP93 \cong NN93 \cong BH93 \cong RN93 \cong ID93 \cong HW93 > EA93 > KE93 > OA93$. La línea que pasa por el origen del biplot y que es perpendicular al eje “fun” separa los ambientes en los cuales el rendimiento de “fun” está por debajo de la

media, es decir OA93 de aquellos en los cuales el rendimiento está por encima de la media, es decir los restantes. La distancia entre el origen del biplot y el marcador de un cultivo se llama vector cultivo. La longitud de este vector es una medida de la capacidad de respuesta del cultivo al medio ambiente. Un vector corto en relación al tamaño del biplot implica que la variedad asociada tiende a ser estable en todos los ambientes.

Este análisis se obtiene mediante la opción de menú Biplot Tools / Examine a Genotype del programa GGEbiplotR.

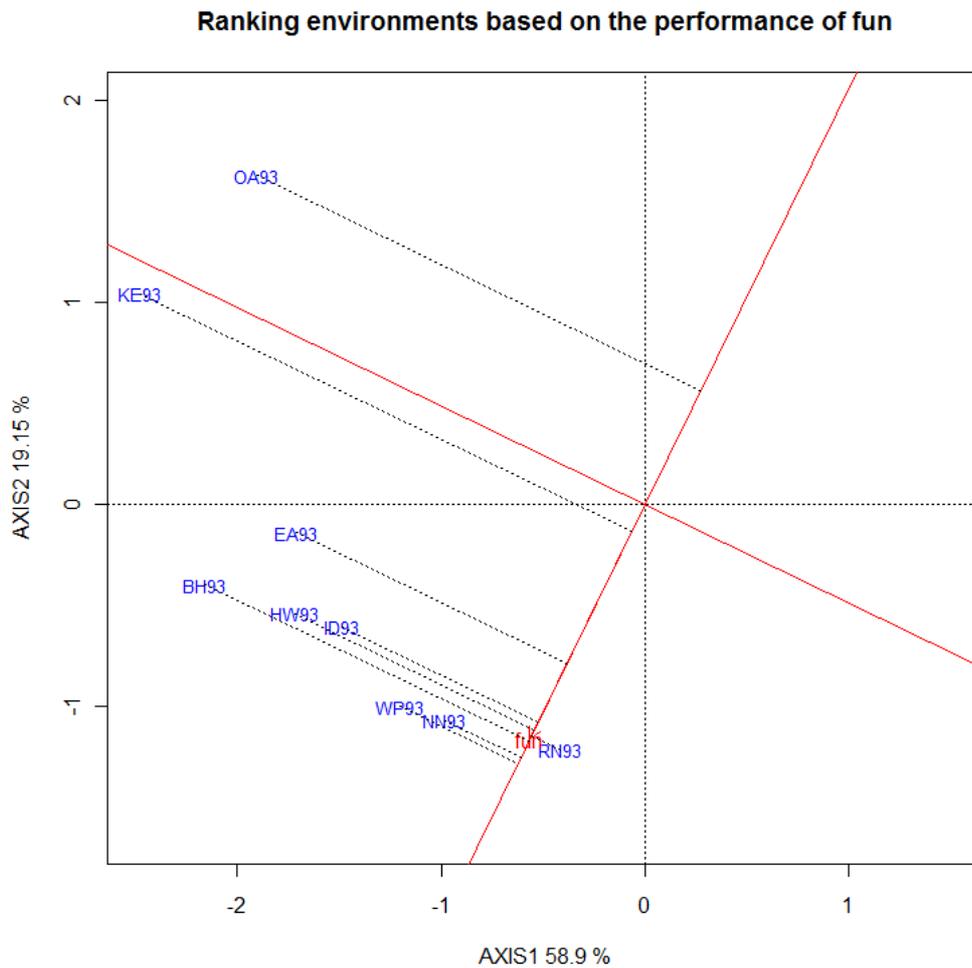


Figura 3.4. Comparación del rendimiento del cultivo “fun” en los diferentes ambientes.

3.3.1.3. Comparación entre dos genotipos

Para comparar dos cultivos, por ejemplo “zav” y “fun” se dibuja una línea que los conecta y una perpendicular a esta que pasa por el origen del biplot.

Existen dos ambientes OA93 y KE93 que están al mismo lado de la perpendicular que “zav” y los otros siete ambientes están al otro lado de la perpendicular, junto con “fun”. Esto indica que el rendimiento de “zav” es mayor que el de “fun” en OA93 y KE93, pero es menor que el de “fun” en los otros siete ambientes.

Este análisis se obtiene mediante la opción de menú Biplot Tools / Compare two Genotypes del programa GGEbiplotR.

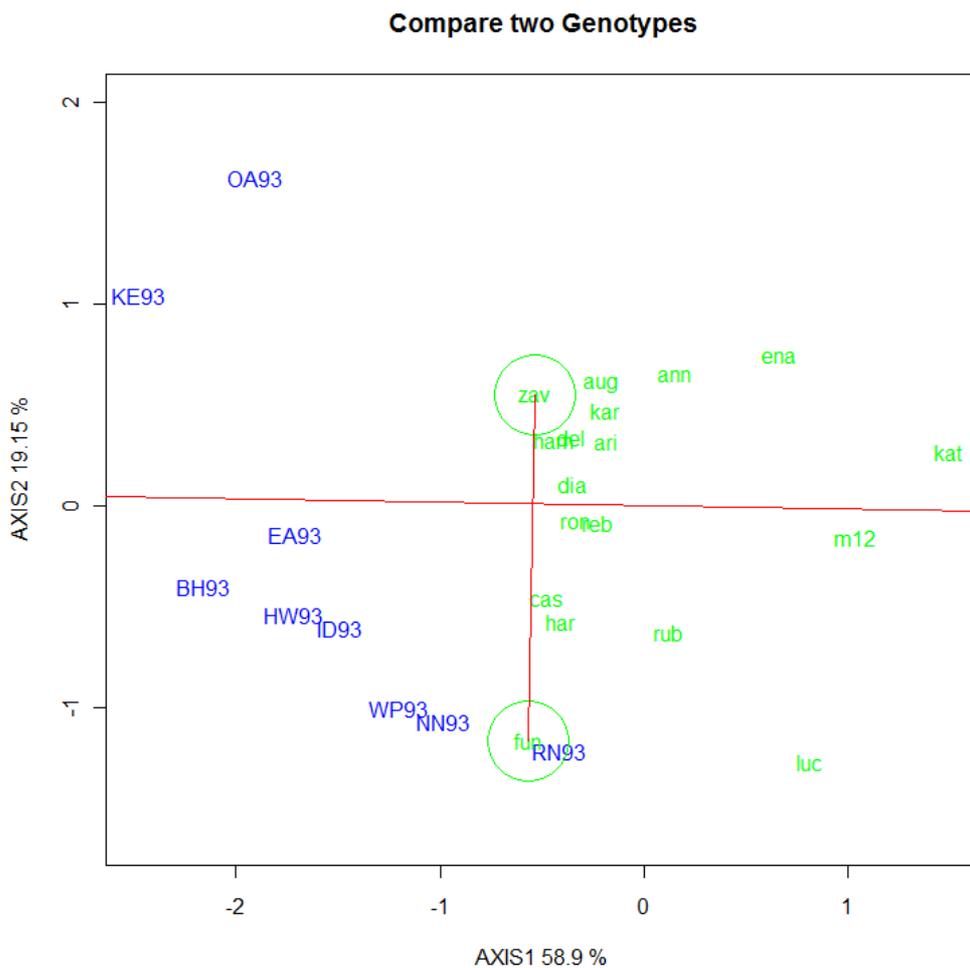


Figura 3.5. Comparación del rendimiento de los cultivos “fun” y “zav” en los distintos entornos de prueba

Esta interpretación, que fue documentada por YAN (1999), se basa en la propiedad del producto interno. Basándonos en esta propiedad los cultivos “fun” y “zav” deberían tener el mismo rendimiento en cualquier ambiente que está localizado sobre la línea perpendicular, porque sus proyecciones sobre la misma son iguales. El cultivo “zav” tiene una proyección más larga sobre los vectores ambiente que están de su lado con respecto a la perpendicular y una proyección más corta sobre los vectores ambiente que están al otro lado respecto a la perpendicular. Una proyección más larga significa mayor rendimiento y viceversa.

