A GxE analysis of Triticale in Spain

1.1 Description of the data set

A series of trials were conducted in 1989 and 1990 to assess yield of 16 Triticale genotypes in 10 locations in Spain. Each experiment was designed as a randomized complete block design (RCBD) with four blocks per site. In this practical we use the adjusted means per genotype (and not the original raw data) as obtained from the 1989 trials. Triticale (2x=AABBRR) is an inter-specific hybrid between Triticum turgidum (2x=AABB) and Secale cereale (2x=RR). In addition, substituted triticales are a variation obtained by replacing the Rye chromosome 2R, by chromosome 2D of Triticum aestivum. Ten of the 16 genotypes used in this research were substituted lines, and the remaining six were complete lines. The locations used in this research differed in their conditions in a number of environmental characteristics, some of which have been included in the data set (soil pH, altitude, and the rainfall). You can find further information in the original publication by Royo et al. (1993).

Open the data file Triticale_ data.xls in Excel (or R if you prefer).

1. We start by obtaining simple summary statistics and plots to explore the GxE in this data set. Obtain tables of means and variances per environment (site), maybe boxplots, and correlations between environments. (You can do this in Excel, or R, whatever your prefer. If you use Excel, you may want to use the file Triticale_data_reorganised.xlsx). Based on the results answer the following questions:
   a) Rank environments from best to worst.
   b) What do you think regarding the variance, is it homogeneous or hetero-geneous? Do you see any relationship between the mean and the variance?
   c) What about the correlations between environments? Are there high, inter-mediate or low, are they positive or negative? What are the implications in terms of GxE?
   d) After your initial assessment, what is your general impression in terms of the GxE in this data set. Do you think it is important or not? Explain.

2. **The Finlay-Wilkinson model**

   A common way to represent GxE is by reaction norm curves. Non-parallel reaction norms are indicative of genotype by environment interaction, the most extreme case being when the reaction norms cross each other (cross-over interaction). A classic example of reaction norms is the Finlay-Wilkinson (FW) regression model (Finlay and Wilkinson, 1963; Yates and Cochran, 1938). In the FW regression model, the environments are characterised by an environmental index that reflects the quality of the environment. The environmental index can be defined as the environmental main effect, E, and can be included in the model as regressor [Equation 1)].

$$y_{ij} = G_i + E_j + \beta_i E_j + \varepsilon_{ij} \qquad (1)$$

   The intercept G_i corresponds to the genotypic main effect, and the slope β_i is a sensitivity parameter that describes how steep the reaction norm is for the particular genotype *i*. A steeper reaction norm is indicative of a higher sensitivity to the environment. The average slope in the whole set of genotypes is $\beta_i = 0$ (interpreted as average sensitivity), so $\beta_i > 0$ is interpreted as

above average sensitivity, and $\beta_i < 0$ as below average sensitivity. If you drop the E_j-term from the model, then the mean beta becomes 1.

The first step is to obtain the main effects of the environments, *E*. For this purpose, fit a model with a main effect for both G and E (no need to fit a variance structure yet),

$$y_{ij} = \mu + G_i + E_j + e_{ij}$$

You can use ASReml file Triticale_main_G_E.as, or R if you prefer. Next, find the solutions for E (from the .sln file if you use ASReml) and add a column to the data set containing the E-values (centre the E-values around zero, by subtracting the mean). The second step is to fit FW-regression according to Equation 1 above. You can use ASReml file Triticale_FW.as, or R if you prefer. Then, from the results, answer the following questions:

a. Inspect the P-values in the output.
   i) Does the FW-model significantly explain the differential genotypic reactions to environmental changes?
   ii) Do the mean β and the P-value change when you fit a model $y_{ij}= G_i + \beta_i E_j + \varepsilon_{ij}$ instead of Equation 1?
   iii) How much of the GxE is explained by the FW model? (Look at the residual variance)
   iv) Why is it important to assess the amount of explained variance by the FW model?

b) Give the parameter estimates of genotype C 1, and give your conclusions in terms of the general adaptation (mean performance across environments), and adaptability for this genotype. (For ease of interpretation, you may remove the intercept and the main-E effect from the model).

c) Compare the relative performance of genotypes C4 and S1 based on their FW parameter estimates. Which of the two do you expect to be more suitable for a high quality environment? Explain. (For ease of interpretation, you may remove the intercept and the main-E effect from the model).

d) Plot the reaction norm of C4 and S1 over the range of the environmental main effect found in the data, and also plot the data points in the same figure. (You can check your answer to question c from this plot). Calculate stability type 3 for both genotypes, as the standard deviation of the residuals. Which genotype has the better stability?

e) Compare the general adaptation and adaptability of complete ('C') genotypes versus substituted ('S') genotypes. Conclude whether their relative advantages depend on the quality of the environment.

