Conference on Applied Statistics in Agriculture   1991 - 3rd Annual Conference Proceedings

ANALYSIS OF GENOTYPE X ENVIRONMENT INTERACTION BY GRAPHICAL TECHNIQUES

George C.J. Fernandez

Recommended Citation

Fernandez, George C.J. (1991). "ANALYSIS OF GENOTYPE X ENVIRONMENT INTERACTION BY GRAPHICAL TECHNIQUES," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1424

ANALYSIS OF GENOTYPE X ENVIRONMENT INTERACTION BY GRAPHICAL TECHNIQUES

George C.J. Fernandez
Department of Agricultural Economics
University Of Nevada- Reno
Nevada Reno NV 89557

Abstract

Genotype x Environment interactions results from the changes in the magnitude of differences among genotypes (non-crossover or quantitative interactions) or changes in the relative ranking of the genotypes (crossover or qualitative interactions) in different environments. Non-crossover interactions are usually associated with variance heterogeneity and non-additivity. The analysis of variance combined with joint regression analysis failed to differentiate between the crossover and non-crossover interactions. Tedious computations are necessary in comparisons of all possible pairs of genotypes in all possible pairs of environments in the crossover detection tests. Therefore, differentiating the non-crossover interaction caused by variance heterogeneity and non-additivity from crossover interaction by simple but effective methods such as exploratory data analysis should be carried out before assessing the stability in GEI studies. The effectiveness of the four graphical methods i) variance heterogeneity diagnostic plot (Box et al. 1978), ii) transformable non-additivity diagnostic plot, (Box et al. 1978) iii) Emerson and Hoaglin's (1983) non-additivity diagnostic plot, and iii) Gabriel's bi-plot (1971) in detecting non-crossover interactions resulting from variance heterogeneity and non-additivity are presented in this paper. Baker's (1990) three simulated and the spring wheat data sets were used to evaluate the effectiveness of these four graphical techniques.

Key words: crossover and non-crossover interactions; variance heterogeneity; transformable non-additivity diagnostic plot; bi-plot; exploratory data analysis.

1. Introductions

Genotype x Environment Interaction (GEI) is an important consideration in multi-environmental plant breeding trials, when the test genotypes exhibit inconsistent relative performance over environments.

Significant GEI results from the changes in the
magnitude of differences among genotypes or changes in
the relative ranking of the genotypes in different
environments. The GEI when present reduces the
progress from selection in any one environment.

A qualitative or crossover interaction arises when
the ranks of the true genotype mean vary among the
environments while the quantitative or non-crossover
interaction involves the variation in the magnitude of
genotype means among environments (Peto, 1982; Gail and
Simon, 1985). Haldane (1946) and Gregorius and Namkoong
(1986) claimed that only the crossover interaction is
practically important. Non-crossover interactions are
usually associated with variance heterogeneity and non-
additivity of the GE data matrix. Thus, differentiating
between the crossover and non-crossover design is
important in decision making relating to plant breeding
strategies, since the presence of crossover interaction
emphasizes the need for breeding programs for specific
adaptation to certain environments.

Several statistical methods are proposed to assess
the GEI or the environmental sensitivity of genotypes
tested under a wide range of environments (Fernandez,
1991). The most popular methods have used analysis of
variance combined with joint regression analysis to
determine the environmental sensitivity of the
genotypes (Finlay and Wilkinson, 1963; Eberhart and
Russell, 1966; Fernandez et al. 1989). When the
genotype mean in an environment is regressed on an
environmental index (environment mean - grand mean),
the regression coefficient and a function of squared
deviation from the regression would provide estimates
of environmental sensitivity statistics. With this
model, the sums of squares due to environments and
genotype x environments are partitioned into
environment-linear, GxE-linear, and deviation from the
regression model. Baker (1990) used the conventional
joint regression analysis on three simulated data sets
and concluded that this method failed to differentiate
between the crossover and non-crossover interactions.

For detecting crossover interaction, Baker (1990)
used the tests developed by Azzalini and Cox (1984) and
Gail and Simon (1985). Tedious computations involving
comparisons of all possible pairs of genotypes in all
possible pairs of environments are required in these
tests. For example, 3620 possible quadruple
comparisons are necessary for a 9 x 10 GE data matrix
(Baker, 1990). Therefore, differentiating the non-
crossover interaction caused by variance heterogeneity
and non-additivity from crossover interaction by simple
but effective methods such as exploratory data analysis

should be carried out before assessing the stability in GEI studies.