3.3.2. Investigación de mega – ambientes

3.3.2.1. Which-won-where

Fue propuesto por YAN (1999). Se construye un polígono uniendo los cultivos (“fun”, “zav”, “ena”, “kat” y “luc”) que se encuentran más alejados del origen del biplot, de manera que todos los cultivos se encuentran dentro del polígono. Este polígono se conoce como “convex hull” y los cultivos situados en los extremos se pueden denominar cultivos vértice. Por tanto los cultivos vértice son aquellos que están más alejados del origen. Estos cultivos son aquellos con vectores más largos en sus respectivas direcciones, lo cual mide su capacidad de respuesta en los ambientes. Por tanto los cultivos situados en los vértices serán aquellos más sensibles.

Se dibuja una línea perpendicular a cada lado del polígono. El polígono se utiliza para comparar cultivos en vértices adyacentes: comparamos “zav” con “ena” y también con “ann” y con “aug”, todos ellos localizados en la línea poligonal que conecta “zav” con “ena”. Se observa que “zav” tiene mayor rendimiento en los ambientes OA93 y KE93. De la comparación de “zav” con “fun” y también con “cas” y “ham” se observa que “zav” tiene mayor rendimiento en OA93 y KE93. A pesar de que no podemos comparar directamente “zav” con “kat”, “kat” es obviamente inferior que “ena” en OA93 y KE93. De igual forma aunque no existe comparación directa entre “zav” y “luc”, “luc” es obviamente inferior a “fun” en OA93 y KE93. Por lo tanto “zav” es el cultivo con mayor rendimiento en OA93 y KE93. Curiosamente OA93 y KE93 están en el mismo sector separado del resto del biplot por dos líneas perpendiculares y “zav” es el vértice del sector. Las perpendiculares a los lados del

polígono dividen al biplot en sectores. Cada sector tiene un vértice cultivo, por lo que podemos denominar al sector con el nombre del cultivo.

Como regla, el cultivo vértice es cultivo con mayor rendimiento en todos los ambientes que pertenecen al sector. Como ejemplo: “fun” es el cultivo con mayor rendimiento en su sector que contiene siete ambientes: “EA93”, “BH93”, “HW93”, “ID93”, “WP93”, “NN93” y “RN93”.

Los cultivos “ena”, “kat” y “luc” no tienen ningún ambiente en sus sectores. Esto indica que estos cultivos no eran los mejores en ninguno de los ambientes de prueba, es más son los más pobres en alguno o todos los ambientes de prueba. Generalizando, los cultivos situados en los vértices son las variedades más sensibles, que son las mejores o las más pobres en algunos o todos los ambientes de prueba.

Este análisis se obtiene mediante la opción de menú Biplot Tools / Which Won Where/ What del programa GGEbiplotR.

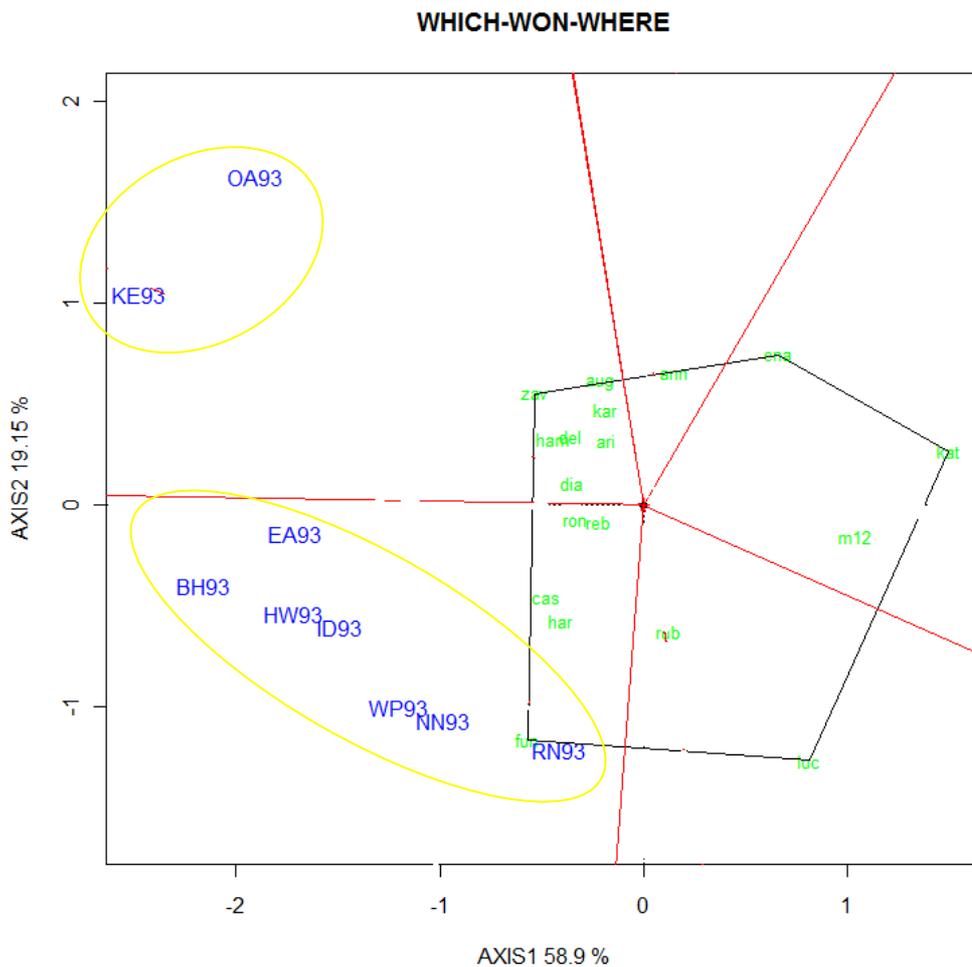


Figura 3.6. ¿Qué genotipo gana dónde?

3.3.2.2. Investigación de mega-ambientes

El polígono no sólo muestra qué cultivo es mejor en cada ambiente de prueba sino que también divide los ambientes de prueba en grupos. En la figura 3.6 se pueden observar claramente dos grupos: OA93 y KE93 pertenecen al sector “zav” y los otros siete ambientes al sector “fun”. Esto sugiere que existen dos mega-ambientes para el trigo de invierno en Ontario.

Son necesarios dos criterios para sugerir la existencia de diferentes mega-ambientes. El primero es que los cultivos con mejor rendimiento sean distintos en los entornos de prueba (GAUCH y ZOBEL 1997). En segundo lugar, la variación entre grupos debe ser significativamente mayor que variación dentro del grupo. Gráficamente, los distintos mega-ambientes deben de ser los grupos de ambientes de prueba que están aparentemente separados en el biplot.

Curiosamente, los dos mega-ambientes sugeridos coinciden con la distribución geográfica de las localidades. OA (Ottawa) y KE (Kemptville) pertenecen al este de Ontario; BH (Bath) también pertenece al este de Ontario pero es mucho más cálido que OA y KE. Los otros seis lugares pertenecen al oeste o al sur de Ontario.

Cualquier sugerencia sobre la existencia de mega-ambientes se debe validar utilizando datos de varios años. En concreto, se debe repetir el patrón “which-won-where” para diferentes años para llegar a la conclusión de la existencia de diferentes mega-ambientes.

3.3.3. Evaluación de cultivos en un mega-ambiente

El patrón “which won where” sugiere por lo tanto que existen dos mega-ambientes: el este de Ontario representado por OA y KE y el sur y oeste de Ontario representado por las otras 7 localizaciones. Un examen de datos MET de varios años confirman esta diferenciación de mega-ambientes (YAN 1999, YAN et al. 2000). El siguiente objetivo es seleccionar cultivos para cada uno de los mega-ambientes. Hasta ahora hemos visto que “zav” es el mejor cultivo en las dos localidades del este de Ontario y “fun” es el mejor en las otras localidades. Los cultivadores normalmente no quieren quedarse con un solo cultivo por lo que es necesario evaluar todos los cultivos desde el punto de vista del rendimiento y de la estabilidad.

3.3.3.1. Rendimiento medio y estabilidad de los genotipos

Para evaluar el rendimiento medio y la estabilidad de los genotipos se dibuja el AEC (average environment coordinate) sobre el JK biplot. En primer lugar se representa el ambiente promedio que se obtiene como la media de las coordenadas de los ambientes (se representa mediante un pequeño círculo). La recta que pasa por el origen del biplot y el ambiente promedio se denomina eje de abscisas del ambiente promedio (abscisa del AEC). Las proyecciones de los marcadores genotipo sobre este eje deben por lo tanto aproximar los rendimientos medios de los genotipos y por tanto los cultivos aparecerán ordenados a lo largo del eje con la flecha apuntando al mayor rendimiento.

