

---

# A guide to interpreting the CAIGE multi-environment pedigree analysis results

Ky L Mathews

February 22, 2021

In 2020 the CAIGE project implemented an across year pedigree analysis for all 3 crops (bread wheat, barley and durum wheat). For the purposes of this document we used the term “variety” to represent a genotype and “environment” to represent the year location combination where a CAIGE evaluation trial was conducted. The statistical model applied to each dataset was a factor analytic linear mixed model where the between variety variance-covariance matrix was estimated by a numerator relationship matrix ( $\mathbf{A}$ ) from the available pedigree information. This model allows the *total* genetic effects to be partitioned into the *additive* and *non-additive* genetic effects, also known as general and specific combining ability, respectively. The additive effects measure the potential of each variety as a parent. The total effects, estimated from the additive and non-additive effects, measure the potential of each variety for varietal release.

The additive and non-additive between environment variance-covariance matrices are modelled with individual factor analytic models which have been shown to parsimoniously model the anticipated variety by environment interaction (VEI). These models produce an unwieldy amount of results where there is a prediction for each variety in each environment, whether it was grown there or not. The factor analytic selection tools (FAST) presented in [Smith & Cullis \(2018\)](#) summarise these results into statistics more easily interpreted by breeders. In essence, there are three measures: overall performance (OP), root mean square deviation which is equivalent to stability (RMSD) and responsiveness. In the CAIGE results these measures are provided for the additive and total effects.

**Overall performance (OP)** provides a measure with which to rank the varieties. It is on the scale of the data and is centred around zero. For yield the scale is t/ha and a high yielding variety will have a high positive OP value and a low yielding variety will have a large negative OP value. OP is based on the first factor in the factor analytic model where the first factor explains the maximum of genetic (additive or total) co-variance in the VE effects, the second factor the next most etc. In the results file, the percentage variance accounted for by the first factor is included in the OP heading name.

**Root Mean Square Deviation (RMSD)** is a measure of stability. As its name implies it measures the scatter of points around the linear regression line for each variety. A small value indicates that the variety is stable and a large value indicates that it is responsive to environmental perturbations. As with OP it is on the scale of the data and is based on the first factor only. In the results file, the percentage variance accounted for by the first factor is included in the RMSD heading name.

**Responsiveness (RESP)** is provided for each factor except factor 1 (generally factor 1 is equivalent to RMSD) and enables the varieties to be ranked based on their performance in the

## REFERENCES

---

contrasting environments for each factor. These statistics are not included in the summary data compilation but are provided in the full set of results available from the CAIGE website and should be interrogated if the percentage variance accounted for by the first factor is not large. The percentage variance accounted for the factor of interest is included in the responsiveness heading name in the results files.

**Common Variety by Environment (CVE) and Variety by Environment (VE)** effects are empirical best linear unbiased predictions (eBLUPS) for each variety in each environment, both additive and total effects. In the data compilation file only the common variety by environment (CVE) effect is provided where 'common' refers to the correlated and therefore repeatable part of the VEI in the dataset. The VE effects (provided in the full results files but not the data compilation file) includes the CVE plus a lack of fit term specific to that VE combination. The VE effects are intrinsically different for varieties with and without data for an environment and this should be well understood when using the VE effects to make decisions. These statistics are provided to enable the breeder to investigate the performance of a variety in a environment of specific interest.

Users are referred to [Smith & Cullis \(2018\)](#) and the references therein for a complete understanding of the use of these statistics.

### References

SMITH, A. & CULLIS, B. R. (2018). Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica* **214**, 1–19.