Exploratory data analysis is the "manipulation, summarization, and display of the data to make them more comprehensible to human minds, thus uncovering the underlying structure in the data and detecting important departure from this structure" (Andrews, 1978). Examining a plot of predicted values and the residuals is a widely used exploratory data analysis to detect the violation of assumptions in ANOVA and regression models (Fernandez, 1992). The feasibility of using some graphical techniques to detect the non-crossover interactions should be evaluated as alternative to the complex crossover detection tests.

The objectives of this study were to investigate the effectiveness of the four graphical methods i) variance heterogeneity diagnostic plot (Box et al. 1978), ii) transformable non-additivity diagnostic plot, (Box et al. 1978) iii) Emerson and Hoaglin's (1983) non-additivity diagnostic plot, and iv) Gabriel's bi-plot (1971) in detecting non-crossover interactions resulting from variance heterogeneity and non-additivity. Baker's (1990) three simulated and the spring wheat data sets were used to evaluate the effectiveness of these four graphical techniques.

2. Baker's (1990) simulated data sets:

2.1 Random effects model:

The statistical model, Baker (1990) used for generating simulated data was:

$$Y_{ij} = \mu + G_i + E_j + GE_{ij} \tag{1}$$

where G_i, E_j, and GE_{ij} represent random genotype, environmental, and GEI effects respectively. The data comprise of 8 genotypes evaluated in 10 environments (table 1). In this model, the changes in genotype rank result from differences in genotype response to unpredictable factors in the environment. This model should show no relationship between the GEI effects and the environmental deviation. However, joint regression analysis produced regression coefficients that vary from -.2 to .2 particularly for this small simulated data with eight genotypes. The range in regression coefficient decreases as the number of genotypes increases. In this simulated data, about 10% of the GEI variance was attributed to the heterogeneity of regression coefficients (Baker 1990).

2.2 Multiplicative model:

Baker (1990) used the following model to generate a simulated data with multiplicative effects:

$$Y_{ij} = \mu + (G_i - G_{min})(E_j - E_{min}) \qquad (2)$$

Where, Y_{ij} is the function of the product of random normal deviate G_i and E_j. This data also comprise of 8 genotypes evaluated in 10 environments (table 1). The important feature of this model is that there was no crossover interaction. This type of non-crossover interaction involves heterogeneity of variance, non-additivity, and high correlation among environments. Significant regression coefficients, insignificant mean square deviation from the joint regression analysis, and a positive correlation between genotype mean and the regression coefficients were characteristic of this data.

2.3 Yield-disease model:

The simulated data for this model were generated by the following statistical model (Baker 1990)

$$Y_{ij} = \mu + G_i + E_j - B_i(E_j - E_{min}) \qquad (3)$$

where B_i is 0 for resistant genotypes and 0.3 for susceptible genotypes. The disease incidence is assumed to be greater under environmental condition that favors higher yield. This data also comprise of 8 genotypes evaluated in 10 environments (table 1). As in the multiplicative model, the interaction was reflecting a non-additive model and there were no deviations from the joint regression analysis. The yield disease model differs from the multiplicative model in that crossover interactions occurred and differed from the random effects model in that the crossover interactions were related to specific causes, the degree of disease resistance.

2.4 Spring wheat yield data (Baker 1990):

This data comprise of nine genotypes grown in 10 environments (table 1). Significant crossover interaction was observed as in yield disease model due to differential disease resistance potentials.

3. Graphical techniques to detect non-crossover interaction:

3.1 Variance heterogeneity diagnostic plot (Box et al. 1978):

A plot between log of environmental standard deviation (S_j) and the log of environmental mean ($\overline{Y}_{.j}$) allows to judge the extent of variance heterogeneity and suggests which power transformation if any would help to remove it. A scatter plot with no consistent trend indicates the absence of variance heterogeneity and a linear trend suggests the presence of variance heterogeneity. If the regression is not significant when $\log(S_j)$ regressed on $\log(\overline{Y}_{.j})$, data transformation is usually not necessary. A significant regression (*P<0.05*) indicates the data should be transformed and the regression coefficient is estimated. The power (λ) of the transformation can be estimated by subtracting the regression coefficient (β) from 1. The value of the power, (λ) indicates the appropriate transformation. For example, if β approximately equals 2, then $\lambda = 1-\beta = -1$. Thus, the appropriate transformation would be reciprocals.