Realizamos el estudio sobre los datos anteriores después de eliminar los ambientes OA93 y KE93 (Figura 3.7). El cultivo “fun” es el cultivo con mayor rendimiento medio en el mega-ambiente, seguido por “cas”, “har”, etc.

La ordenada del AEC es la recta que pasa por el origen del biplot y es perpendicular al eje de abscisas anterior. Perpendicular significa ortogonal. Por tanto, si la abscisa del AEC representa a G, la ordenada del AEC aproxima la GEI asociada a cada genotipo y por tanto mide la variabilidad o inestabilidad de los genotipos. Mayor proyección sobre el eje de ordenadas significa mayor inestabilidad. Por lo tanto “rub” y “dia” son más variables y menos estables que los otros cultivos. Los cultivos más cercanos al eje de abscisas son los más estables: “cas”, “zav”, “reb”, “del”, “ari” y “kar”.

Este análisis se obtiene mediante la opción de menú Biplot Tools / Mean vs. Stability del programa GGEBiplotR.

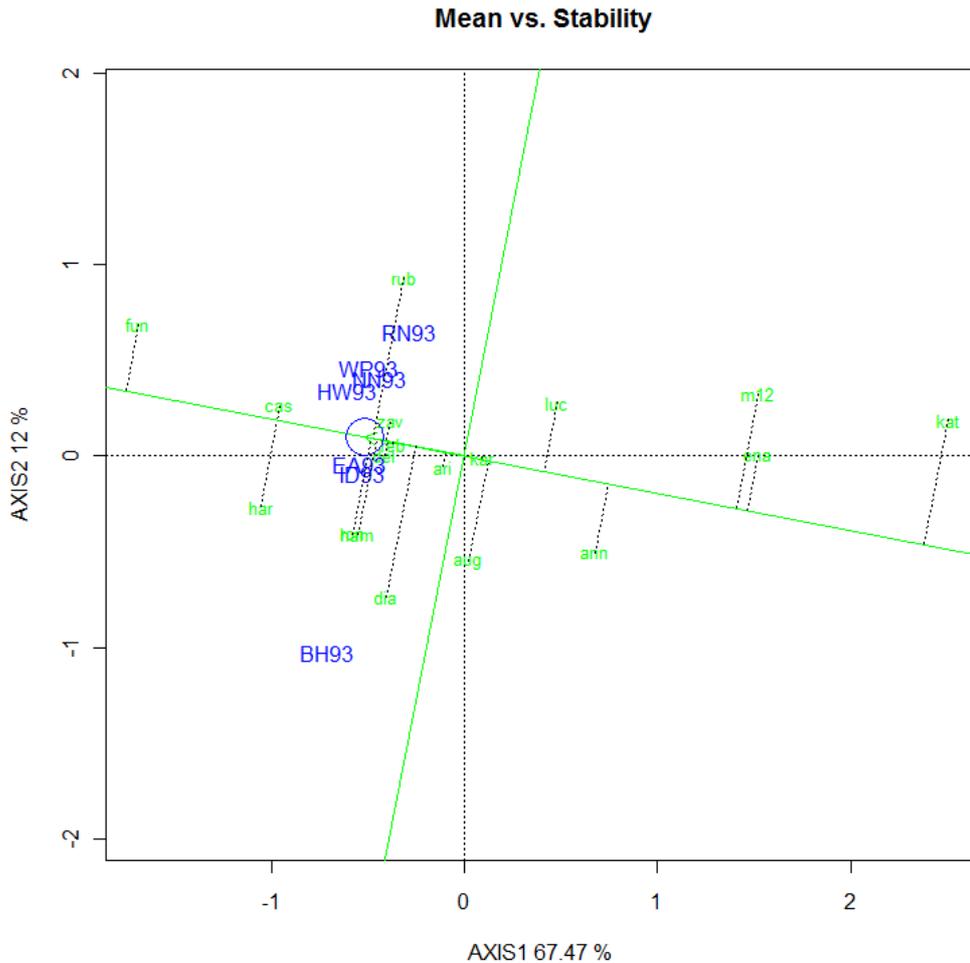


Figura 3.7. Ranking de genotipos basado en el rendimiento medio

3.3.3.2. Ordenación de genotipos respecto al Genotipo “Ideal”

El genotipo ideal debe tener al mismo tiempo el mayor rendimiento medio y alta estabilidad en todos los ambientes. Por tanto el genotipo ideal es un punto del AEA (absolutamente estable) en la dirección positiva y la longitud de su vector asociado debe ser igual a la del mayor vector genotipo del lado positivo del AEA (mayor rendimiento medio). El genotipo ideal normalmente no existe, pero a pesar de ello se puede utilizar como referencia en la evaluación de cultivos. Se puede utilizar la distancia en el gráfico entre cualquier genotipo y el genotipo ideal como una medida de su deseabilidad. Los círculos concéntricos con centro el genotipo ideal permiten visualizar la distancia entre todos los genotipos y el genotipo ideal (Figura 3.8). Por lo tanto “fun” es el más cercano al cultivo ideal, y de ahí que sea el más deseable de

todos los cultivos testeados. Está seguido de “cas” y “har”, los cuales a la vez están seguidos de “ron”, “ham”, “rub”, “zav”, “del”, etc.

Este análisis se obtiene mediante la opción de menú Biplot Tools / Rank Environment – Genotypes / with ref. to the ‘Ideal’ Genotype del programa GGEBiplotR.

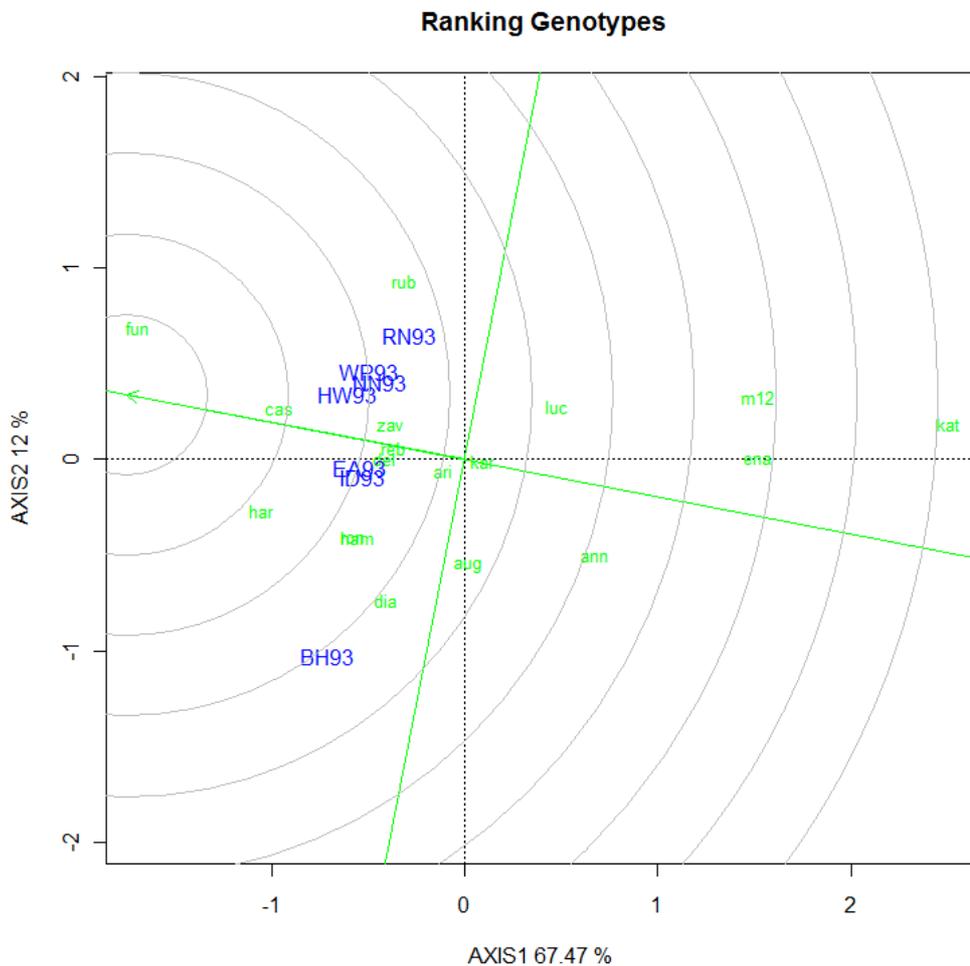


Figura 3.8. Ranking de genotipos con respecto al genotipo “Ideal”

Es interesante observar que el ranking de genotipos en la Figura 3.7, en función del rendimiento medio sólo y la clasificación de cultivares en la Figura 3.8, basado en el rendimiento medio y la estabilidad son casi idénticos. Esto se debe a que G es mucho mayor que GEI en este conjunto de datos. La magnitud de G frente a GE se puede ver por el rango de valores que toma PC1 frente a PC2.

