

PAULO CANAS RODRIGUES^{1,2}

STANISLAW MEJZA¹

JOÃO TIAGO MEXIA²

¹ Poznan University of Life Sciences, Poland

² Nova University of Lisbon, Portugal

Structuring genotype \times environment interaction — an overview*

Badanie struktury interakcji genotypowo-środowiskowej — przegląd metod

The phenotype of an individual is determined by both the genotype and environment. Farmers and scientists aim to determine a superior genotype over a wide range of environmental conditions but also over years. The basic cause of differences between genotypes in their yield stability is when these two effects are not only additive, i.e. when genotype \times environment interaction (GEI) is present in the data. Multi-location trials play an important role in plant breeding and agronomic research. The data from such trials have three main points: (i) to accurately estimate and predict yield based on limited experimental data; (ii) to determine yield stability and the pattern of response of genotypes across environments; and (iii) to provide reliable guidance for selecting the best genotypes or agronomic treatments for planting in future years and at new sites (Crossa, 1990). The purpose of the present paper is (i) to describe various multivariate statistical methods for analyzing interactions in general and GEI in particular, and (ii) to present a selected bibliography of 142 references to previous work.

Key words: genotype \times environment interaction, additive main effects and multiplicative interaction model, principal component analysis, cluster analysis, biplot

Celem programów genetyczno-hodowlanych, jak i praktyki, jest uzyskanie genotypów (odmian), które wykazują korzystne właściwości w różnych środowiskach. Podstawowym źródłem różnorodności fenotypu odmian jest nieaddytywność tych dwóch czynników, tzn. genotypu i środowiska. Nieaddytywność tę charakteryzujemy poprzez interakcję genotypowo-środowiskową. W badaniach hodowlanych bardzo ważną rolę pełnią doświadczenia wielokrotne i wieloletnie. Wykorzystujemy je głównie do oceny odmian w zakresie estymacji i predykcji plonów, oceny stabilności plonu odmian w różnych środowiskach oraz do rekomendacji uprawowych odmian ze względu na wartość hodowlaną lub rolniczą (Crossa, 1990). W pracy tej dokonujemy przeglądu wielowymiarowych metod analizowania interakcji podwójnej oraz, w szczególności, interakcji genotypowo – środowiskowej. Podajemy także literaturę dotyczącą wyżej wymienionych zagadnień.

Słowa kluczowe: interakcja genotypowo środowiskowa, model addytywny ze względu na efekty główne i multiplikatywny ze względu na interakcję, analiza składowych głównych, metody grupowania, biplot

* This paper is an extension of a paper presented at XVIIIth Summer School of Biometrics, Rackove dolina, June, Slovakia, 2008 A. Corresponding author: e-mail – paulocanas@ft.unl.pt

1. INTRODUCTION

Farmers and scientists aim at determining a genotype which is superior over a wide range of environmental conditions and also over a number of years (multi-environment trials). The basic cause of differences between genotypes in their yield stability is when these two effects (genotype and environment) are not simply additive, i.e. when GEI is present in the data.

Various statistical methods have been proposed for the analysis of interaction in general, and GEI in particular and these will be reviewed in this paper. First of all it is necessary to find if there is an interaction effect present in the data, then one has to consider this effect and importance as a subsequent work. Freeman (1972) described the aim of structuring GEI as: “This is all part of an attempt to discover more about the data than the crude tests reveal; thus, assuming that $data = pattern + noise$, one wants to find as much as possible of the pattern while eliminating the maximum noise”.

To the best of our knowledge there is no recent review paper which provides an up-to-date and extensive list of references about the most used techniques for analyzing and structuring GEI.

The decision as to which papers are most important and relevant is necessarily subjective and doubtful, a lot of very good papers are not included. The aim of this selection is to give some key references as a possible basis for a further reading. Some of the references are very recent and with high quality which means that this subject is still under development.

2. MOTIVATION AND PREVIEW OF THE REVIEW PAPERS

Literally hundreds (even thousands) of papers have been written on this subject by statisticians, breeders, agronomists and geneticists. Over time some authors presented helpful works with many references about GEI: the paper of Aastveit and Mejza (1992) presents a collection with more than 120 references. Also the review works by Denis and Vincourt (1982), Cox (1984), Westcott (1986) and Crossa (1990) and the books by Kang (1990), Gauch (1992) and Kang and Gauch (1996) contain many references. The MSc. Thesis by Alberts (2004) also provides some useful references. However it is not easy to keep up with the expanding literature on this subject and the main reasons for this were pointed out by Kang and Gauch Jr. (1996), (i) “many important developments on interaction are published in statistical journals not read routinely by many agricultural researchers, and relevant statistical applications appear in an even wider assortment of medical, engineering, and other journals”; and (ii) “merely the agricultural literature itself is enormous, appearing in various languages and published in numerous countries. Some important items are published as departmental monographs or as chapters in books that may not come to some researcher’s attention.”

4. MULTIVARIATE ANALYSIS

Crossa (1990) pointed out that multivariate analysis has three main purposes: (i) to eliminate noise from the data pattern (i.e., to distinguish systematic from nonsystematic variation); (ii) to summarize the data; and (iii) to reveal a structure in the data. He also divided the multivariate techniques used to structuring GEI in two groups: (i) ordination techniques — PCA, biplots and graphical representation, principal coordinates analysis and factor analysis — which assume that the data is continuous and attempt to represent genotype and environment relationships in a low-dimensional space; and (ii) classification techniques — cluster analysis and discriminant analysis — which seek for discontinuities in the data, group similar entities in clusters and summarize redundancy in the data. Here we follow the same division as Crossa (1990) to present the most used multivariate techniques for structuring GEI.

4.1. ORDINATION TECHNIQUES

4.1.1. Principal component analysis

When dealing with large data sets, one of the main problems is how to extract the information. Principal component analysis (PCA) is the most used technique for reducing the data set while preserving significant features (e.g. Jolliffe, 2002).

The most immediate objective of PCA is to verify if a small number of the first principal components exists which explains a high proportion of the variation in the original data set. If so, a few principal components can be used to represent the original data set without a big loss of information. This procedure corresponds to the dimension reduction of the original data set.

Another goal of PCA is the visualization of variables underlying the original structure (the principal components) which have physical meaning and, therefore, help to see the initial structure from another point of view.

This analysis can effectively reduce the structure of a two-way genotype \times environment data matrix of G (genotypes) points in E (environments) dimensions in a subspace of fewer dimensions (Crossa, 1990).