3.2 The transformable additivity diagnostic plot (Box et al. 1978):

In a two-way data structure, a value of the response is observed for each combination of genotype and environment. Where:

$$Y_{ij} = \mu + G_i + E_j + GE_{ij}. \tag{4}$$

First an additive model is fitted and the μ, G_i, and E_j are estimated. The transformable additivity diagnostic plot (Box et al. 1978) plot illustrates the relationship between the interaction component, GE_{ij} and $\overline{Y}_{ij}$, where

$$\overline{Y}_{ij} = \mu + G_i + E_j \tag{5}$$

$$GE_{ij} = Y_{ij} - \overline{Y}_{ij} \tag{6}$$

A scatter plot with no consistent trend indicates the absence of transformable non-additivity and a curvilinear plot suggests the existence of non-crossover interaction resulting from transformable non-additivity.

3.3 Emerson and Hoaglin's (1983) non-additivity diagnostic plot:

This plot allows the extent of any non-additivity to be judged and suggests which power transformation if any would help to remove it. The diagnostic plot illustrates the relationship between the interaction component, GE_{ij} and the comparison value (cv_{ij}), where:

$$cv_{ij} = (G_i \times E_j)/\mu \qquad\qquad (7)$$

A scatter plot with no consistent trend indicates that the data do not depart systematically from an additive model. Alternatively, if we see a linear plot, the slope of this diagnostic plot guides us to a transformation that should help to remove the non-additivity. If the slope is β, the power near $(1-\beta)$ provide a useful power transformation.

3.4 Gabriel's biplot (1971):

The GE_{ij} component is partitioned into a sum of multiplicative terms using principal component analysis:

$$Y_{ij} = \mu + G_i + E_j + \sum a_{iu} b_{ju} + \varepsilon_{ij} \qquad\qquad (8)$$

Let λ_u is the uth eigenvalue and b_{1u}, b_{2u}, . . is the uth eigenvector. When scaled so that $\sum_j b^2_{ju} = 1$, the b_{1u}, b_{2u}, .. represents the environmental loadings for the uth principal component while a_{1u}, a_{2u}, . . represent the genotype scores on that principal component axis, and $\lambda_u = \sum_i a^2_{iu}$ is the contribution to the GEI sums of squares arising from the uth component (Kempton, 1984). Thus, the biplot display of principal component analysis where two dimensional approximation to a two way table (rows x column) can be obtained from the first two principal components. For a table consisting of GEI components, the genotypes (row points) and environments (vector coordinates) can be plotted in the same graph (the biplot).

In all four data sets, number of environments (columns) were greater than the number of genotypes. To obtain unique principal components, the number of row points (observations) should be greater than the columns (variables). To satisfy this requirement and to be consistent with all four data analyses, certain environments were not included in all four graphical analyses (table 1).

4. Results and Discussions

The Box et al. (1978) variance heterogeneity diagnostic plots (Fig.1) revealed the presence of significant heterogeneous variances among environments in the multiplicative and yield disease models. A scatter plot with no consistent trend in the random effects model confirmed variance homogeneity. In addition, 1-slope of the linear regression model between the log S_j and Y_j can be useful in deciding the suitable power transformation that can be used to remove the variance heterogeneity. The suggested power transformations that can remove the variance heterogeneity were $Y^{-2.92}$, $Y^{-2.66}$, and Y^0 for the multiplicative, yield disease, and the spring wheat data respectively. However, the suggested power transformations based on the slope were not appropriate in the context of the data for the multiplicative and the yield disease models.

A curvilinear plot in the transformable additivity diagnostic plots (Box et al.1978) (Fig.2) confirmed the presence of non-crossover interaction in addition to variance heterogeneity in the multiplicative model. Scatter plots were observed in other three models confirmed the presence of crossover interactions.

Scatter plots were observed in Emerson and Hoaglin's non-additivity plots for the random effects and spring wheat data (Fig. 3) confirmed the absence of non-additivity in these models. A highly significant linear relationship in the multiplicative model confirmed the presence of non-additivity or the non-crossover interaction. The recommended power transformation to remove non-additivity based on the slope was -3.7. However, this is not an appropriate power transformation within the context of the data. In the yield disease model, a significant linear trend was observed. This could be attributed to the variance heterogeneity of the yield disease data. Thus, Emerson and Hoaglin's non-additivity plot failed to separate non-additivity and variance heterogeneity when they are present together.

The biplot displays points for genotypes and vector coordinates for environments on the same plot so that the expected response of a genotype in a particular environment may be derived from visual inspection of their relative positions on the biplot. The biplot will display a major portion the variation of the two-way GEI data matrix. The relative length of the vector line will represent the variance of the environments and their angles represent their correlations. The ranking of genotypes in a particular environment will be achieved by ordering the projected points.

In the multiplicative model, the presence of transformable non-additivity was revealed by the highly significant correlations (0.999) among the high yielding and the low yielding environments. About 99.9% variation in the data could be explained by the first dimension in the biplot display which separated the low yielding environments (Env1-Env4) from the high yielding environments (Env5-Env10)(Fig. 4).

