

A GxE analysis of Triticale in Spain - 2: modelling the VCOV

Answers.

1. Raw data: The variance is very heterogeneous between environments, and the correlations between the performances of a genotype in different environments deviate greatly from 1. Hence, we need a model that allows for heterogeneous variances and a varying, and non-unity correlation, between environments.
2. CS: *a.* The log-likelihood is $\text{LogL} = -38.1547$ and $\text{AIC} = 80.31$. The genetic correlation between environments is 1 in the CS-model, as there is only one genetic effect that is the same in all environments. *b.* The phenotypic correlation equals $r_p = \sigma_G^2 / (\sigma_G^2 + \sigma^2) = 0.18$. This is a quite low value; so the phenotypic performance of a genotype in one environment does not predict performance in another environment accurately.
3. CSH: *a.* The log-likelihood is 27.5001, and $\text{AIC} = -33.00$. This is much lower than for the CS-model, so this model is clearly better. *b)* The AIC of this model is much better than the previous, so this is evidence that the residual variance is heterogeneous. Also, from the asr-file, you see that estimated residual variances are very different, and from the SE you can eyeball that those differences are significant. *c.)* As there is only a single genotypic effect, that is the same for all environments, the genetic correlation is 1. The phenotypic correlation between Coruna (E3) and Cordoba (E4) equals $r_p = 0.01327 / \sqrt{[(0.01327 + 2.195)(0.01327 + 0.1253)]} = 0.024$. So phenotypic performances of genotypes in Coruna and Cordoba show hardly any similarity based on this model.
4. DIAG: *a.* The log-likelihood is -44.2672, and $\text{AIC} = 90.53$. Hence, this model has very poor fit. *b.* There is no genetic, nor phenotypic correlation between environments (i.e. both are assumed zero).
5. DIAGH: *a.* The log-likelihood is 23.1355, and $\text{AIC} = -26.27$. *b.* Hence, this model is much better than DIAG and CS, but not as good as CSH. The change in AIC due to fitting heterogeneous variances (e.g. CSH vs CS, or DIAGH vs DIAG) is much larger than that due to fitting a correlation (e.g. CS vs DIAG or CSH vs DIAGH). So fitting the variance correctly is most important.
6. FULL: *a.* The model has 55 parameters, 10 variances, and $(10^2 - 10)/2 = 45$ covariances. This is a lot. Hence, fitting this model may be a challenge, and the model may overfit the data. *b.)* The log-likelihood is 86.4310, $\text{AIC} = -62.86$. *c.)* Hence, this is the best model so far, but it has very many parameters, we jumped from 11 parameters (CSH) to 55; maybe there is a better model somewhere in-between. *d)* To judge the degree of GxE-interaction, you should look at the estimated correlations (and ideally at their SE, which you can get if you make a PIN-file). Quite a number of the estimated correlations are negative, suggesting substantial GxE.
7. FA: *a.)* the FA-model has two parameters for each environment, one for the variance (ψ) and one for the covariance (λ), so the total number of parameters is $2n = 20$, n being the number of environments (here). Hence, this model is considerably simpler than the full model. *b.)* The log-likelihood = 55.3732 and $\text{AIC} = -70.75$. As expected, the likelihood is lower than from the full model (because the FA-model fits fewer parameters), but the fit judged by AIC is better than the full model. *c.)* variance in Coruna: $1.58452^2 - 0.103880 = 2.407$. Covariance Coruna and Cordoba = $1.58452 \times 0.243047 = 0.385$. Variance in Cordoba: $0.243047^2 + 0.107050 = 0.166$. Correlation between Coruna and Cordoba = $0.385 / \sqrt{(2.407 \times 0.166)} = 0.6091$. These values agree with the (co)variance/correlation matrix given by Asreml. *d.)* Below is the matrix

of the differences between the correlations from the full model and the factor analysis model. For many pairs of environments, the difference in the estimates is small. However, for certain pairs, there is a substantial deviation between the FA and full model. These appear to be concentrated on the Sevilla environments.

	Badajoz	Coruna	Cordoba	Lleida	Orense	Salamanca	Sevilla_C	Sevilla_E	Toledo
Albacete	-0.09	-0.01	-0.08	-0.22	0.04	-0.29	-0.09	-0.16	-0.07
Badajoz		0.00	-0.18	0.00	0.04	0.07	-0.18	-0.09	0.06
Coruna			-0.01	-0.03	0.01	0.00	0.01	-0.02	-0.01
Cordoba				-0.27	-0.15	0.00	-0.05	-0.38	-0.04
Lleida					-0.09	-0.01	-0.54	-0.60	-0.27
Orense						0.21	-0.17	-0.13	-0.12
Salamanca							0.32	0.05	0.01
Sevilla_C								-0.53	0.03
Sevilla_E									-0.02