The model can be written as

$$Y_{ij} = \mu + \sum_{n=1}^h k_n v_{ni} s_{nj} \quad (1)$$

where Y_{ij} is the observed mean yield of the i -th genotype at the j -th environment, μ is the general mean, k_n is the singular value of the n -th axis/principal component (k_n^2 is the eigenvalue), v_{ni} is the eigenvector of the i -th genotype for the n -th axis, s_{nj} is the eigenvector of the j -th environment for the n -th axis, and $\sum_{n=1}^h v_{ni} = \sum_{n=1}^h s_{nj} = 1$.

Johnson and Graybill (1972) and Corsten and van Eijnsbergen (1972) gave the full statistical analysis underlying the principal component approach and derived the appropriate distribution for a test of significance of the first component.

The model (1), under certain conditions, can be seen as a generalization of the linear regression analysis (Williams, 1952; Mandel, 1969; Johnson, 1977; Digby, 1979; Crossa, 1990).

Another generalization which includes PCA (together with the analysis of variance - ANOVA) is the Additive Main effects and Multiplicative Interaction (AMMI) model (Fisher and Mackenzie, 1923; Mandel, 1971; Gauch, 1988; Zobel, Wright and Gauch 1988; Gauch and Zobel, 1997; Gauch, 2006). The AMMI model, also called doubly-centered PCA, applies the singular value decomposition (SVD) to the data minus the genotype and environment means (Gauch, 2006). In the AMMI model, ANOVA estimates the additive effects of genotypes and environments and SVD or PCA estimates the GEI.

Applications of PCA can be found, among others, in: Mandel (1971); Freeman and Dowker (1973); Hill and Goodchild (1981); Annicchiarico (1992); Jackson *et al.* (1994); Bull *et al.* (1994); Romagosa *et al.* (1996); Martinez *et al.* (1997); Ibanez *et al.* (2001); Laurentin *et al.* (2007).

Three mode PCA (Kroonenberg, 1983) has been used to analyse and structuring GEI when the data are considered as a three dimensional matrix, e.g. Genotype \times Environment \times Year: Basford *et al.*, 1991; van Eeuwijk and Kroonenberg, 1998; de la Vega and Kroonenberg, 2002). Furthermore, one can find some more applications on Kroonenberg and Basford (1989); and de la Vega and Chapman (2001, 2006. Varela *et al.* (2006) presented as a generalization of three mode PCA for AMMI models.

Although PCA is more efficient than the linear regression method in describing genotypic performance (Crossa, 1990), the interpretation of resulting principal components is difficult (Aastveit and Mejza, 1992). This weakness can be overcome by the use of graphical visualisations and biplots which are described in the next subsection. More weak points of PCA can be found in the literature: (i) the reduction of dimensionality of multivariate data may lead to distortions on interpretations (Crossa, 1990); (ii) if the proportion of variance explained by the first principal components is small, the individuals that are quite different may be represented by points that are close together (Gower, 1967); (iii) as a multiplicative model it has the problem of not describing the additive main effects (Zobel *et al.*, 1998) and confuses the additive (main effects of genotypes and environments) structure of the data with nonadditivity effects (GEI) (Crossa, 1990); (iv) and nonlinear association in the data prevents PCA from efficiently describing the real relationships between entities (Williams, 1976). Perkins (1972) also pointed out that PCA was not useful for studying the adaptation of a group of inbred lines of *Nicotiana rustica*.

4.1.2. Biplot and graphical visualization

The display of genotypes and environments along the first two principal component axes for the interaction table of residuals is called a biplot (Gabriel, 1971; Bradu and Gabriel, 1978; Kempton, 1984; Gower and Hand, 1996). This technique can provide useful information on grouping similar genotypes and/or environments and can also provide some useful information about the GEI to identify genotypes which are well adapted to a particular environment (Zobel *et al.*, 1988; Crossa, 1990; Crossa *et al.*, 2002). Three dimensional biplots were presented by Gower and Digby (1981).

Since Bradu and Gabriel (1978) explored the use of the biplot as a diagnostic tool for choosing an appropriate model for the analysis of two-way data it has been used in multi-environment trial data analysis (Yan and Hunt, 2002). Consequently some generalizations and applications have been proposed: GGE biplots display both genotype main effects (G) and genotype \times environment interaction (GE) which are two sources of yield variation that are relevant in cultivar evaluation (Cooper *et al.*, 1997; Yan, 1999, 2001; Yan *et al.*, 2001; Yan and Kang, 2003); Yan and Tinker (2005) presented a third type of biplot (after biplot and GGE biplot) which displays the yield-trait relations in individual environments and addresses whether and how the GEI for yield can be explored by indirect selection for the other traits. A genotype \times trait biplot (Yan and Rajcan, 2002; Yan and Kang, 2003) graphically approximates a genotype \times trait two-way table and can be used to visualize the genetic correlations among traits.

Applications of the biplot methodology to diallel data (Yan and Hunt, 2002), to host-by-pathogen data (Yan and Falk, 2002), to barley yield (Dehghani *et al.*, 2006) and to linear-bilinear models (Crossa *et al.*, 2002), among others, can be found in the literature.

4.1.3. Principal coordinates analysis

Principal coordinates analysis (PCO) was pioneered by Gower (1966) as an alternative to PCA, which is a metric multidimensional scaling method based on projection and uses spectral decomposition to approximate a matrix of distances/dissimilarities by the distances between a set of points in several dimensions. It can be seen as the dual of PCA, i.e. instead of the use of the matrix WW' whose size is the number of variables or environments (PCA), it considers the matrix $W'W$ whose size is the number of individuals or genotypes, where W represents all the variation within genotypes, including that for the error (Freeman, 1972).

In contrast to PCA, PCO employs a broader range of distances (PCO is equivalent to PCA on a covariance matrix of a transposed data matrix if the distance matrix is Euclidean; and PCO is equivalent to PCA on a correlation matrix of a transposed data matrix if the distance matrix is Penrose) or dissimilarity coefficients between objects (Crossa, 1990). The objectives and limitations of PCO are similar to those of PCA with some advantages as pointed out by Crossa (1988, 1990): (i) it is trustworthy when used for data that include extremely low or high yielding sites; (ii) it does not depend on the set of genotypes included in the analysis; and (iii) it is simple to identify stable varieties from the sequence of graphic displays.

Applications of PCO can be found, for example, on Fox and Rosielle (1982); Crossa *et al.* (1988, 1999) and Ibanez *et al.* (2001).

4.1.4. Factor Analysis

Factor Analysis (FA) is similar to PCA where the principal components are replaced by "factors". As in PCA, in FA a large number of correlated variables is reduced to a small number of main (uncorrelated) factors (Spearman, 1904; Cattell, 1965), and assumed that each variable can be expressed as a linear compound of k (lower than the number of individuals) hypothetical variables, the common factors, plus an additional term depending only on the particular variable and known as the specific factor (Gower, 1966). Another difference between PCA and FA is that in FA the rotation of the obtained factors is possible.