In the yield disease model, the first dimension which explains 88% of the variation separated the low yielding environments (Env1-Env4) from the high yielding environments (Env5-Env10). The nature of the crossover interactions in the yield disease model was revealed by the pattern of grouping of similar environments and genotypes and the separation of dissimilar environments by the relative positions of the genotypes in the biplot. The relatively higher performance of disease resistant genotypes, B, D, and E in high yielding environments where the disease pressure is expected to be higher is an indication of crossover interaction.

In the spring wheat data, the first dimension which explains 76% of the variation separated the high yielding environments (Env3, Env6, and Env10) from the low yielding environments. The nature of the crossover interactions in this data was revealed by the position of interacting genotypes and environments in the biplot. Relatively higher yields of genotypes C and G in the high yielding environments and relatively lower yields of genotypes D and I in the low yielding environments clearly indicate the presence of crossover interaction in this data.

5. Summary

The variance heterogeneity diagnostic plot effectively detects the heterogeneity of variance. The transformable non-additivity could be separated from crossover interaction by the transformable non-additivity diagnostic plots (Box et al.1978) and the biplots. Non-crossover interaction attributed to both variance heterogeneity can be detected by the presence of significant linear plot in the Emerson and Hoaglin's non-additivity plots. In addition, the biplot technique is useful in showing inter-unit distances among the genotypes and environments and thus facilitate the clustering of genotypes as well as display variances and correlations among the environments. Thus, these exploratory graphical techniques are valuable tools in the GEI studies before estimating and interpreting the traditional stability statistics.

6. References

Andrews, D. 1978. Exploratory data analysis. In:
     International Encyclopedia of statistics. PP. 97-107.
     Krypton W. H. (eds.). N.Y. Free Press.
Azzalini, A. and D.R. Cox. 1984. Two new tests
     associated with analysis of variance. J. R. Stat.
     Soc. B. 46:335-343.
Baker, R.J. 1990. Crossover genotype-environmental
     interaction in spring wheat. In genotype by
     environment interaction and plant breeding. M. S.
     Kang (ed.) Louisiana Agricultural Experiment Station.
     LSU. Baton Rouge. Box, G.E.P., W.G. Hunter, and J.S.
Hunter. 1978. Statistics for Experimenters. John Wiley
     & Sons. N.Y.
Eberhart, S.A, and W.A. Russell. 1963. Stability
     parameters for comparing varieties. Crop Sci. 6: 36-
     40.
Emerson J.D. and D.C. Hoaglin.1983. Analysis of two-way
     tables by medians. In: Understanding Robust and
     exploratory data analysis. D.C. Hoaglin, F.
     Mosteller, and J.H. Tukey. (eds). John Wiley & Sons.
     Inc. N.Y.
Fernandez, G.C.J., H.K. Chen and J.C. Miller, Jr. 1989.
     Adaptation and environmental sensitivity of
     mungbean genotypes evaluated in the International
     Mungbean Nursery. Euphytica 41:253-261.
Fernandez, G.C.J. 1991. Analysis of genotype x
     environment interaction by stability estimates.
     HortSci. 26:947-950.
Fernandez, G.C.J. 1992. Residual analysis and
     transformations — Important tools in statistical
     analysis. HortSci. 27:(in press).
Finlay, K.W. and G.N. Wilkinson, 1963. The analysis of
     adaptation in plant breeding programs. Aust. J.
     Agric. Sci. 14 742-757.
Gabriel, K.R. 1971. Biplot display of multivariate
     means with application to principal component
     analysis. Biometrika 58:453-467.
Gail, M. and R. Simon. 1965. Testing for qualitative
     interactions between treatment effects and patient
     subsets. Biometrics 41:361-372.
Gregorius, H.R. and G. Namkoong. 1986. Joint analysis of
     genotype x environmental effects. Theor. Appl. Genet.
     72:413-422.
Haldane, J.B.S. 1946. The interaction of nature and
     nurture. Ann. Eugenics. 13:197-205.
Kempton, R.A. 1984. The use of biplots in interpreting
     variety by environment interactions. J. Agric. Sci.
     103:123-135.
Peto, R. 1982. statistical aspects of cancer trials.
     PP:867-871. In: E.E. Halnan. (ed.) Treatment of
     cancer. Chapman and hall. London . U.K.

Table 1. Baker's (1990) data sets.