3.3.4. Evaluación de ambientes de prueba

3.3.4.1. Relaciones entre los ambientes

La Figura 3.9 se obtiene según la parametrización siguiente: modelo GGE, sin escalado y SVP = CMP (GH Biplot) y se obtiene a través de la opción de menú Biplot Tools / Relation among Environments del programa GGEBiplotR. La figura se puede interpretar según sigue:

- Las líneas que conectan los marcadores ambiente con el origen del Biplot se denominan vectores ambiente. Los cosenos de los ángulos formados por dos vectores ambiente aproximan la correlación entre ellos. Por ejemplo: OA93 y KE93 están positivamente relacionadas (ángulo agudo), OA93 y RN93 están negativamente correlacionadas (ángulo obtuso) y OA93 y NN93 no están correlacionadas (ángulo recto).
- La presencia de ángulos obtusos “grandes” (es decir gran correlación negativa) entre los vectores ambiente indica una fuerte interacción de cruce GE. En este caso el mayor ángulo es ligeramente superior a 90° (entre RN93 y OA93), lo cual implica que la interacción no es muy grande.
- La distancia entre dos ambientes mide su disimilitud a la hora de discriminar genotipos. Por lo tanto los nueve ambientes se pueden dividir en dos grupos: OA93 y KE93 forman un grupo y el resto de los genotipos forman otro.
- La presencia de fuerte correlación entre dos ambientes de prueba sugiere que la misma información sobre los genotipos se puede obtener a partir de menos ambientes de prueba y por tanto se pueden reducir costos. Si dos entornos de prueba están estrechamente correlacionados constantemente a través de los años, uno de ellos se puede quitar sin pérdida de mucha información.

3.3.4.2. Poder discriminante de los entornos de prueba

- Los círculos concéntricos ayudan a visualizar la longitud de los vectores ambiente, la cual es proporcional a la desviación típica de los respectivos ambientes y por tanto es una medida del poder discriminante de estos. Por lo tanto de entre los nueve ambientes OA93 y KE93 son los más discriminantes y RN93 el menos discriminante.

- Los entornos de prueba con poco poder discriminante proporcionan poca información sobre los genotipos y por lo tanto no deberían utilizarse como entornos de prueba.

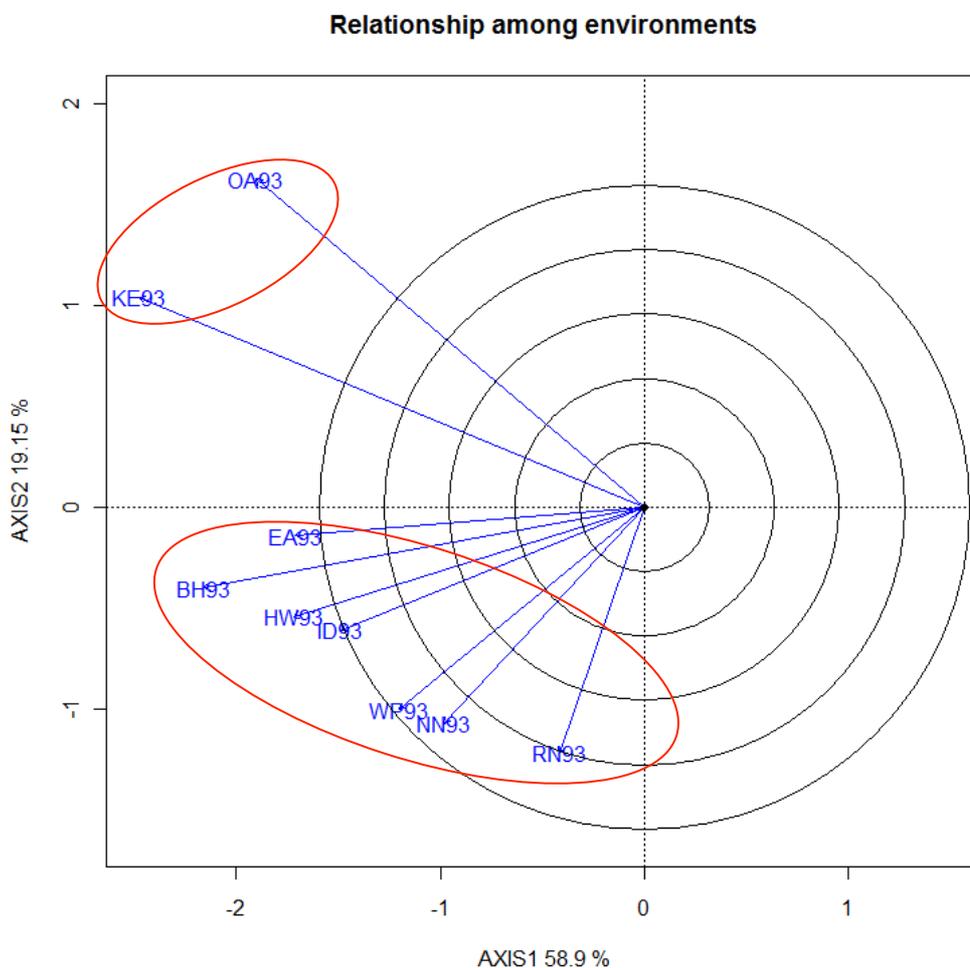


Figura 3.9. Relación entre los entornos de prueba

3.3.4.3. Representatividad de los entornos de prueba

La figura 3.10 presenta el mismo Biplot que la figura anterior salvo que se ha añadido un eje denominado AEA (YAN 2001). El “ambiente medio” (representado por un pequeño círculo en el extremo de la flecha) se obtiene promediando las coordenadas de todos los entornos de prueba y el eje AEA es la recta que pasa por el origen del Biplot y el ambiente medio. Este gráfico se puede interpretar de la siguiente forma:

- Los ambientes de prueba que forman un ángulo pequeño con el eje AEA son más representativos que el resto de los ambientes de prueba. Por lo tanto BH93 es el más representativo y OA93 y KE93 los menos representativos.
- Los ambientes de prueba que son al mismo tiempo discriminantes y representativos (por ejemplo BH93) son buenos entornos de prueba para la selección de genotipos adaptados en general.
- Los ambientes de prueba que son discriminantes pero no representativos (por ejemplo OA93 y KE93) son útiles para seleccionar genotipos específicamente adaptados siempre y cuando se puedan distinguir distintos mega-ambientes.
- Los ambientes de prueba que son discriminantes pero no representativos son útiles para eliminar genotipos inestables siempre y cuando el entorno esté constituido por un solo mega-ambiente.

