Theor Appl Genet (1992) 84: 735–738

THEORETICAL AND APPLIED GENETICS

Trimmed joint regression: a new approach to the joint regression analysis for cultivar relative-performance evaluation

L. Gusmão[1], **J.T. Mexia**[2], **and J. Baeta**[2]

[1] Departamento de Genética, Estação Agronómica Nacional, 2780 Oeiras, Portugal
[2] Faculdade de Ciências e Tecnologida, Universidade Nova de Lisboa, 2825 Monte da Caparica, Portugal

Received October 21, 1991; Accepted February 26, 1992
Communicated by J. Mac Key

Summary. A new approach to joint regression analysis, entitled Trimmed Joint Regression (TJR), is proposed in which the adjustment of the linear relative-yield pattern of the cultivars is trimmed from the residues which can be attributed to a "specific interaction". The ranking of the residues to the joint regressions for each cultivar, within each trial, is analysed by the Friedman test to ascertain if they belong to the same parental distribution of the population of residues, for the different genotypes. The rejection of the null hypothesis is envisaged as the result of an organized pattern of the residues, due to a "specific interaction", and the genotypes responsible for such interactions are identified. The trimming method consists of the assessment of the linear regressions, after a reassessment of data related with "specific interactions". The increased accuracy which can be achieved is shown in a numerical example where the high repeatability of the method is illustrated by means of a comparison of the estimated yields for 1- and 2-year trials.

Key words: "Genotype × environment" interaction – Joint regression analysis – Specific interaction – Specific instability X *Triticosecale* Wittmack

Introduction

As stressed by Simmonds (1981) "yields from trials are often so much higher than agricultural yields that the consistency of extrapolating GE effects from one to the other may be doubtful". On the other hand, in some instances, the differences between some production environments within one farm may be so extreme that, as pointed out by Atlin and Frey (1990), "substantially different gene complements are required to achieve the highest yields in each".

The linear regression method, initially suggested by Mooers (1921) and subsequently developed by Finlay and Wilkinson (1963), seemed to be adequate in this instance since, besides the assessment of "Genotypic Value" (Pinthus 1973), it allows for the evaluation of relative cultivar behaviour at different environmental indices.

However, as pointed out by Becker and Léon (1988), in the joint regression analysis "only a minor part of the GE interactions can be attributed to known environmental determinants" and, as commented by Zobel et al. (1988), "linear regression (LR) analysis is able to effectively analyse interaction terms only where the pattern fits a specific regression model". The required conditions which we have already advanced in order to make this method a robust one (Gusmão 1985; Gusmão et al. 1989) may ensure a better agreement of the data with a specific regression model, which increases the precision of the method. Nevertheless, the joint regression still leaves unexplained that part of the interaction which, after Eberhart and Russell (1966), has been ascribed to "specific instability".

In the present work, we found that the specific instability can be further analysed for the detection of "specific interactions". In view of this, we suggest a new approach to the Joint Regression Analysis (JRA), where the relative yield pattern of cultivars is estimated after reassessment of data related to "specific interactions".

Statistical approach

Assessment of specific interactions

In a network of j yield trials, designed in k randomized complete blocks, within the same equipotential zone for yield pattern evaluation, for the same cohort of i genotypes, we start by obtaining the ijk residues for the JRA. For each trial and each cultivar there is, then, a sample of k residues (one per block).

To test the hypothesis of equal parental distribution for the residues, pertaining to the different cultivars within a trial, we use the non-parametric Friedman test, since the specific interaction could render the usual assumption of normality rather disputable.

When the hypothesis of equal parental distributions is rejected we eliminate the cultivar for which the sum of ranks lies furthest away from the corresponding mean, and repeat the test with the remaining cultivars.

In this way it is possible to identify the specific combinations of genotype and trial which are responsible for an interaction. With further knowledge of both genotype and environmental features it may be possible to identify the particular genetic system and environmental factor responsible for that specific interaction.

Trimming method

The trimming method consists of the following iterating procedure:

(1) Adjustment of the joint regressions – the first adjustment of the joint regressions is based on the block to block results (Gusmão 1985).
(2) Identification of the situations of specific interaction – we proceed with the identification of situations of specific interaction as described above.
(3) Adjustment of the first iteration regressions – the first iteration regressions are then evaluated, based on a reduced data matrix, built up by exclusion of trials where specific interaction was detected.
(4) Assessment of the corrected data matrix – the first iteration regressions are used to evaluate the corrected data matrix, according to the procedure described below.
(5) Adjustment of the TJRs – we finally use the corrected data matrix, in which we have replaced the yields corresponding to specific interactions by the corresponding estimated values, to adjust the TJRs

Assessment of the corrected data matrix

Let α_i^* and β_i^* ($i=1, \dots, l$) be the estimates of the coefficients of the first iteration regressions, the true values of these coefficients being α_i and β_i ($i=1, \dots, l$); these estimates are obtained from the reduced data matrix and will be used to calculated the corrected values for situations of "specific interaction", according to the procedure we describe below.