1. Random effects model

Gen.	E1	E2	E3^z	E4	E5	E6^z	E7	E8^z	E9	E10
A	170	173	180	184	202	193	171	218	207	235
B	154	190	172	191	187	192	192	209	211	241
C	185	183	181	214	197	211	203	213	214	239
D	177	195	194	208	203	192	216	213	201	250
E	190	204	210	193	199	211	204	239	235	262
F	192	176	206	218	213	213	195	221	244	251
G	170	204	196	214	221	213	214	240	243	251
H	202	228	209	218	235	239	246	254	263	285

2. Multiplicative model

Gen.	E1	E2^z	E3	E4^z	E5	E6	E7	E8	E9^z	E10
A	200	200	200	200	200	200	200	200	200	200
B	200	203	204	206	207	207	209	211	213	215
C	200	215	227	236	241	244	256	268	279	291
D	200	216	229	239	245	248	261	273	285	299
E	200	220	236	248	255	258	275	290	304	321
F	200	232	257	278	289	294	321	345	368	395
G	200	236	263	286	297	303	333	359	385	415
H	200	236	264	287	299	305	335	362	388	419

3. Yield-disease model

Gen.	E1	E2^z	E3	E4^z	E5	E6	E7	E8	E9^z	E10
A	163	168	169	170	177	177	179	181	188	189
B	164	174	175	176	191	192	195	200	214	215
C	172	177	177	178	186	186	188	190	197	198
D	176	186	187	188	203	204	207	212	226	228
E	177	186	187	189	204	204	208	212	226	228
F	178	183	184	184	192	192	194	196	203	204
G	181	186	186	187	194	194	196	199	206	206
H	181	186	187	187	195	195	197	199	206	207

4. Spring wheat (9) x Environment (10) — grain yield

Gen.	E1	E2	E3	E4	E5^z	E6^z	E7	E8	E9	E10
A	294	348	428	432	485	478	213	317	204	354
B	325	404	421	452	548	448	240	345	226	355
C	305	314	427	463	486	506	214	316	196	353
D	332	380	402	403	424	376	214	326	219	349
E	284	290	361	396	418	386	194	273	189	320
F	270	289	344	365	422	387	169	270	175	306
G	227	324	399	424	450	464	175	272	177	343
H	307	285	373	410	453	396	204	282	184	326
I	260	314	409	316	351	364	176	291	183	328