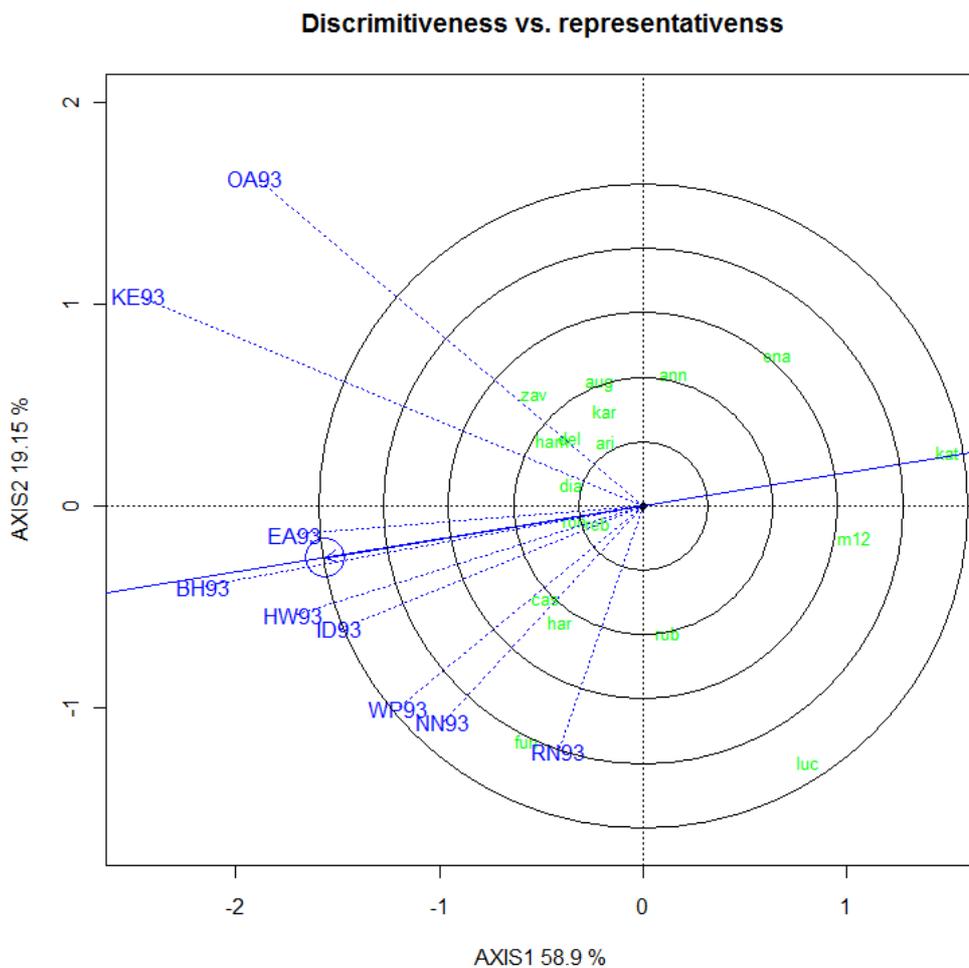


Figura 3.10. Representatividad de los entornos de prueba

Este análisis se obtiene mediante la opción de menú Biplot Tools / Discrimitiveness vs. Representativenss del programa GGEBiplotR.

3.3.4.4. Ambiente “ideal” de prueba para la selección de Genotipos:

Dentro de un solo mega-ambiente el entorno de prueba ideal debe ser el más discriminativo (informativo) y además representativo del entorno de destino. La figura 3.11 define un entorno de prueba “ideal” que es el centro de los círculos concéntricos. Es un punto sobre el eje AEA en la dirección positiva (más representativo) y con una distancia al origen del biplot que es la máxima de todas las distancias de los vectores ambiente (más informativo). BH93 es el más cercano a este punto y por lo tanto el mejor y KE93 y OA93 son los peores ambientes de prueba para la selección de cultivos adaptados a la región completa. De cualquier forma es necesaria información de más años para confirmar que este es el entorno de prueba ideal.

Este análisis se obtiene mediante la opción de menú Biplot Tools / Rank Environment – Genotypes / with ref. to the ‘Ideal’ Environment del programa GGEBiplotR.

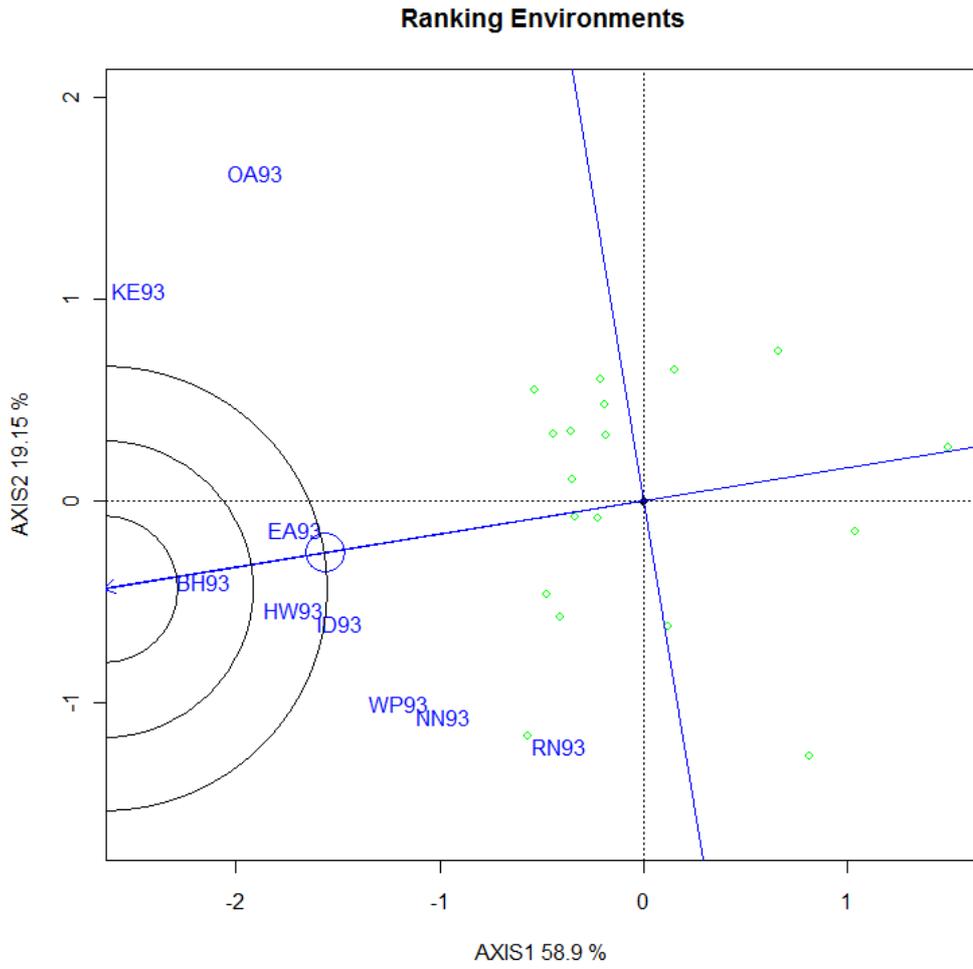


Figura 3.11. Ranking de ambientes basado en representatividad y discriminación

CAPÍTULO IV: PROGRAMA EN R PARA EL ANÁLISIS GGE

En este trabajo se presenta el programa GGEBiplotR, un programa que estará disponible en el CRAN de R en <http://CRAN.R-project.org/>.

Una vez cargado el paquete para su ejecución bastará con escribir en la línea de comandos de R lo siguiente:

```
R> GGEBiplotR()
```

En ese momento aparece una pantalla inicial como la que se muestra en la Figura 4.1. Al pinchar en el botón Start se abre una pantalla desde la cual se solicita el fichero Excel (formato .xls) y la hoja del mismo donde se encuentran los datos con los que se va a trabajar.

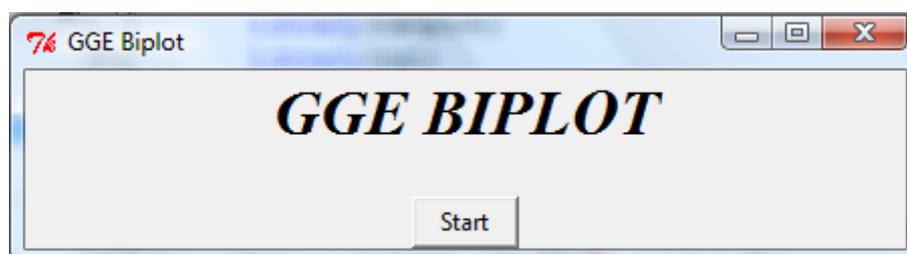


Figura 4.1. Pantalla inicial

Una vez seleccionados los datos aparece una ventana para la selección del modelo (Figura 4.2):

- SVP: Permite seleccionar la factorización que se quiere realizar de la matriz de datos. Las opciones posibles son:
 - GH-Biplot
 - Jk-Biplot
 - SQRT-Biplot
 - HJ-Biplot
- Centrado: Con las siguientes opciones:
 - No centrado
 - Centrado Global (E+G+GE)
 - GGE

- Doble centrado (GE)
- Escalado: Con las opciones:
 - No escalado
 - Desviación típica

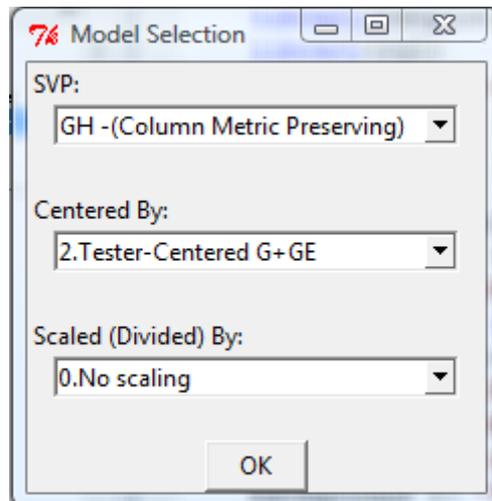


Figura 4.2. Pantalla de selección de modelo

Al pulsar el botón OK aparece una pantalla con el GGE Biplot de acuerdo al modelo seleccionado y un menú que se detalla a continuación:

- **Menú File (Figura 4.3):**
 - Open log file: permite abrir el fichero de log donde se guarda toda la información relativa el Biplot: descomposición en valores singulares, coordenadas de genotipos y ambientes, contribuciones, etc.
 - Copy image: copia el gráfico que se muestra al portapapeles para posteriormente poder pegarlo en cualquier documento.
 - Save image: Permite guardar el gráfico en los formatos PDF, Postscript, Metafile, Bmp, Png, Jpg/Jpeg y TeX file.
 - Print image: imprime el gráfico.
 - Exit: cierra la aplicación.

