Cross validation of multiplicative terms of AMMI models for blocked designs

The Additive Main Effects and Multiplicative Interaction (AMMI) model is one of the most popular models for the analysis of multi-environment trials (MET) in variety testing and plant breeding (Gauch, 1992). An important step in the analysis using AMMI models is the determination of the number of multiplicative terms. Different procedures have been proposed for this task, including significance tests and cross validation (CV) schemes.

The most commonly used cross-validation scheme divides the data of each trial into one replicate for validation and the remaining replicates for estimation. The models of the AMMI family (AMMI0, AMMI1, …, AMMIF) are estimated from the estimation data, and cross validation error is then computed by comparing model predictions with the held out validation data.

The MATMODEL software does cross validation (CV) by independently sampling a replicate for every genotype and environment for cross validation. This procedure is not optimal in case the field experimental design involved blocking. As pointed out in Piepho (1994), this procedure does not fully exploit the gain in precision from blocking, and, in fact, CV error is inflated by the block variance. Thus, Piepho (1994) suggested keeping blocks intact in CV.

A very simple alternative idea is to simply pre-process the data by subtraction of the estimated block effect from every observation prior to CV. In an RCBD, the estimated block effect is just the block mean minus the trial mean. Pre-processing is expected to have about the same effect on CV error as the modified CV scheme suggested in 1994, though, of course, the two approaches are not identical. The nice thing about this simple pre-processing is that people could use the popular MATMODEL package to obtain a CV accounting for blocking. Also, one might consider extending the idea to incomplete blocks.

There is an issue with this simple idea because corrected data are not strictly independent due to the correction, but we do not expect this to be a relevant problem in practice. The objective of this Master thesis is therefore to empirically compare the new idea based on pre-processing to the CV procedure proposed in Piepho (1994).

References

Gauch, H.G. (1992): Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs. Elsevier, New York.

Piepho, H.P. (1994): Best linear unbiased prediction (BLUP) for regional yield trials: A comparison to additive main effects multiplicative interaction (AMMI) analysis. *Theoretical and Applied Genetics* **89**, 647-654.