A rotation which requires the factors to remain uncorrelated is an *orthogonal* rotation, while others are *oblique* rotations. Oblique rotations often achieve greater simple structure, though at the cost that one must also consider the matrix of factor intercorrelations when interpreting results.

FA has been used to analyse GEI. Grafius and Kiesling (1960) constructed orthogonal vectors representing environmental effects using FA and predicted genotype responses in terms of these vectors. Walton (1972), and Seiler and Stafford (1985) used FA to understand relationships between yield components and morphological characteristics. Gollob (1968) refers to the fixed two-way models

$$Y_{ij} = \mu + \tau_i + \beta_j + \sum_{k=1}^c \theta_k u_{ki} v_{kj} + \varepsilon_{ij}, \quad i = 1, \dots, t; j = 1, \dots, b, \quad (2)$$

where μ is the general mean, τ_i and β_j represent the main effects, and the $\varepsilon_{ij} \sim N(0, \sigma^2)$, as factor analysis of variance or simply FANOVA models because they incorporate the benefits of data reduction from a factor decomposition of residuals and the ease of interpretation permitted by the analysis of variance. Latter, Yochmowitz and Cornell (1978) proposed a stepwise test base upon likelihood ratio statistics to determine the number of multiplicative components in the MANOVA (Multivariate ANOVA) model.

Applications of FA to GEI can be found, among others, in Peterson and Pfeiffer (1989), Oliveira *et al.* (2005) and Garbuglio *et al.* (2007).

4.2. CLASSIFICATION TECHNIQUES

4.2.1. Cluster analysis

Cluster analysis (CA) (Cormack, 1971; Everitt *et al.*, 2001) seeks to identify a set of groups which both minimize within-group variation and maximize between-group variation.

Two types of clustering can be distinguished: (i) hierarchical and (ii) partitional/nonhierarchical. Hierarchical clustering algorithms find successive clusters using previously established clusters and do not require preset knowledge of the number of groups (the traditional representation is called a hierarchical tree diagram or dendrogram). Hierarchical algorithms can be agglomerative — beginning with each element as a separate cluster and merging them into successively larger clusters; or divisive — beginning with the whole set and proceeding to divide it into successively smaller clusters.

Partitional/nonhierarchical algorithms determine all clusters at once and one has to decide in the beginning of the analysis the number of clusters (the most well known partitional algorithm is K-means which uses the Euclidian distance). CA also requires a measure of distance between the individuals which determines how the similarity/dissimilarity of two objects is calculated. The most known and used distance functions are: Euclidean; Mahalanobis, Manhattan, maximum norm and Hamming.

CA is one of the most important methods to analyzing and structuring GEI (Aastveit and Mejza, 1992) and the most widely used technique for classifying environments or genotypes into homogeneous groups (Huhn and Truberg, 2002). CA has been used to study

genotype adaptation by simplifying the pattern of responses and to subdivide genotypes and environments into more homogeneous groups (Crossa, 1990). Lin *et al.* (1986) and Westcott (1987) presented review papers about the application of classification methods to GEI and discussed their problems.

Crossa (1990) pointed out some of the disadvantages of CA: (i) numerous hierarchical groupings exist, and each of them may produce different cluster groups; (ii) the truncation level of the classificatory hierarchies may be decided arbitrarily; (iii) many different similarity measures can be used yielding different results; and (iv) CA may produce misleading results by showing structures and patterns in the data when they do not exist (Gordon, 1999).

With the two-stage Ward-Modified Location Model (Ward-MLM), proposed by Franco *et al.* (1998) where the initial groups are generated by the Ward (1963) minimum variance within-groups hierarchical method, it is possible to avoid the “independence across sites” assumption and therefore to estimate variance of the traits within sites and their covariances across sites (Franco *et al.*, 2003).

Franco *et al.* (1999) extended the Ward-MLM strategy to the case of clustering three-way data of cultivar x environment x trait. This strategy considers the same trait measured in different environments as different variables, and it considers the clustering of n individuals on the basis of p continuous and q categorical traits measured in r different environments (Franco *et al.*, 2003).

Multiplicative models for multi-environment cultivar trials that consider one trait at a time (unvaried case) have been used for developing methods for clustering environments or genotypes into groups with statistically negligible crossover interaction (Cornelius *et al.*, 1992, 1993; Crossa *et al.*, 1993; Kang and Gauch, 1996; Crossa and Cornelius, 1997; Abdalla *et al.*, 1997). The multivariate case is when one considers all the traits (continuous and categorical) simultaneously (three-way Ward-MLM methodology) (Franco *et al.*, 2003).

Basford and McLachlan (1985) also proposed a classification method for three-way cluster analysis when all measured traits are continuous, and an association between categorical and continuous variables can be found in Franco and Crossa (2002) and Franco *et al.*, 2002) with application in Franco *et al.* (2002).

The first attempt to apply CA to GEI was by Abou-El-Fittouh *et al.* (1969) using cotton. Lin (1982) proposed a cluster method to group genotypes according to their response to environments when the GEI is large. Ramey and Rosielle (1983) proposed a new method of cluster analysis, termed the hierarchical agglomerative sums of squares method (HASS) to facilitate clustering of genotypes and environments where GEI exists. Corsten and Denis (1990) proposed an agglomerative hierarchical clustering procedure for an orthogonal two-way table. Franco *et al.* (1998) proposed the two-stage clustering Ward-Modified Location Model, and Lefkovitch (1985) discussed the problems of multi-criteria clustering in GEI.

Huhn and Truberg (2002) compare different cluster techniques, present theoretical results and give wide list of references to preview work in CA, in particular applied to GEI. The experimental results are given by Truberg and Huhn (2002).

Many references to the application of CA to GEI can be found: Robert (1997) used a clustering procedure to identify groups of environments which were homogeneous for interaction in bread wheat, in two multi-location series of trials. Berger *et al.* (2002) used the K-means clustering algorithm to group genotypes of undomesticated Mediterranean Vicia. CA was undertaken by Adugna and Labuschagne (2003) for 10 genotypes of linseed that were tested in a four-times replicated randomized block design across 18 environments (six localities by 3 years) of Ethiopia. Gruneberg *et al.* (2005) derived three environmental groups, for sweet potato clones from CA. A CA was used by Isik and Kleinschmit (2005) in order to illustrate the general pattern of similarities among the test sites of *Picea abies* in northern Germany. Padi (2007) used hierarchical CA to identify discrete groups of cowpea genotypes.