Given a block in which a "specific interaction" has occurred, let C be the set of indices for the genotypes without "specific interaction" in that block. Let $\#(C)$ be the number of indices in C; we can take

$$\alpha^*(C)=\frac{1}{\#(C)}\sum_{i\in C}\alpha_i^* \quad \text{and} \quad \beta^*(C)=\frac{1}{\#(C)}\sum_{i\in C}\beta_i^*$$

to estimate

$$\alpha(C)=\frac{1}{\#(C)}\sum_{i\in C}\alpha_i \quad \text{and} \quad \beta(C)=\frac{1}{\#(C)}\sum_{i\in C}\beta_i$$

The occurrence of a "specific interaction" precludes the use of the average yield of all the cultivars in the block as a measure of the corresponding environmental index (x). To estimate x we point out that if $\bar{y}(C)$ is the average yield for the cultivars without "specific interaction", then we should have

$$\bar{y}(C)\approx\alpha^*(C)+\beta^*(C)x$$

so that for the environmental index (x) we will have the estimate

$$x^*=\frac{\bar{y}(C)-\alpha^*(C)}{\beta^*(C)}$$

In the blocks where "specific interactions" occurred, the corrected values for the cultivars with "specific interaction" ($i\notin C$) will now be

$$y_i^*=\alpha_i^*+\beta_i^*x^*;\ i\notin C$$

These values will be used in the corrected data matrix.

Numerical example

As a numerical example we use the results of a cohort of eight triticale genotypes in 20 yield trials.

The triticale genotypes were cultivars 'Bacum', 'Arabian' 'Juanilho' and 'Borba' and advanced lines TTE8701, TTE8702, TTE8703 and TTE8704.

The design for each field trial had four randomized complete blocks.

Trials were carried out in the centre and south of Portugal, within the same "equipotential zone for relative

Table 1. Specific genotype-environment interaction and significance level (S.L. %) as assessed by the Friedman test of the residues to the joint regression lines

Trial	S.L. %	After exclusion of	
		Genotype	S.L. %
Abrantes/88	99.0	'Arabian'	95.0
		TTE8702	95.0
		TTE8701	n.s.
Beja/88	99.9	'Bacum'	99.0
		'Arabian'	95.0
		TTE8701	95.0
		TTE8702	n.s.
Santarém/89	99.0	'Arabian'	95.0
		TTE8702	95.0
		'Bacum'	n.s.
Almeirim/89	95.0	'Arabian'	n.s.

Table 2. Estimated parameters, α^* and β^*, to the joint regressions ($y=\alpha+\beta x$), assessed by the established procedure (JRA) and by the trimming procedure (TJR), based on the trial results obtained in 1 year (1987/88) and in 2 years (1987/88+1988/89) in a triticale yield trials network in the center and south of Portugal

Genotype	1987/88				1987/88+1988/89			
	JRA		TJR		JRA		TJR	
	α^*	β^*	α^*	β^*	α^*	β^*	α^*	β^*
'Bacum'	−439	1.02	−484	1.01	−578	1.04	−471	1.01
'Arabian'	−558	1.05	−657	1.04	−763	1.03	−596	1.00
'Juanilho'	155	0.97	151	0.97	345	0.97	275	0.97
'Borba'	230	1.00	200	0.99	221	0.99	185	1.00
TTE8701	137	0.97	190	0.98	186	1.00	121	1.01
TTE8702	252	0.91	339	0.92	262	0.95	217	0.97
TTE8703	102	1.09	103	1.09	80	1.06	80	1.06
TTE8704	120	1.00	158	1.00	245	0.96	189	0.98

Table 3. Comparison of the predictions for three environmental indices (1,500, 3,500 and 5,500 kg/ha), based on the joint regressions estimated, by the established procedure, from the results of the first year trial (1987/88) and from the results of the 2-year trials (1987/88 + 1988/89)

