**Appendix I: Extended explanation of the statistical methodology**

**Standard multi-environment trial models**

The vectorof the observed yield responses across the trials (environments) was analysed using the following linear mixed model:

|  |  |
| --- | --- |
|  | (1) |

where =is the vector of fixed effects containing the trial means and terms to capture global trends at each trial, is its associated design matrix, = is the vector of random genetic effects of the lines present in the trials, is its associated design matrix, is the vector of random non-genetic effects consisting in block effects for each trial and additional model terms to capture extraneous specific trial variation, with associated design matrix, and is the vector of residual errors partitioned by trials.

Initially, a separable variance-covariance of the form was assumed where the variance-covariance of the genetic effects between environments (was a diagonal (DIAG) matrix which accounted for different variances between environments and no genetic correlation between pairs of environments and the lines were independent. Ideally, an unstructured form of is desired, which enables modelling the heterogeneity of genetic variances between environments and heterogeneity of genetic covariances between pairs of environments. Due to the large number of trials, a parsimonious approximation of the unstructured was sought using the factor analytic (FA) approach of Smith et al. (2001). With this approach becomes:

|  |  |
| --- | --- |
|  | (2) |

where is the matrix of loadings and is the diagonal matrix of specific trial variances.

The joint distribution of the random effects was assumed to be Gaussian with zero mean and pairwise independent components. The and , where denotes the variance associated to , denotes the identity matrix,is the trial error variance and refers to an autoregressive process of order one in the column and row directions in the trial.

**Pedigree multi-environment trial models**

Following Oakey et al (2006, 2007), the (total) genetic effects of the lines present in the pedigree were partitioned into the additive genetic effect, , and residual non-additive genetic effects, . The linear mixed model to analyze the data including the pedigree information was:

|  |  |
| --- | --- |
|  | (3) |

where all model terms are defined as in Eqn. 1 and is replaced by .

Separable variance-covariance matrices of the form and were assumed, where the and are between trial unstructured variance-covariance matrices for the additive and non-additive genetic effects, respectively. is the known additive relationship matrix and accounts for the inter-line relationships of the additive genetic effects. The non-additive genetic effects were assumed independent between lines.

The matrix is defined as:

|  |  |
| --- | --- |
| = | (4) |

where is the inbreeding coefficient of line adjusted to generations of sel-fertilization, and is the coefficient of coancestry of line and . The inverse of the matrix was calculated from the pedigree data set following the iterative method described in (Meuwissen and Luo, 1992) but incorporating the adjustment of the inbreeding coefficient for the level of selfing (Oakey et al., 2006).

A parsimonious approximation of the and genetic effects was sought using the FA approach of Smith et al. (2001). With this approach the between trial variance-covariances of the and effects become:

|  |  |
| --- | --- |
|  | (5) |

where and are matrices of loadings and and are diagonal matrices of specific trial variances.

The joint distribution of the random effects was assumed to be Gaussian with zero mean and pairwise independent components. The variance covariance structures for and were identical to the structures defined for these effects in the models without pedigree.

**Factor analytic selection tools**

The predictions obtained of the additive effects were summarized using the factor analytic selection tools (FAST) derived in Smith and Cullis (2018), which provided measures of overall performance and stability for each line. These measures are based on the multiple regression representation of the additive genetic effects. Let be the predicted additive genetic effects of the variety:

|  |  |
| --- | --- |
|  | (6) |

where are the estimated loadings (Eqn. 5) rotated to a principal component solution for a meaningful interpretation and represent the independent variables of the regression, are the predicted rotated scores and represent the regression coefficients, and is a lack of fit term. The FAST tools separate the common additive effects () into the effects associated with the first factor and the rest, i.e. The overall performance (OP) is the fitted value at the mean of the first loading and the stability (Eqn. 7) is the root mean square deviation of (RMSD) from the regression line associated to the first loading.

|  |  |
| --- | --- |
|  | (7) |

**Bibliography**

Meuwissen, T. H. E. & Luo, Z. (1992). Computing inbreeding coefficients in large populations. *Genetics Selection Evolution,* 24**,** 305-313, doi:10.1051/gse:19920402.

Oakey, H., Verbyla, A., Cullis, B., Wei, X. & Pitchford, W. (2007). Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics,* 114**,** 1319-1332, doi:10.1007/s00122-007-0515-3.

Oakey, H., Verbyla, A., Pitchford, W., Cullis, B. & Kuchel, H. (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics,* 113**,** 809-819, doi:10.1007/s00122-006-0333-z.

Smith, A., Cullis, B. & Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics,* 57**,** 1138-1147, doi:10.1111/j.0006-341X.2001.01138.x.

Smith, A. B. & Cullis, B. R. (2018). Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica,* 214**,** 143-161, doi:10.1007/s10681-018-2220-5.