Sometimes, the CA is combined with regression analysis (Tai *et al.*, 1994, Yang *et al.*, 2006) and with the analysis of variance (Lin, 1982; Lefkovitch, 1985; Sanchez-Dominguez *et al.*, 2006).

Many times the multivariate methods constitute a basis for a CA (e.g. Seif *et al.*, 1979; Calinski *et al.*, 1989 a, b; Adugna and Labuschagne, 2003 — for canonical analysis).

4.2.2. Linear discriminant analysis

Linear discriminant function analysis or discriminant analysis (LDA) (Fisher, 1936) aims to find the linear combinations of features which best separate two or more predefined classes of objects or events. The multivariate case, known as Multiple Discriminant Analysis, Discriminant Factor Analysis or Canonical Discriminant Analysis (CDA), derives the canonical coefficients and finds linear combinations of the quantitative variables that provide maximal separation between the classes or groups in much the same way that principal components summarize total variation (from the n original variables it constructs n linear functions — canonical discriminant functions - and the discriminant functions may be used to calculate a set of n discriminant scores for each discriminant function).

Applications of DA and CDA may be found in the literature about GEI. Gallacher (1997) used both a monothetic and a polythetic discriminant analysis on selected characteristics of Australian sugarcane germplasm. Martinez *et al.* (1997) used DA on Grenache N grapevine, and Salehuzzaman and Joarder (1979) on soybean. Crown shape of four different clones planted out in six experimental fields located in five European countries are described and compared using discriminant analysis (Santini and Camussi, 2000). Johnston *et al.* (2001) used CDA to infer the environmental factors most strongly associated with the different genotypic groups in a Louisiana iris hybrid population.

Canonical analysis is concerned with determining the relationships between groups of variables in a data set (Hotelling, 1936; Cliff and Krus, 1976) and has also been used in GEI. Applications of this methodology to GEI can be found in Calinski *et al.* (1987a, 1987b); Lejeune and Calinski (2000) and Adugna and Labuschagne (2003).

5. AMMI MODELS

Work by Williams (1952), Gollob (1968), Mandel (1971) and Bradu and Gabriel (1978) has made an important contribution to the development of additive main effects and multiplicative interaction (AMMI) models (Annicchiarico, 2002). According to Gauch (1988), Zobel *et al.* (1988) and Crossa *et al.* (1990), AMMI models can be used to analyze multilocation trials.

The AMMI method first applies the additive analysis of variance model to a two-way data (e.g. genotypes and environments), and then applies the multiplicative principal component analysis model to the residual from the additive model, i.e. to the interaction (e.g. GEI) (Gauch, 1988, 1992). The AMMI model is given by, see Gauch (1988),

$$Y_{ge} = \mu + \alpha_g + \beta_e + \sum_{n=1}^N \lambda_n \gamma_{gn} \delta_{en} + \theta_{ge}, \quad (3)$$

where:

Y_{ge} is the yield of genotype g in environment e ;

μ is the grand mean;

α_g are the genotype mean deviations (the genotype means minus the grand mean);

β_e are the environment mean deviations;

λ_n is the eigenvalue of PCA axis n ;

γ_{gn} and δ_{en} are the genotype and environment PCA scores for PCA axis n ;

N is the number of PCA axes retained in the model;

θ_{ge} is the residual.

If the experiment is replicated, an error term ε_{ger} , which is the difference between the Y_{ge} mean and the single observation for replicate r , may be added.

Crossa (1990) pointed out three main purposes of AMMI models: (i) model diagnosis — AMMI is more appropriate in the initial statistical analysis of yield trials because it provides an analytical tool for diagnosing other models as subclasses when these are better for a particular data set (Bradu and Gabriel, 1978); (ii) to clarify GEI — AMMI models summarize patterns and relations of genotypes and environments (Kempton, 1984; Zobel *et al.*, 1988; Crossa *et al.* 1990); and (iii) to improve the accuracy of yield estimates — gains have been obtained in the accuracy of yield estimates that are equivalent to increasing the number of replicates by a factor of two to five (Zobel *et al.*, 1988; Crossa *et al.*, 1990) which can be used to reduce the costs by reducing the number of replications, to include more treatments in the experiment, or to improve efficiency in selecting the best genotypes.

There are several possible AMMI models characterized by a number of significant principal component axes ranging from zero (AMMI-0, i.e. additive model) to $\min(g-1, e-1)$, where g is the number of genotypes and e is the number of environments. The full model (AMMI-F), with the highest number of principal component axes, provides a perfect fit between expected and observed data. Models including one (AMMI-1) or two (AMMI-2) PC axes are usually the most appropriate where there is significant GEI. Due to their simplicity, they provide a notable reduction of dimensionality for the adaptation patterns relative to observed data (Annicchiarico, 2002).

6. OTHER TECHNIQUES

6.1. Multivariate analysis of variance

Multivariate analysis of variance (MANOVA) is an extension of ANOVA for several dependent variables. The main aims of MANOVA are identifying whether changes in the independent variables have a significant effect on the dependent variables, and the interactions among the independent variables and the association between dependent variables, if any.

MANOVA has been used to analyse GEI. Yang and Baker (1991) proposed two heuristic tests for the significance of the two causes of GEI. A comparison between MANOVA and REML — restricted maximum likelihood analysis — was presented by Yang (2002).

Application of MANOVA to the analysis of GEI can also be found in Shukla (1972), Calinski *et al.* (1979, 1987 a, 1987 b, 1989 a, 1989 b), Finnie *et al.* (2006), Aggoun *et al.* (2006), among others.

6.2. Additional information — environmental variables

Many times the environment part of GEI can be characterized with some additional variables (Wood, 1976) such as soil characteristics and climate (e.g. temperature, rainfall, radiation level, altitude). Usually the available data is composed of two-way data for yields to structure GEI and of data characterizing the environment (environmental variables), with the authors aiming to relate these two data sets to find relationships.

The most usual way to introduce environmental variables in the analysis of GEI is using linear regression and related techniques. The multilinear regression method can be used for predictive purpose when environmental data are used as independent variables (Knight, 1970).

Factorial regression analysis (FA) (van Eeuwijk *et al.*, 1996) was used by Signor *et al.* (2001) to model the interaction effect using additional genotypic and environmental information for grain yield of early maize, and by Campbell *et al.* (2004) to explain GEI and quantitative trait \times environment interaction (QEI) on wheat. Crossa *et al.* (1999) found that temperature differences across environments accounted for a large portion of the GEI detected in tropical maize (*Zea mays* L.). Ceretta and van Eeuwlik (2008) used FA to model GEI directly in relation to measured environmental variables in malting barley cultivars.