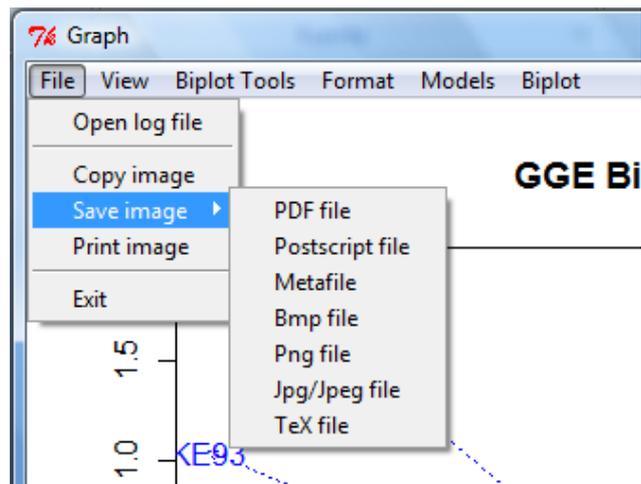


Figura 4.3. Menú File

- **Menú View (Figura 4.4):**

- Show Both: muestra simultáneamente en el Biplot los genotipos y los ambientes (es la opción por defecto).
- Show Genotypes: muestra únicamente los genotipos.
- Show Enviroments: muestra únicamente los ambientes.
- Show/Hide Title: Muestra/ Oculta el título (por defecto aparece el título)
- Show/Hide Gidelines: Muestra/ Oculta los ejes (por defecto se muestran los ejes).

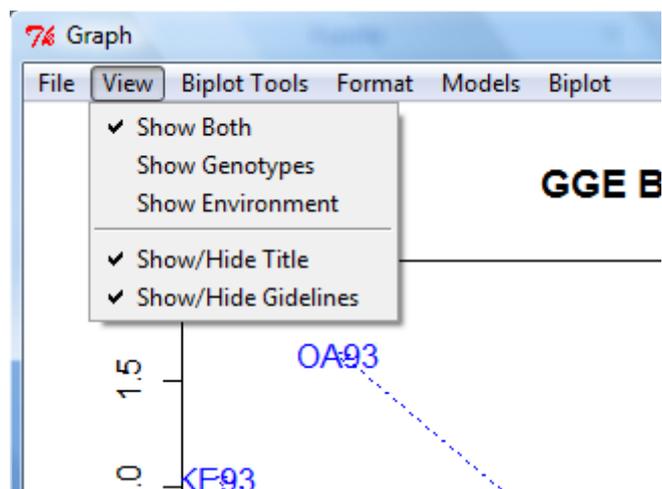


Figura 4.4. Menú View

- **Menú Biplot Tools (Figura 4.5):**
 - Examine a Genotype: Permite seleccionar un genotipo y evaluar su rendimiento en los diferentes ambientes.
 - Examine an Environment: Permite seleccionar un ambiente y evaluar el rendimiento de los distintos cultivos en el ambiente seleccionado.
 - Relation among Environments: Permite visualizar las relaciones entre los distintos entornos de prueba.
 - Compare two Genotypes: Permite la comparación visual de dos genotipos con respecto a todos los entornos de prueba. Al seleccionar esta opción aparece una pantalla desde la cual se pueden seleccionar los dos genotipos que se quieren comparar.
 - Which Won Where/ What: Muestra la línea poligonal que permite realizar el análisis de qué variedades son las más sensibles, es decir las mejores o peores en algún o todos los entornos de prueba.
 - Discrimitiveness vs. Representativenss: Permite realizar un análisis visual de la representatividad de los entornos de prueba.
 - Mean vs. Stability: Permite realizar un estudio visual del rendimiento medio y la estabilidad de los genotipos.
 - Rank Environment with ref. to the Ideal Environment: Permite ordenar los ambientes con respecto a un ambiente de prueba ideal.
 - Rank Genotype with ref.to the Ideal Genotype: Permite ordenar los genotipos con respecto a un genotipo ideal.
 - Back to original data: Muestra el GGE Biplot con el modelo que esté seleccionado en ese momento.

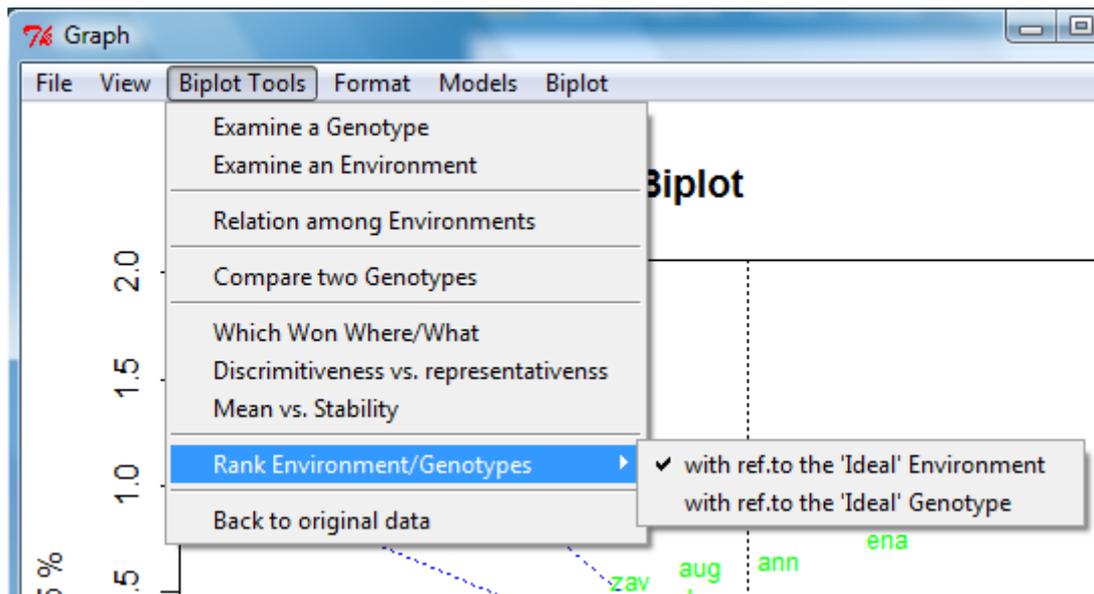


Figura 4.5. Menú Biplot Tools

- **Menú Format (Figura 4.6):**

- Plot Title: Permite modificar el título del gráfico
- Change Color Background: Cambia el color del fondo del gráfico.
- Change Color Genotype labels: Cambia el color de las etiquetas de los genotipos.
- Change Color Environment labels: Cambia el color de las etiquetas de los ambientes.
- Change Color Biplot Title: Cambia el color del título del gráfico.