^{z} Yield data from these environments were not analyzed.
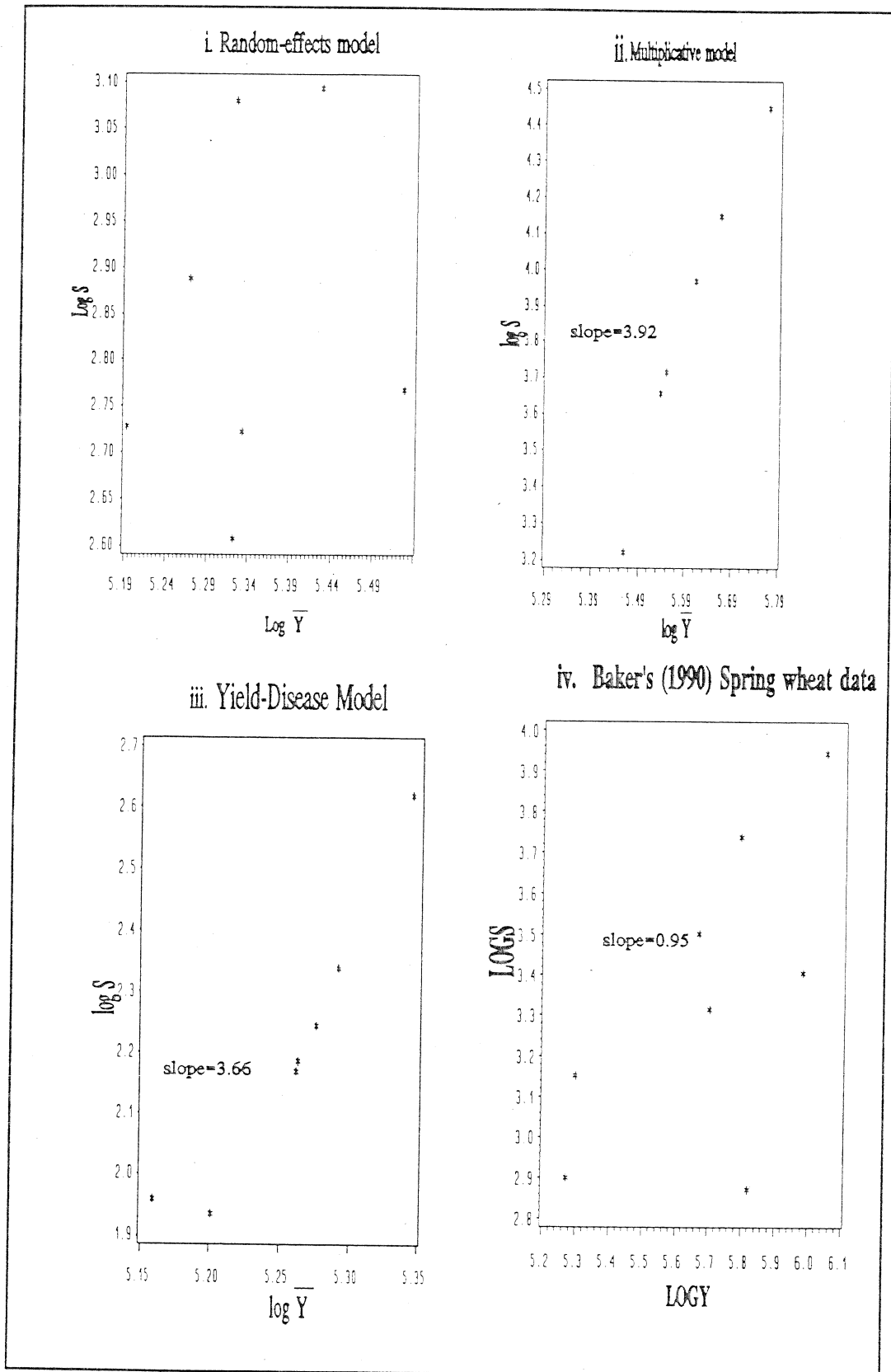
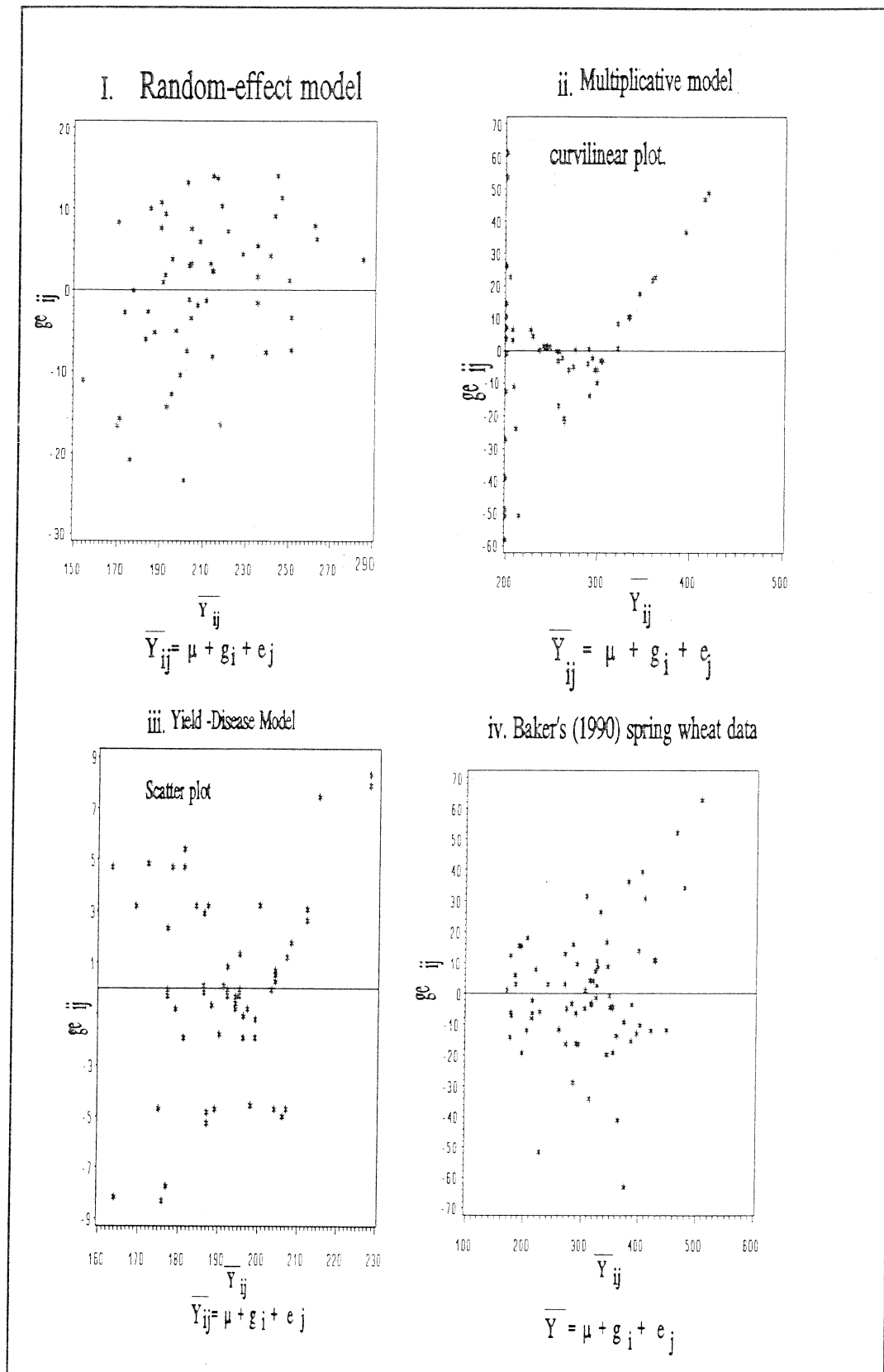Fig.1 Variance heterogeneity diagnostic plots (Box et al. 1978)