Aastveit and Martens (1986) used the Partial Least Squares (PLS) regression (Wold, 1966; Wold, 1975) to relate year differences in straw length of 15 barley genotypes to climate variation over 9 years. A data set of rainfall, temperature, and global radiation at different stages throughout the growth season over years was related with the residual table of GEI after the factors' effects were removed by an ANOVA. Other applications of PLS regression can be found, for example, in Vargas *et al.* (1999, 2001); Reynolds *et al.* (2004); and Ortiz *et al.* (2007).

Vargas *et al.* (1999) compared results obtained from PLS and FR in two large multi-environment trials and showed that both methods identified similar environmental and genotypic covariates that explain a large proportion of the GEI. Ortiz *et al.* (2007) also presented a comparison between PLS and FR with application to tomato genotypes.

PCA, combined with multiple regression, may be useful for reducing the number of environmental variables to be included in the final analysis (Perkins, 1972; Crossa, 1990).

Reynolds *et al.* (2004) added environmental variables in an AMMI biplot. Yan and Hunt (2001) presented another approach for incorporating genetic and environmental covariates in multi-environment trials data analysis, in which the observed GGE (combination between genotypes and genotype \times environment interaction) patterns were explained as interactions between genetic covariates and environmental covariates. This was achieved by relating the genetic–environmental covariates to the genotypic–environmental scores of the first two principal components derived from GGE biplot analysis (Yan and Tinker, 2005).

7. CONCLUDING REMARKS

A wide range of multivariate methods can be used to structure and analyze GEI on multilocation trial data. Although some of these have limitations on interpretation of the achieved results, they often yield suggestions into particular complex situations and they usually overcome the limitations of linear regression. Nowadays researchers combine different techniques to get better results (e.g. AMMI models joint together ANOVA and PCA).

In this paper we reviewed the most used multivariate techniques and we gave a selection with some references to be considered as a possible basis for further reader. If the reader finds out some important reference missing please let us know.

REFERENCES

- Aastveit A. H., Martens, H. 1986. ANOVA interactions interpreted by partial least squares regression. *Biometrics* 42: 829 — 844.
- Aastveit A. H., Mejza S. 1992. A selected bibliography on statistical methods for the analysis of genotype \times environment interaction. *Biul. Oc. Odm.* 25: 83 — 97.
- Abdalla O. S., Crossa, J. P. L. 1997. Cornelius. Results and biological interpretation of shifted multiplicative model clustering of durum wheat cultivars and testing sites. *Crop Sci.* (37):88 — 97.
- Abou-El-Fittouh H. A., Rawlings J. O., Miller P. A. 1969. Classification of environments to control genotype by environment interactions with an application to cotton. *Crop Sci.* 9: 135 — 140.
- Adugna, W., Labuschagne M. T. 2003. Cluster and canonical variate analyses in multilocation trials of linseed. *J. Agric. Sci.* 140: 297 — 304.
- Aggoun A., Benmahammed K., Dekhili, M. 2006. Effects of environmental factors on kernel weight in five Algerian *durum* wheat populations (*Triticum durum* Desf.). *Cahiers Agriculture* 15: 425 — 431.
- Alberts M. J. A. 2004. A comparison of statistical methods to describe genotype \times environment interaction and yield stability in multi-location maize trials. Master's thesis, University of Free State, South Africa.
- Annicchiarico P. 2002. Genotype \times environment interactions — challenges and opportunities for plant breeding and cultivar recommendations. FAO. Plant Production and Protection Papers.
- Annicchiarico P. 1992. Cultivar adaptation and recommendation from alfalfa trials in northern Italy. *Journal of Genetics and Breeding* 46: 269 — 278.
- Basford K., McLachlan G. J. 1985. The mixture method of clustering applied to three-way data. *J. Classif.*, (2): 109 — 125.
- Basford K. E., Kroonenberg P. M., DeLacy I.H. 1991. Three-way methods for multivariate genotype \times environment data: an illustrated partial survey. *Field Crops Res.* 27: 131 — 157.

- Berger J. D., Robertson L. D., Cocks, P. S. 2002. Genotype \times environment interaction for yield and other plant attributes among undomesticated Mediterranean *Vicia* species. *Euphytica* 126: 421 — 435.
- Bradu D., Gabriel K. R. 1978. The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20: 47 — 68.
- Bull J. K., Basford K. E., Cooper M., Delacy I. H. 1994. Enhanced interpretation of pattern analysis of environments — The use of blocks. *Field Crops Res.* 37: 25 — 32.
- Calinski T., Czajka S., Kaczmarek Z. 1979. On some methods for studying the genotype \times environment interaction. *Quaderni di Epidemiologia* 1: 11 — 30.
- Calinski T., Czajka S., Kaczmarek Z. 1987. A model for the analysis of a series of experiments repeated at several places over a period of years. II. Example. *Biul. Oc. Odm.* 17–18: 35 — 71.
- Calinski T., Czajka S., Kaczmarek Z. 1987. A model for the analysis of a series of experiments repeated at several places over a period of years. I. Theory. *Biul. Oc. Odm.* 17–18: 7 — 33.
- Calinski T., Czajka S., Kaczmarek Z. 1989. A model for the analysis of a series of variety trials repeated at places subject to grouping. II. Example. *Biul. Oc. Odm.* 21–22: 45 — 64.
- Calinski T., Czajka S., Kaczmarek Z. 1989. A model for the analysis of a series of variety trials repeated at places subject to grouping. I. Theory. *Biul. Oc. Odm.* 21–22: 27 — 44.
- Campbell B. T., Baenziger P. S., Eskridge K. M., Budak H., Streck N. A., Weiss A., Gill K. S., Erayman M. 2004. Using environmental covariates to explain genotype \times environment and QTL \times environment interactions for agronomic traits on chromosome 3A of wheat. *Crop Sci.* 44: 620 — 627.
- Cattell R. F. 1965. Factor analysis: An introduction to essentials. I. The purpose and underlying models. *Biometrics* 21: 190 — 215.
- Ceretta S., van Eeuwlik F. 2008. Grain yield variation in malting barley cultivars in Uruguay and its consequences for the design of a trials network. *Crop Sci.* 48: 167 — 180.
- Cliff N., Krus D. J. 1976. Interpretation of canonical variate analysis: Rotated vs. Unrotated solutions. *Psychometrika* 41: 35 — 42.
- Cooper M., Stucker R. E., DeLacy I. H., Harch B. D. 1997. Wheat breeding nurseries, target environments, and indirect selection for grain yield. *Crop Sci.* 37: 1168 — 1176.
- Cormack R. M. 1971. A review of classification. *Journal of the Royal Statistical Society A*, 134: 321 — 367.
- Cornelius P. L., Seyedsadr M., Crossa J. 1992. Using the shifted multiplicative model to search for “separability” in crop cultivar trials. *Theor. Appl. Gen.* (84): 161 — 172.
- Cornelius P. L., Van Sanford D. A., Seyedsadr M. S. 1993. Clustering cultivars into groups without rank-change interactions. *Crop Sci.* (33): 1193 — 1200.
- Corsten L. C. A., Denis J. B. 1990. Structuring interaction in two-way tables by clustering. *Biometrics* 46: 207 — 215.
- Corsten L. C. A., van Eijnsbergen A. C. 1972. Multiplicative effects in two-way analysis of variance. *Statistica Neerlandica* 26: 61 — 68.
- Cox D. R. 1984. Interaction. *International Statistical Review* 52: 1 — 31.
- Crossa J. 1988. A comparison of results obtained with two methods for assessing yield stability. *Theor. Appl. Gen.* 75: 460 — 467.
- Crossa J. 1990. Statistical analyses of multilocation trials. *Advances in Agronomy* 44: 55 — 85.
- Crossa J., Cornelius P. L. 1997. Site regression and shifted multiplicative model clustering of cultivar trials site under heterogeneity of error variances. *Crop Sci.* (37): 406 — 415.
- Crossa J., Cornelius P. L., Seyedsadr M., Byrne P. 1993. A shifted multiplicative model cluster analysis for grouping environments without genotypic rank change. *Theor. Appl. Gen.* (85): 577 — 586.
- Crossa J., Cornelius, P. L., Yan W. 2002. Biplots of linear-bilinear models for studying crossover genotype \times environment interaction. *Crop Sci.* 42: 619 — 633.
- Crossa J., Gauch H. G., Zobel R. W. 1990. Addictive main effects and multiplicative interaction analysis of two international maize cultivar trials. *Crop Sci.* (6): 36 — 40.
- Crossa J., Vargas M., van Eeuwijk F. A., Jiang C., Edmeades G. O., Hoisington D. 1999. Interpreting genotype \times environment interaction in tropical maize using linked molecular markers and environmental covariates. *Theor. Appl. Gen.* 99: 611 — 625.