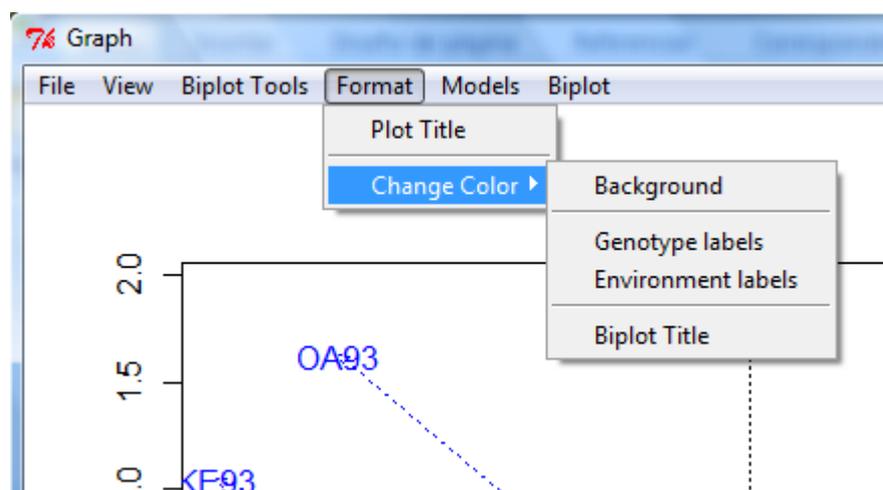


Figura 4.6. Menú Format

- **Menú Models (Figura 4.7):**

- Scaled (divided by):
 - No scaling
 - Std Deviation (SD): Divide por la desviación estándar.
- Centered by:
 - No centering
 - Global – Centered E+G+GE
 - Tester – Centered G+GE
 - Double – Centered GE
- SVP:
 - JK – Row Metric Preserving
 - GH – Column Metric Preserving
 - HJ – Dual Metric Preserving
 - SQ - Symmetrical

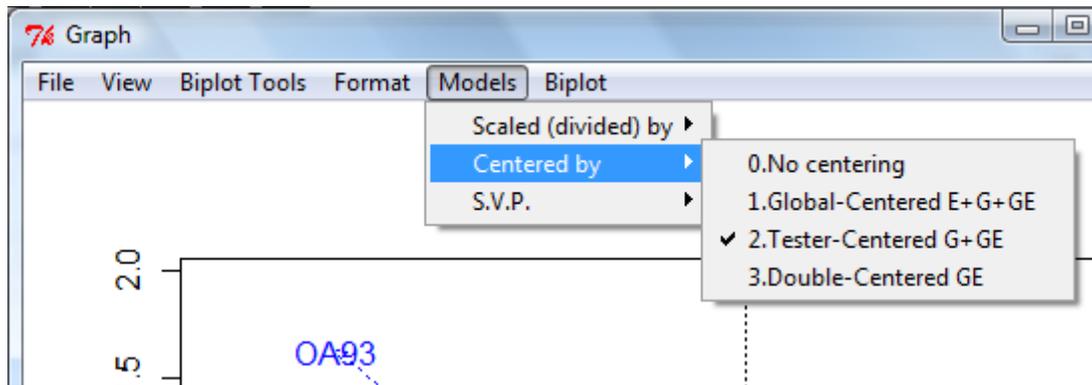


Figura 4.7. Menú Models

- **Menú Biplot (Figura 4.8):**

- PC1 vs. PC2 (Primary). Aparece seleccionado por defecto.
- PC3 vs. PC4
- PC5 vs. PC6
- PC1 vs. PC3
- PC2 vs. PC3
- Biplot 3D: Muestra el biplot en tres dimensiones (Figura 4.9).

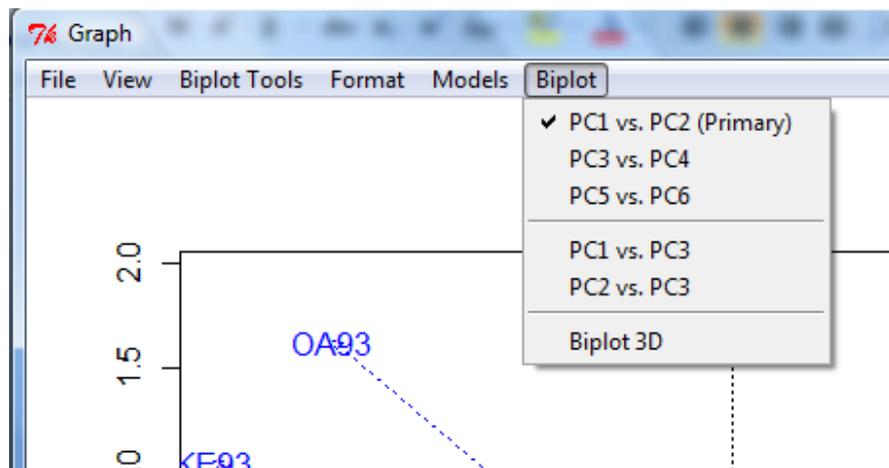


Figura 4.8. Menú Biplot

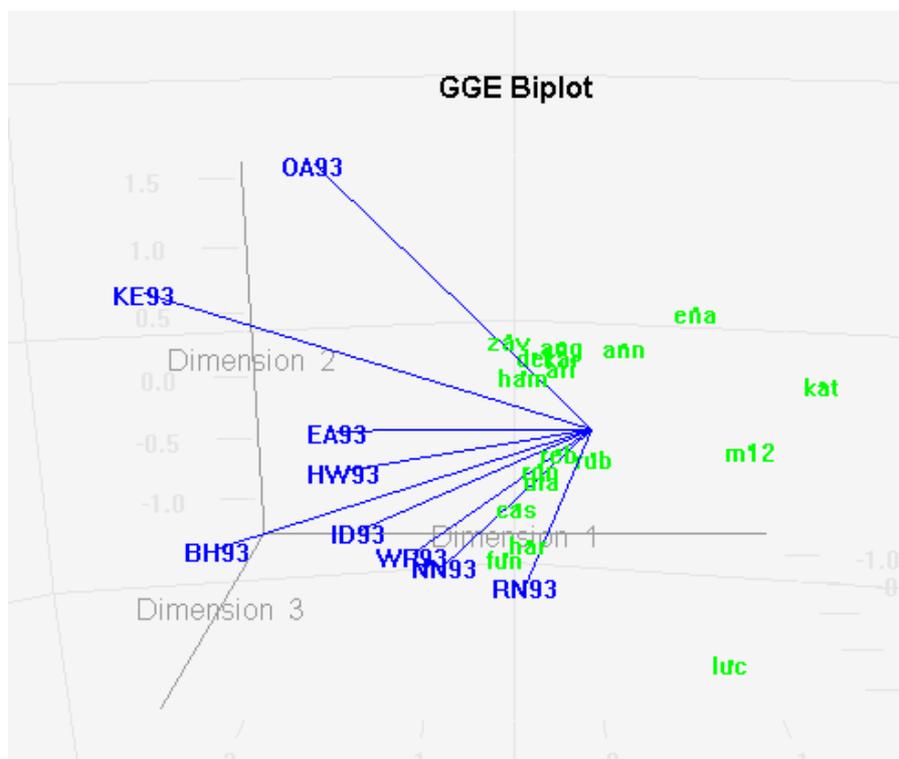


Figura 4.9. Biplot 3D

CONCLUSIONES

El análisis Biplot se ha convertido en una técnica importante dentro del análisis de cultivos, en concreto el GGE Biplot proporciona una solución fácil para el análisis de datos genotipo ambiente. No sólo permite la evaluación de genotipos sino que también permite evaluar los entornos de prueba detectando la presencia o no de mega-ambientes. Dentro de un mega-ambiente permite evaluar si los entornos de prueba son discriminantes y/o representativos. El análisis biplot permite también evaluar los genotipos en términos de rendimiento medio y estabilidad en los ambientes de prueba. Para poder llevar a cabo todos estos análisis se ha presentado en este trabajo el programa GGEbiplotR desarrollado en el lenguaje R.

GLOSARIO

ACP:	Análisis de componentes principales
AMMI:	Modelo de efectos principales aditivos e interacción multiplicativa
CMP-Biplot:	Término equivalente a GH-Biplot
E:	Ambiente
G:	Genotipo
GEI:	Interacción genotipo ambiente
GGE Biplot:	Biplot de la matriz en la que se combina el efecto principal del genotipo y la interacción genotipo ambiente (G+GE).
GH-Biplot:	Biplot que preserva la métrica para las columnas.
HJ-Biplot:	Biplot que superpone los marcadores fila y columna sobre un mismo sistema de referencia con máxima calidad de representación
JK-Biplot:	Biplot que preserva la métrica para las filas.
MET:	Datos de ensayos multambiente
P:	Fenotipo
RMP_Biplot:	Término equivalente a JK-Biplot
SQRT-Biplot:	Biplot simétrico: asigna papeles simétricos a las filas y columnas de la matriz X
SVD:	Descomposición en valores y vectores singulares.

BIBLIOGRAFIA

BAKER, R.J. (1990). Crossover genotype-environmental interaction in spring wheat, p. 42–51. In: M.S. Kang (Ed.), *Genotype-by-environment interaction and plant breeding*, Louisiana State University Agricultural Center, Baton Rouge, LA.