3. Factorial regression on specific environmental parameters.

a. Using linear regression, investigate whether pH, altitude or rainfall explain a significant amount of GxE-interaction variance (judge based on residual variance of the model; fit one factor at a time, at least initially). Are there advantages of using specific environmental factors, rather than FW-regression? (You may use Triticale_FR.as as a start).

4. AMMI model and biplots

The AMMI model is an extension of the FW model, in which more than one environmental indexes are included. In addition, the environmental indexes are optimal in the sense that they maximise the explained variance (are 'latent' variables from a principal components

analysis).The first principal component explain most of the GxE, the second one is the second best, etc. The AMMI model can be written as:

$$\underline{y}_{ij} = \mu + G_i + E_j + \sum_{k=1}^{K} \lambda_k u_i v_j + \underline{\varepsilon}_{ij}$$

Note that GxE is explained by K principal components, and that that parameters are either related with genotypes (genotype scores or sensitivities ui), or with the environments (environmental scores vj). The λ_k are the corresponding eigenvalues. An added value of this model is that genotypic scores and environmental scores can be visualized together in a biplot, which reveals the structure in the GxE, both with respect to environments, genotypes and their combination.

To save time and struggle ☺, results of the AMMI model are already given below (From GenSTAT software). Based on these results, answer the following questions:

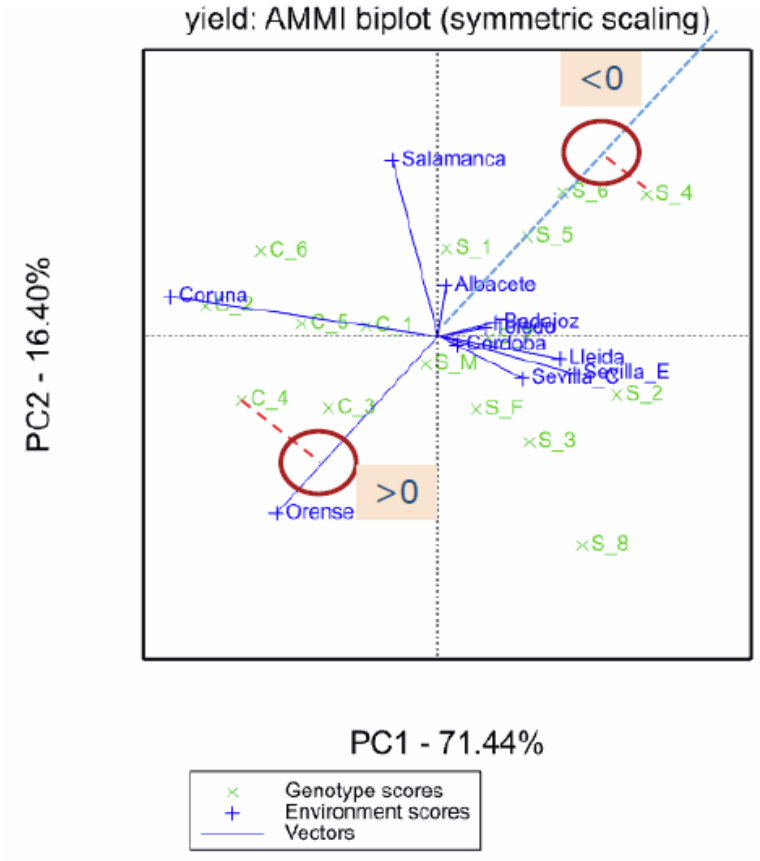
ANOVA table for AMMI model

Source	d.f.	s.s.	m.s.	v.r.	F pr
Genotypes	15	21.6	1.440	3.18	<0.001
Environments	9	480.8	53.426	117.93	<0.001
Interactions	135	61.2	0.453		
IPCA 1	23	43.7	1.900	23.26	<0.001
IPCA 2	21	10.0	0.478	5.85	<0.001
Residuals	91	7.4	0.082		

a) Inspect the ANOVA table.

i) Do the first two principal components (PC) explain a significant part of the GxE?

ii) Compare the explained variance of the AMMI model with that of the FW model.



b) Inspect the AMMI biplot. In the biplot, both genotypes and environments are displayed based on their corresponding scores. Environments are displayed as vectors, genotypes as symbols.

i) Do you recognize patterns of distribution of the genotypes in the biplot? Hint: recall that genotypes labelled with a prefix C are 'Complete', and those with S are 'Substituted'.

ii) Based on what you observe in the biplot, what type of interaction (positive/negative) does C4 show with environment Orense? And what about S4 and Orense? And what about C4 and S4 with Salamanca? To help you get started, projections are shown for Orense. Do projections for Salamanca yourselves. Explain both the direction and the strength of the interaction.

iii) Identify one or two genotypes that seems to have a high positive interaction with Sevilla and Lleida.

c) Is a higher positive interaction (relative better adaptation) necessarily indicative of a higher performance of the genotype in that environment? Explain why or why not.