- Crossa J., Westcott B., Gonzalez C. 1988. Analyzing yield stability of maize genotypes using a spatial model. *Theor. Appl. Gen.* 75: 863 — 868.
- de la Vega A. J., Chapman S. C. 2001. Genotype by environment interaction and indirect selection for yield in sunflower II. Three-mode principal component analysis of oil and biomass yield across environments in Argentina. *Field Crops Res.* 72: 39 — 50.
- de la Vega A. J., Chapman S. C. 2006. Multivariate analyses to display interactions between environment and general or specific combining ability in hybrid crops. *Crop Sci.* 46: 957 — 967.
- de la Vega, A. J., Kroonenberg, P. M. 2002. Investigating the physiological bases of predictable and unpredictable genotype by environment interactions using three-mode pattern analysis. *Field Crops Res.* 78: 165 — 183.
- Dehghani H., Ebadi A., Yousefi A. 2006. Biplot analysis of genotype by environment interaction for barley yield in Iran. *Agronomy J.* 98: 388 — 393.
- Denis, J. B., Vincour T. P. 1982. Panorama des méthodes statistiques d'analyse des interactions génotype × milieu. *Agronomie* 2: 219 — 230.
- Digby P. G. N. 1979. Modified joint regression analysis for incomplete variety × environment data. *J. Agric. Sci.* 93: 81 — 86.
- Everitt B. S., Landau S., Leese M. 2001. Cluster analysis. Oxford University Press Inc., New York, 2001.
- Finnie S. M., Bettge A. D., Morris C. F. 2006. Influence of cultivar and environment on water-soluble and water-insoluble arabinoxylans in soft wheat. *Cereal Chem.* 83: 617 — 623.
- Fisher R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7: 179 — 188.
- Fisher R. A., Mackenzie W. A. 1971. Studies in crop variation. II. The manorial response of different potato varieties. *J. Agric. Sci.* 23:311 — 320.
- Fox P. N., Rosielle A. A. 1982. Reference sets of genotypes and selection for yield in unpredictable environments. *Crop Sci.* 22: 1171 — 1175.
- Franco J., Crossa J. 2002. The modified location model for classifying genetic resources: I. Association between categorical and continuous variables. *Crop Sci.* (42): 1719 — 1726.
- Franco J., Crossa J., Taba S., Eberhart S. A. 2002. The modified location model for classifying genetic resources: II. unrestricted variance covariance matrices. *Crop Sci.* (42): 1727 — 1736.
- Franco J., Crossa J., Taba S., Shands H. 2003. A multivariate method for classifying cultivars and studying group × environment × trait interaction. *Crop Sci.* (43): 1249 — 1258.
- Franco J., Crossa J., Villasenor J., Taba S., Eberhart S. A. 1998. Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38: 1688 — 1696.
- Franco J., Crossa J., Villasenor J., Taba S., Eberhart S. A. 1999. A two-stage, three-way method for classifying genetic resources in multiple environments. *Crop Sci.* (39): 259 — 267.
- Freeman G. H. 1972. Statistical methods for the analysis of genotype-environment interactions. *Heredity* 31: 339 — 254.
- Freeman G. H., Dowker B. D. 1973. The analysis of variation between and within genotypes and environments. *Heredity* 30: 97 — 109.
- Gabriel K. R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58: 453 — 467.
- Gallacher D. J. 1997. Optimised descriptors recommended for Australian sugarcane germplasm (*Saccharum* spp. hybrid). *Australian J. Agric. Res.* 48: 775 — 779.
- Garbuglio D. D., Gerage A. C., de Araujo P. M., Fonseca N. D., Shioga P. S. 2007. Factor analysis and bissegmented regression for studies about environmental stratification and maize adaptability. *Pesquisa Agropecuaria Brasileira* 42: 183 — 191.
- Gauch, H. G. 1988. Model selection and validation for yield trials with interaction. *Biometrics* 44: 705 — 715.
- Gauch H. G. 1992. *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*. Elsevier, Amsterdam.
- Gauch H. G. 2006. Statistical analysis of yield trials by AMMI and GGE. *Crop Sci.* 46: 1488 — 1500.
- Gauch H. G., Zobel R. W. 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 analysis of variance techniques. *Psychometrika* 33: 73 — 116.
- Gordon A. D. 1999. *Classification*. Chapman and Hall/CRC, Boca Raton, FL.
- Gower J. C. 1999. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53: 325 — 338.
- Gower J. C. 1967. Multivariate analysis and multidimensional geometry. *The Statistician* 17: 13 — 28.
- Gower J. C., Digby P. G. N. 1981. Three dimensional biplots. *Biometrika* 77: 773 — 785.
- Gower J. C., Hand D. J. 1996. *Biplots*. Chapman and Hall, United Kingdom.
- Grafius J. E., Kiesling R. L. 1960. The prediction of the relative yields of different oat varieties based on known environmental variables. *Agron. J.* 52: 396 — 399.
- Gruneberg W. J., Manrique K., Zhang, D., Hermann M. 2005. Genotype \times environment interactions for a diverse set of Sweetpotato Clones evaluated across varying ecogeographic conditions in Peru. *Crop Sci.* 45: 2160 — 2171.
- Hill J., Goodchild N. A. 1981. Analysing environments for plant breeding purposes as exemplified by multivariate analyses of long term wheat yields. *Theor. Appl. Gen.* 58: 317 — 325.
- Hotelling H. 1936. Relations between two sets of varieties. *Biometrika* 28: 321 — 337.
- Huhn M., Truberg B. 2002. Contributions to the analysis of genotype \times environment interactions: Theoretical results of the application and comparison of clustering techniques for the stratification of field test sites. *Journal of Agronomy and Crop Sci.* 188: 65 — 72.
- Ibanez M. A., Di Renzo M. A., Samame S. S., Bonamico N. C., Poverne M. M. 2001. Genotype-Environment interaction of Lovegrass forage yield in the semi-arid region of Argentina. *The Journal of Agricultural Science* 137: 329 — 336.
- Isik, K., Kleinschmit J. 2005. Similarities and effectiveness of test environments in selecting and deploying desirable genotypes. *Theor. Appl. Gen.* 110: 311 — 322.
- Jackson, P. A., Byth D. E., Fischer K. S., Johnston R. P. 1994. Genotype \times environment interaction in progeny from a barley cross. 2. Variation in grain-yield, yield components and dry-matter production among lines with similar times to anthesis. *Field Crops Res.* 37: 11 — 23.
- Johnson D. E., Graybill F. A. 1972. An analysis of a two-way model with interaction and no replication. *Statistica Neerlandica* 67: 862 — 868.
- Johnson G. R. 1977. Analysis of genotypic similarity in terms of mean yield and stability of environmental response in a set of maize hybrids. *Crop Sci.* 17: 837 — 842.
- Johnston J. A., Wesselingh R. A., Bouck A. C., Donovan L. A., Arnold M. L. 2001. Intimately linked or hardly speaking? The relationship between genotype and environmental gradients in a Louisiana iris hybrid population. *Molecular Ecology* 10: 673 — 681.
- Jolliffe I. T. 2002. *Principal component analysis*. Second ed. Springer Verlag, New York Inc.
- Kang M. S., Gauch, H. G. Jr. (ed.). 1996. *Genotype-by-environment interaction*. CRC Press, New York.
- Kang M. S. (ed.). 1990. *Genotype-by-environment interaction and plant breeding*. Louisiana State University Agricultural Centre, Baton Rouge, Louisiana.
- Kempton R. A. 1984. The use of biplot in interpreting variety by environment interactions. *The Journal of Agricultural Science* 103: 123 — 135.
- Knigh R. 1970. The measurement and interpretation of genotype environment interactions. *Euphytica* 19: 225 — 235.
- Kroonenberg P. M. 1983. *Three-mode Principal Component Analysis: Theory and Applications*. DWO Press, Leiden, The Netherlands.
- Kroonenberg P. M., Basford K. E. 1989. An investigation of multi-attribute genotype response across environments using 3-mode principal component analysis. *Euphytica* 44:109 — 123.
- Laurentin H., Montilla, D., Garcia V. 2007. Interpreting genotype \times environment interaction in sesame (*Sesamum indicum* L.). *The Journal of Agricultural Science* 145: 263 — 271.
- Lefkovitch L. P. 1985. Multi-criteria clustering in genotype-environment interaction problems. *Theor. Appl. Gen.* 70: 585 — 589.
- Lejeune M., Calinski, T. 2000. Canonical analysis applied to multivariate analysis of variance. *Journal of Multivariate Analysis* 72: 100 — 119.