BLÁZQUEZ, A. (1998). Análisis Biplot basado en Modelos Lineales Generalizados. Tesis doctoral. Universidad de Salamanca.

BRADU, D. and GABRIEL, K.R. (1974). Simultaneous statistical inference on interactions in two-way analysis of variance. *Journal of the American Statistical Association*, 29: 428-436.

BRADU, D. and GABRIEL, K.R. (1978). The Biplot as a diagnostic tool for models of two-way tables. *Technometrics*, 20(1): 47-68.

COMSTOCK R.E. and R.H. MOLL. (1963). *Statistical genetics and plant breeding*, National Academy of Science —National Research Council Publishing NAS-NRC, Washington, D.C.

CROSSA, J. and P.L. CORNELIUS. (1997). Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances, *Crop Sci.*, 37: 405–415.

DELACY, I.H., M. COOPER, and K.E. BASFORD. (1996). Relationships among analytical methods used to study genotype-by-environment interactions and evaluation of their impact on response to selection, p.51–84. In: M.S. Kang and H.G. Gauch, Jr. (Eds.), *Genotype-by-environment interaction*, CRC Press, Boca Raton, FL.

DENIS, J.B. (1991). Ajustements de modèles linéaires et bilinéaires sous contraintes linéaires avec données manquantes. *Revue de Statistique Appliquée*, 29(2): 5-24.

DENIS, J.B. and GOWER, J.C. (1992). Biadditive models. Technical Report. Laboratoire de Biométrie, INRA-Versailles.

DENIS, J.B. and GOWER, J.C. (1994). Biadditive models. Letter to the editor. Biometrics, 50: 310-311.

DIAZ-LENO, M.S. (1995). Los métodos Biplot como herramienta de diagnóstico en la modelización de datos multidimensionales. Tesis Doctoral. Universidad de Salamanca.

EBERHART, S.A. and RUSSELL, W.A. (1966). Stability parameters for comparing varieties. Crop Science, 6: 36-40.

ECKART, C. and YOUNG, G. (1936). The approximation of one matrix by another of lower rank. Psychometrika, 1: 211-218.

ECKART, C. and YOUNG, G. (1939). A principal axis transformation for non-Hermitian matrices. Am.Math.Soc.Bull, 45: 118-121.

FINLAY, K.W. and WILKINSON, G.N. (1963). The analysis of adaptation in a plant breeding programme. Australian Journal of Agricultural Research. 14: 742-754.

GABRIEL, K.R. (1971). The Biplot graphic display of matrices with applications to principal components analysis. Biometrika, 58(3): 453-467.

GABRIEL, K.R. (1978). Least Squares Approximation of Matrices by Additive and Multiplicative Models. Journal of the Royal Statistical Society, Series B. 40: 186-196.

GALINDO, M.P. (1985). Contribuciones a la representación simultánea de datos multidimensionales. Tesis Doctoral. Universidad de Salamanca.

GALINDO, M.P. (1986). Una alternativa de representación simultánea: HJ-Biplot. Qüestió, 10(1): 13-23.

GALINDO, P., CUADRAS, C. (1986). Una extensión del método Biplot y su relación con otras técnicas. Publicaciones de Bioestadística y Biomatemática 17, Universidad de Barcelona, España.

GAUCH, H.G. (1988). Model Selection and Validation for Yield Trials with Interaction. *Biometrics*, 44: 705-715.

GAUCH, H.G. and ZOBEL, R.W. (1989). Accuracy and selection success in yield trial analyses. *Theoretical and Applied Genetics*, 77: 473-481.

GAUCH, H.G. (1992). Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam, Netherlands.

GAUCH, H.G. and ZOBEL, R.W. (1996). AMMI analysis of yield trials, p. 85–122. In: M.S. Kang and H.G. Gauch, Jr. (Eds.), *Genotype-by-environment interaction*, CRC Press, Boca Raton, FL.

GAUCH H.G. and R.W. ZOBEL. (1997). Identifying mega-environments and targeting genotypes, *Crop Sci.*, 37: 311–326.

GOLLOB, H.F. (1968). A statistical model which combines features of factor analytic and analyses of variance techniques. *Psychometrika* 33: 73-115.

GOLUB, G.H. and REINSCH, C. (1971). The singular value decomposition. En J:H Wilkinson y C. Reinsch. (eds.). *Handbook of Automatic Computation*. Springer Verlag. Berlin.

GOWER, J.C. (1990). Three-dimensional biplots. *Biometrika*, 77(4): 773-785.

GREENACRE, M.J. (1984). Theory and applications of Correspondence Analysis. Academic Press. London.

GUTIERREZ, J.C., M. LOPEZ, and K.M. EL-ZIK. (1994). AMMI (additive main effects and multiplicative interactions analysis): A tool to determine adaptability of upland cotton genotypes in Spain, Cotton Improvement Conference, Beltwide Cotton Conf. Proc., 2:688–689.

KANG, M.S. (1993). Simultaneous selection for yield and stability in crop performance trials: Consequences for growers, *Agron. J.*, 85: 754–757.

KANG, M.S. (1998). Using genotype-by-environment interaction for crop cultivar development, *Adv. Agron.*, 62: 199–252.

KANG, M.S. and GAUCH, H.G. (1996). Genotype by Environment Interaction. CRC Press. New York.

KEMPTON, R.A. (1984). The use of Biplots in interpreting variety by environment interactions. *Journal of Agricultural Science. Cambridge* 103: 123-135.

MANDEL, J. (1961). Non-additivity in two-way analysis of variance. *Journal of the American Statistical Associations*, 56: 878-888.

MILLIKEN, G.A. and JOHNSON, D.E. (1989). Analysis of Messy Data. Volume 2: Nonreplicated Experiments. New York: Van Nostrand Reinhold.

PITTELKOW, Y.E. and WILSON, S.R. (2003). The GE-biplot for microarray data. *Proc. Virtual Conf. Genomics Bioinformatics* 2: 8-11.

ROMAGOSA, I.; SE ULLRICH, F HAN and PM HAYES (1996). Use of the AMMI model in QTL mapping for adaptation in barley. *Theory Applied Genetic*, 93: 30-37.

TUKEY, J.W. (1949). One degree of freedom for non-additivity. *Biometrics*, 5: 232-242.

VAN EEUWIJK, F.A. (1995a). Linear and bilinear models for the analysis of multi-environment trials: An inventory of models. *Euphytica*, 84: 1-7.

VAN EEUWIJK, F.A. (1995b). Linear and bilinear models for the analysis of multi-environment trials: An application to data from the Dutch Maize Variety Trials. *Euphytica*, 84: 9-22.

VAN EEUWIJK, F.A. and KROONENBERG, P.M. (1998). Multiplicative Models for Interaction in Three-Way ANOVA, with Applications to Plant Breeding. *Biometrics*, 54: 1315-1333.

WOUTERS, L., GOHLMANN, H. W., BIJNENS, L., KSS, S. U., MOLENBERGHS, G. and LEWI, P.J. (2003). Graphical exploration of gene expression data; a comparative study of three multivariate methods. *Biometrics* 59: 1131-1139.

YAN, W. and L.A. HUNT. (1998). Genotype by environment interaction and crop yield, *Plant Breed. Rev.*, 16:135–178.

YAN, W. (1999). A study on the methodology of cultivar evaluation based on yield trial data — with special reference to winter wheat in Ontario, Ph.D. thesis, University of Guelph, Guelph, Ontario, Canada.

YAN, W., HUNT L.A., SHENG, Q. and SZLAVNICS, Z. (2000). Cultivar evaluation and mega-environment investigation based on GGE biplot. *Crop Sci.* 40: 597-605.

YAN, W. (2001). GGEbiplot - a Windows application for graphical analysis of multi-environment trial data and other types of two-way data, *Agron. J.*, 93: 1111–1118.

YAN, W. and I. RAJCAN. (2002). Biplot analysis of sites and trait relations of soybean in Ontario, *Crop Sci.*, 42: 11–20.

YAN, W., and M.S. KANG. (2003). GGE Biplot Analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL.

YAN, W. and TINKER, N.A. (2005). An integrated biplot analysis system for displaying, interpreting, and exploring genotype-by-environment interactions. *Crop Sci.* 45: 1004-1016.

YATES, F. and COCHRAN, W.G. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3: 19-22.

YOUNG, G. and HOUSEHOLDER, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3: 19-22.