Fig.2 The transformable additivity diagnostic plots (Box
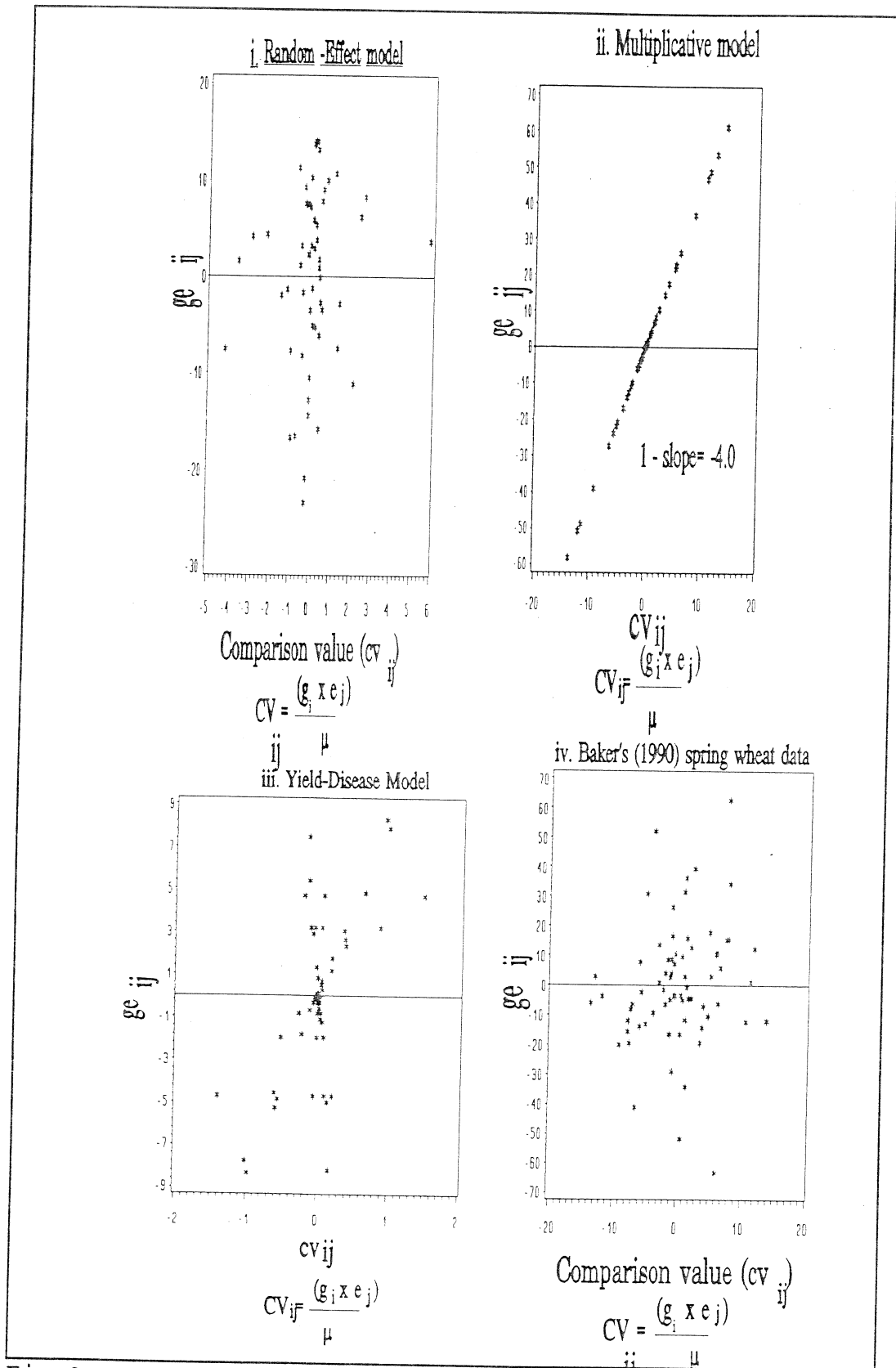   et al. 1978)