- Lin C. S. 1982. Grouping genotypes by a cluster method directly related to genotype-environment interaction mean square. *Theor. Appl. Gen.* 62: 277 — 280.
- Lin C. S., Binns, M. R., Lefkovitch L. P. 1986. Stability analysis: Where do we stand? *Crop Sci.* 5: 894 — 900.
- Mandel J. 1969. A method for fitting empirical surfaces to physical or chemical data. *Technometrics* 3: 411 — 429.
- Mandel J. 1971. A new analysis of variance model for non-additive data. *Technometrics* 13: 1 — 18.
- Martinez M. C., Boursiquot J. M., Grenan S., Boidron R. 1997. Ampelometric study of adult leaves from somaclones of the Grenache — N grapevine (*Vitis vinifera* L.). *Canadian Journal of Botany — Revue Canadienne de Botanique* 75: 333 — 345.
- Oliveira G. V., Carneiro P. C., Dias L. A., Carneiro J. E., Cruz, C. D. 2005. Factor analysis in the environment stratification for the evaluation of common bean cultivars. *Crop Breeding and Applied Biotechnology* 5: 166 — 173.
- Ortiz R., Crossa, J., Vargas, M., Izquierdo J. 2007. Studying the effect of environmental variables on the genotype x environment interaction of tomato. *Euphytica* 153: 119 — 134.
- Padi F. K. 2007. Genotype × environment interaction and yield stability in a cowpea-based cropping system. *Euphytica* 158: 11 — 25.
- Perkins J. M. 1972. The principal component analysis of genotype-environmental interactions and physical measures of the environment. *Heredity* 29: 51 — 70.
- Peterson C. J., Pfeiffer W. H. 1989. International winter wheat evaluation: Relationships among test sites based on cultivar performance. *Crop Sci.* 29: 276 — 282.
- Ramey T. B., Rosielle A. A. 1983. Hass cluster analysis: a new method of grouping genotypes or environments in plant breeding. *Theor. Appl. Gen.* 66: 131 — 133.
- Reynolds M. P., Trethowan R., Crossa J., Vargas, M., Sayre K. D. 2004. Physiological factors associated with genotype by environment interaction in wheat. *Field Crops Research* 85: 253 — 274.
- Robert N. 1997. Structuring genotype × environment interaction for quality traits in bread wheat, in two multi-location series of trials. *Euphytica* 97: 53 — 66.
- Romagosa I., Ullrich S. E., Han F., Hayes P. M. 1996. Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theor. Appl. Gen.* 93: 30 — 37.
- Salehuzzaman M. Joarder, O. I. 1979. Genotype × environment interaction, diversity estimates and application of discriminated function selection in soybean. *Genet. Pol.* 20: 89 — 101.
- Sanchez-Dominguez S., Munoz-Orozco A., Gonzalez-Hernandez V. A., Martinez-Garza A. 2006. Characterization and classification of Mexican peanut (*Arachis hypogaea* L.) germplasm. *Agrociencia* 40: 171 — 182.
- Santini A., Camussi A. 2000. The environmental effect on crown shape of common cypress clones in the Mediterranean countries. *Annals of Forest Science* 57: 277 — 286.
- Seif E., Evans J. C., Calaam L. N. 1979. A multivariate procedure for classifying environments according to their interaction with genotypes. *Australian Journal of Agricultural Research* 30: 1021 — 1026.
- Seiler G. P., Stafford R. E. 1985. Factor analysis of components of yield in guar. *Crop Sci.* 25: 905 — 908.
- Shukla G. K. 1972. Some statistical aspects of partitioning genotype-environment components of variability. *Heredity* 29: 237 — 245.
- Signor C. E., Dousse S., Lorgeou J., Denis J. B., Bonhomme R., Carolo P., Charcosset C. 2001. Interpretation of genotype × environment interactions for early maize hybrids over 12 years. *Crop Sci.* 41: 663 — 669.
- Spearman C. 1904. General intelligence objectively determined and measured. *American Journal of Psychology* 15: 201 — 293.
- Tai G. C. C., Levy D., Coleman W. K. 1994. Path analysis of genotype-environment interactions of potatoes exposed to increasing warm-climate constraints. *Euphytica* 75: 49 — 61.
- Truberg B. Huhn M. 2002. Contributions to the analysis of genotype × environment interactions: Experimental results of the application and comparison of clustering techniques for the stratification of field test sites. *Journal of Agronomy and Crop Science* (188): 113 — 122.
- Van Eeuwijk F. A., Denis J. B., Kang M. S. 1996. Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In: *Genotype by environment interaction*. CRC Press, Boca Raton, Kang, F. L., Gauch, H. G. Jr. (ed.): 15 — 49.