Genotype	Environmental index											
	1,500 kg/ha				3,500 kg/ha				5,500 kg/ha			
	Yield (kg)		Difference		Yield (kg)		Difference		Yield (kg)		Difference	
	a	b	kg	%[c]	a	b	kg	%[c]	a	b	kg	%[c]
'Bacum'	1,091	982	−109	**7**	3,131	3,062	− 69	**1**	5,171	5,142	− 29	**0**
'Arabian'	1,017	782	−235	**15**	3,117	2,842	−275	**7**	5,217	4,902	−315	**5**
'Juanilho'	1,610	1,800	190	**12**	3,550	3,740	190	**5**	5,490	5,680	190	**3**
'Borba'	1,730	1,706	− 24	**1**	3,730	3,686	− 44	**1**	5,730	5,666	− 64	**1**
TTE8701	1,592	1,686	94	**6**	3,532	3,686	154	**4**	5,472	5,686	214	**3**
TTE8702	1,617	1,687	70	**4**	3,437	3,587	150	**4**	5,257	5,487	230	**4**
TTE8703	1,737	1,670	− 67	**4**	3,917	3,790	−127	**3**	6,097	5,910	−187	**3**
TTE8704	1,620	1,685	65	**4**	3,620	3,605	− 15	**0**	5,620	5,525	− 95	**1**

[a] Based on the experimental results in 1987/88
[b] Based on the experimental results in 1987/88 + 1988/89
[c] In relation to the assumed environmental index

Table 4. Comparison of the predictions for three environmental indices (1,500, 3,500 and 5,500 kg/ha), based on the TJRs estimated from the results of the first year trials (1987/88) and from the results of the 2-year trials (1987/88 + 1988/89).

Genotype	Environmental index											
	1,500 kg/ha				3,500 kg/ha				5,500 kg/ha			
	Yield (kg)		Difference		Yield (kg)		Difference		Yield (kg)		Difference	
	a	b	kg	%[c]	a	b	kg	%[c]	a	b	kg	%[c]
'Bacum'	1,031	1,044	13	**1**	3,051	3,064	13	**0**	5,171	5,084	13	**0**
'Arabian'	903	904	1	**0**	2,983	2,904	− 79	**2**	5,063	4,904	−159	**3**
'Juanilho'	1,606	1,730	124	**8**	3,546	3,670	124	**4**	5,486	5,610	124	**2**
'Borba'	1,685	1,685	0	**0**	3,665	3,685	20	**1**	5,645	5,685	40	**1**
TTE8701	1,660	1,636	− 24	**2**	3,620	3,656	36	**1**	5,580	5,676	96	**2**
TTE8702	1,719	1,672	− 47	**3**	3,559	3,612	53	**2**	5,399	5,552	153	**3**
TTE8703	1,738	1,670	− 68	**5**	3,918	3,790	−128	**4**	6,098	5,910	−188	**3**
TTE8704	1,658	1,659	1	**0**	3,658	3,619	− 39	**1**	5,658	5,579	− 79	**1**

[a] Based on the experimental results in 1987/88
[b] Based on the experimental results in 1987/88 + 1988/89
[c] In relation to the assumed environmental index

yield pattern evaluation" (Gusmão et al. 1989), by the Department of Cereals of the National Plant Breeding Station, and were included in the network of adaptation trials for 1987/88 and 1988/89.

Each year, a total of two trials where "specific interaction" occurs were identified, as can be seen in Table 1.

In Table 2 are presented the results for the joint regressions, in the first year and in the two years, using the established procedure and the TJR method.

The increase in the repeatability of the yield estimation, from the TJR method to the established joint regression method, more particularly in the genotypes for which "specific interaction" was detected (as is the case for 'Arabian', TTE8702, TTE8701 and 'Bacum'), can be seen by comparing the values in Tables 3 and 4, where the estimates at different environmental indices for the first year and for the two years are included.

Discussion

Specific instability can be regarded as the result of specific genotype-environment interactions which may not show clearly within a particular set of trials. On the other hand, in our recent experience (Mexia et al. 1990a), we verified that the residues to the regressions, once trimmed from the outliers, are not significantly heterocedastic, and no significant departures from the Gauss-Markov assumptions were detected (Gusmão et al. 1990; Mexia et al. 1990b). We may reason, as did Belsley et al. (1980), that "unusual influential data points, of course, are not necessarily bad data points; they may contain some of the most interesting sample information".

In the present data it can be seen that some residues showed organized patterns, specifically for individual genotypes within particular trials. We may thus infer that these residues are related with a specific interaction that make them behave as outliers to the joint regression.