Fig.3 Emerson and Hoaglin's (1983) non-additivity diagnostic plots

i. Random-Effects Model

ii. Multiplicative model
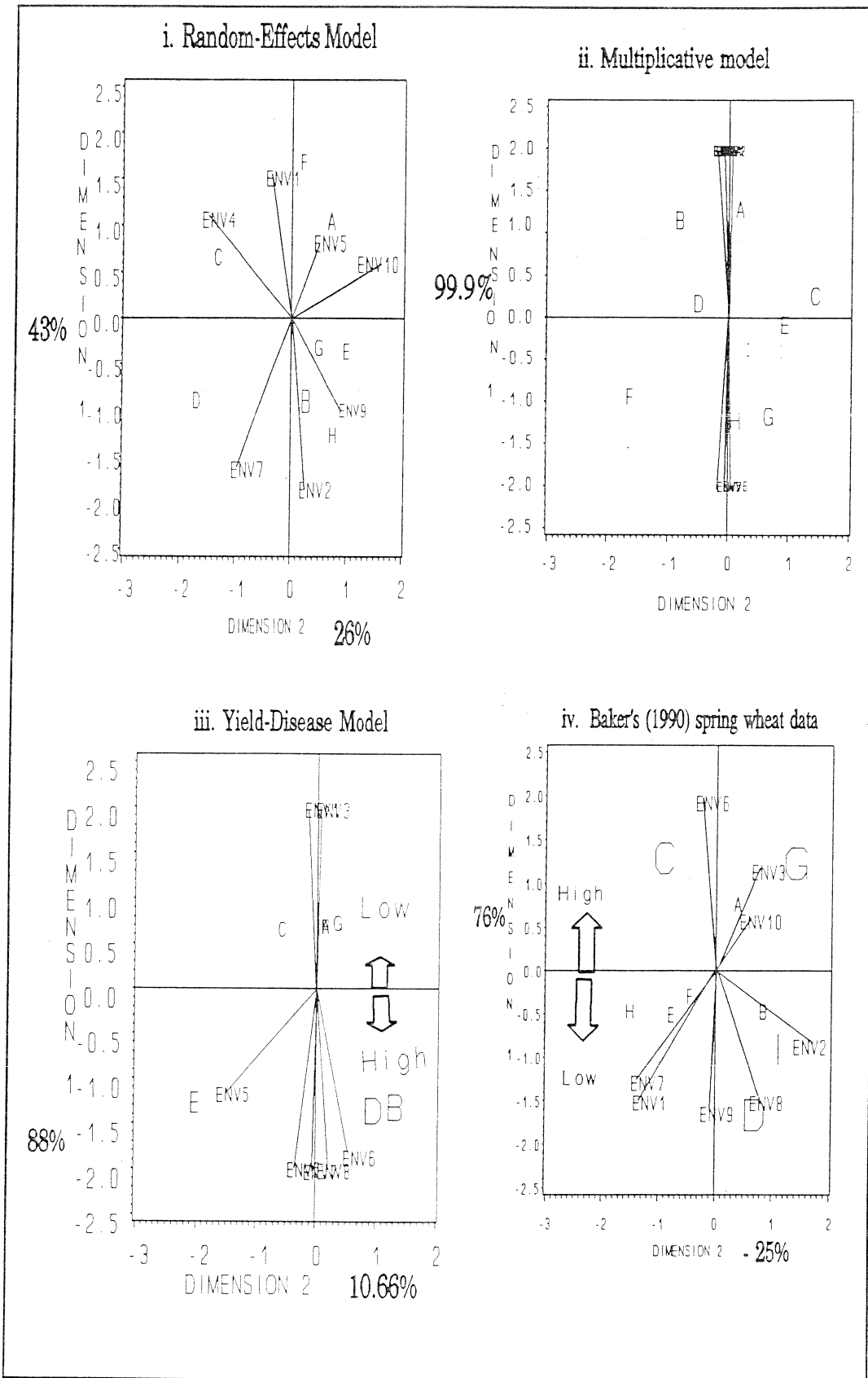
iii. Yield-Disease Model

iv. Baker's (1990) spring wheat data

Fig.4   Gabriel's (1971) biplot representation of GEI.