- Van Eeuwijk F. A., Kroonenberg P. M. 1998. Multiplicative models for interaction in three-way ANOVA, with applications to plant breeding. *Biometrics* 54: 1315 — 1333.
- Varela M., Crossa J., Rane J., Joshi A. K., Trethowan R. 2006. Analysis of a three-way interaction including multi-attributes. *Australian Journal of Agricultural Research* 55: 1185 — 1193.
- Vargas M., Crossa, J., Van Eeuwijk, F. A., Ramirez, M. E., Sayre K. 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype \times environment interaction. *Crop Sci.* 39: 955 — 967.
- Vargas M., Crossa J., Van Eeuwijk F. A., Sayre, K., Reynolds M. P. 2001. Interpreting treatment \times environment interaction in agronomy trials. *Agron. J.* 93: 949 — 960.
- Walton P. D. 1972. Factor analysis of yield in spring wheat (*Triticum aestivum* L.). *Crop Sci.* 12: 731 — 733.
- Ward J. H. 1963. Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* (58): 236 — 244.
- Westcott B. 1986. Some methods of analysing genotype-environment interactions. *Heredity* 56: 243 — 253.
- Westcott B. 1987. A method of assessing the yield stability of crop genotypes. *The Journal of Agricultural Science* 108: 267 — 274.
- Williams E. J. 1952. The interpretation of interactions in factorial experiments. *Biometrika* 39: 65 — 81.
- Williams W. T. 1976. *Pattern Analysis in Agriculture Science*. CSIRO, Melbourne and Elsevier, Amsterdam.
- Wold H. 1966. Estimation of principal components and related models by iterative least squares. In: *Multivariate analysis*. Krishnaiah P. R. (ed.). Academic Press, New York: 391 — 420.
- Wold H. 1975. Soft modelling by latent variables: The nonlinear iterative partial least squares approach. In: *Perspectives in probability and statistics: Papers in honour of Bartlett, M. S., Gani, J.* (ed.). Academic Press, London: 117 — 142.
- Wood J. T. 1976. The use of environmental variables in the interpretation of genotype-environment interaction. *Heredity* 37: 1 — 7.
- Yan W. 1999. Methodology of cultivar evaluation based on yield trial data with special reference to winter wheat in Ontario. PhD thesis, University of Guelph, Guelph, ON, Canada.
- Yan W. 2001. GGE biplot - 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., Cornelius P. L., Crossa J., Hunt L. A. 2001. Two types of GGE biplots for analyzing multi-environment trial data. *Crop Sci.* 41: 656 — 663.
- Yan W., Falk D. E. 2002. Biplot analysis of host-by-pathogen data. *Plant Dis.* 86: 1396 — 1401.
- Yan W., Hunt L. A. 2001. Interpretation of genotype \times environment interaction for winter wheat yield in Ontario. *Crop Sci.* 41: 19 — 25.
- Yan W., Hunt L. A. 2002. Biplot analysis of diallel data. *Crop Sci.* 42: 21 — 30.
- Yan W., Kang M. S. 2003. *GGEbiplot analysis: A graphical tool for breeders, geneticists, and agronomists*. CRC Press, Boca Raton.
- Yan W., Rajcan I. 2002. Biplot evaluation of test sites and trait relations of soybean in Ontario. *Crop Sci.* 42: 11 — 20.
- Yan W., Tinker N. A. 2005. An integrated biplot analysis system for displaying, interpreting, and exploring genotype \times environment interaction. *Crop Sci.* 45: 1004 — 1016.
- Yang R.-C. 2002. Likelihood-based analysis of genotype \times environment interactions. *Crop Sci.* 42: 1434 — 1440.
- Yang R.-C., Beker R. J. 1991. Genotype-environment interactions in two wheat crosses. *Crop Sci.* 31: 83 — 87.
- Yang R. C., Stanton D., Blade S. F., Helm J., Spaner D., Wright S., Domitruk D. 2006. Is yield analysis of barley cultivar trials in the Canadian prairies. *Journal of Agronomy and Crop Science* 192: 184 — 294.
- Yochmowitz M. G., Cornell R. G. 1978. Stepwise tests for multiplicative components of interaction. *Technometrics* 20: 79 — 84.
- Zobel R. W., Wright M. J., Gauch H. G. 1988. Statistical analysis of yield trial. *Agron. J.* 80: 388 — 393.

ACKNOWLEDGMENTS

The first author wishes to express his thanks to Fundação para a Ciência e Tecnologia (Portuguese foundation for science and technology), Grant Number SFRH/BD/35994/2007, for its financial support.