It is obvious that some genotypes were more susceptible to specific interaction. However, within the same equipotential zone for relative yield pattern evaluation, years and sites do not show a clear differentiation in its manifestation.

Further analysis of the causes of specific interaction, should be sought in the particular genotype-environment combination where it occurs. For the same genotype, the frequency of this phenomenon will depend upon its particular nature, and may vary with the year (within the same site) or site (within the same year).

Since the genotype-environment combination which is responsible for a specific interaction can be recognized, we may say that there are no bad years or site for relative yield evaluation, and particular environments where specific interactions are bound to occur can yield additional information on the genotypic value of the cultivars. We may, even, in some trials introduce possible factors of specific interaction, either biotic or abiotic, in order to ascertain if they interfere with the relative pattern response.

A genotypic susceptibility or tolerance/resistance to a particular environmental factor is here envisaged as a non-conformity with the relative yield pattern, assessed by the joint regression method, under such an environmental factor, even if such a genotype does not correspond, on average, to the lowest or the highest ranking cultivar. When testing for susceptibility or tolerance/resistance, it cannot be concluded that susceptibility or tolerance/resistance, assessed by the differential response to the environmental variable factor, is different from the differential response to the environmental index, simply because the genotype ranks lower or higher than the other genotypes.

Since the outliers can be attributed to a well-defined interaction (being related to the agronomic concept of stability; Becker 1981), they should not be included in the evaluation of the yield pattern (which is related with the biological concept of stability; Becker 1981).

The trimming of the joint linear regression is, thus, a legitimate procedure, able to bring accuracy and robustness to the method, and giving a more comprehensive explanation of genotype-environment interaction.

This is well illustrated in Tables 3 and 4 where we included the estimates at different environmental indices for JRA and TJR, respectively, based both on 1 year and 2 years. The increasing accuracy, from JRA to TJR is expressed in the repeatability of the estimated yields for the second method, particularly in respect of the genotypes for which specific interaction was detected, as is the case for 'Arabian', TTE8702, TTE8701 and 'Bacum'.

As far as the assumptions for the correct use of the joint regression methods are concerned, we may expect, as in the present example, that TJR in 1 year gives an identical estimation as in 2 years, with the further advantage of the possible definition of some "specific interactions", whenever factors of interaction are included in the trial network.

Acknowledgements. We are grateful to Prof. M. Mota, for the critical revision of the manuscript, and to our colleagues F. Bagulho, B. Maçãs and J. Coutinho who kindly made available the results used in this paper.

References

Atlin GN, Frey KJ (1990) Selecting oat lines for yield in low-productivity environments. Crop Sci 30:556–561

Becker HC (1981) Correlations among some statistical measures of phenotypic stability. Euphytica 30:835–840

Becker HC, Léon J (1988) Stability analysis in plant breeding. Plant Breed 101:1–23

Belsley DA, Kuh E and Welsch RE (1980) Regression diagnostics: Identifying influential data and sources of colinearity. John Wiley and Sons, New York

Eberhart SA, Russell WA (1966) Stability parameters for comparing varieties. Crop Sci 6:36–40

Finlay KW, Wilkinson GW (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14:742–754

Gusmão L (1985) An adequate design for regression analysis of yield trials. Theor Appl Genet 71:314–319

Gusmão L, Mexia JT, Gomes ML (1989) Mapping of equipotential zones for cultivar yield pattern evaluation. Plant Breed 103:293–298

Gusmão L, Baeta J, Mexia JT (1990) Ad análise de resíduos extremos para a regressão conjunta na identificação de situações de "interacção especifica". XXV Jornadas de Genética Luso Españolas. Programa y Resumenes de las comunicaciones, Alcalá de Henares, p 74

Mexia JT, Gusmão L and Baeta J (1990a) Homogeneidade numa zona equipotencial de adaptação e suas implicações no delineamento de ensaios de produção de cultivares de cereais. Garcia de Orta 16:65–70

Mexia JT, Gusmão L and Baeta J (1990b) Regressão conjunta rectificada. XXV Jornadas de genética Luso Españolas. Programa y resumenes de las comunicaciones. Alcalá de Henares, p 101

Mooers CA (1921). The agronomic placement of varieties. J Amer Soc Agron 13:337–352

Pinthus MJ (1973) Estimate of genotypic value: a proposed method. Euphytica 22:121–123

Simmonds NW (1981) Genotype (G), environment (E) and GE components of crop yields. Expl Agric 17:355–362

Zobel RW, Wright MJ, Gauch HGJr (1988) Statistical analysis of a yield trial. Agron J 80:388–393