Electronic Appendix

We here describe the implementation of our proposed models and methods using the SAS procedures MIXED or GLIMMIX.

Single-stage analysis: One location, one trial

With YLD as the response variable, the SAS code for a model using the CS structure for both the block effect and the plot effect is given in Figure 1. The statements for the MIXED and GLIMMIX procedures yield equivalent model fits. A particular advantage of the GLIMMIX procedure is that sums of adjusted ("least square") means for yield across harvest years can be conveniently computed using the LSMESTIMATE statement, as is illustrated for the first two genotypes.

```
proc mixed;
  class gen har blk plt;
  model yld=gen har gen*har;
  random har/subject=blk type=cs;
  repeated har/subject=blk*plt type=cs;
  lsmeans gen*har;
run;

proc glimmix;
  class gen har blk plt;
  model yld=gen har gen*har;
  random har/subject=blk type=cs;
  random har/subject=blk*plt type=cs;
  lsmeans gen*har;
  lsmestimate gen*har 'sum genotype 1' 1 1 1;
  lsmestimate gen*har 'sum genotype 2' 0 0 0 1 1 1;
run;
```

Figure 1: MIXED and GLIMMIX statements for a single trial (first harvest year 2003) and location Kalteneber assuming a compound symmetry structure for both random within-trial terms. The GLIMMIX code illustrates computation of total yields for the first two genotypes. Note that GEN must be listed before HAR in the CLASS statement so that GEN•HAR means are first ordered by genotypes and then by harvest years within genotypes. This ordering is important for the appropriate order of coefficients in the LSMESTIMATE statement.

Single-stage analysis: Several locations, one trial

The model can be fitted in MIXED using the code given in Figure 2, if the serial correlation model for all random effects is CS.

```
proc mixed;
  class gen loc har blk plt;
  model yld=gen har gen*har;
  random har/subject=loc type=cs;
  random har/subject=loc*gen type=cs;
  random har/subject=loc*blk type=cs;
  repeated har/subject=loc*blk*plt type=cs;
  lsmeans gen*har;
run;
```

Figure 2: MIXED statements for a single trial (first harvest year 2003) and six locations assuming homogeneity of variance between locations and a compound symmetry structure for both random within-trial terms.

Note that we have used the factor LOC as a subject effect in the three RANDOM statements and the REPEATED statement. This allows SAS to identify locations as independent blocks on the diagonal of the variance-covariance matrix for the data, which may considerably reduce computing time, because an inverse of the variance-covariance matrix can be computed by location (Piepho and Möhring, 2011).

For some of these models we had to experiment with different starting values using the PARMS statement before getting convergence. The AR(1) model gave unrealistic estimates of some auto-correlations in the heterogeneous case, with some estimates tending towards minus one. We therefore constrained all parameters to be non-negative using the LOWERB option to the PARMS statement.

Using the group option to model heterogeneity between locations

The program code in Figure 2 assumes that the serial correlation parameter is the same in each location. This assumption can be relaxed by using the GROUP option as shown in Figure 3.

```
proc mixed;
  class gen loc har blk plt;
  model yld=gen har gen*har;
  random har/subject=loc type=cs;
  random har/subject=loc*gen type=cs;
  random har/subject=loc*blk type=cs group=loc;
  repeated har/subject=loc*blk*plt type=cs group=loc;
  lsmeans gen*har;
run;
```

Figure 3: MIXED statements for a single trial (first harvest year 2003) and six locations assuming heterogeneity of variance between locations and a compound symmetry structure for both random within-trial terms.

Single-stage analysis: One location, several trials

The SAS code to fit the model, again assuming (homogeneous) CS structures for random design effects, is given in Figure 4. Note that because of the repeated measures structure of the model, observations from different years are not independent, so there are no independent blocks in the variance-covariance structure as in the single-trial multi-location case. This is reflected in the program code by the fact that there is no common subject effect or factor in

the RANDOM and REPEATED statements. So effectively there is only a single subject, meaning that all observations are correlated, and this prolongs computing time, because the full variance-covariance matrix of the data needs to be inverted during each iteration of the REML algorithm.

```
proc mixed data=v142;
  class gen yr har trl blk plt;
  model yld=gen har gen*har;
  random yr yr*gen yr*gen*har;
  random   har/subject=trl type=cs;
  random   har/subject=trl*blk type=cs;
  repeated har/subject=trl*blk*plt type=cs;
  lsmeans gen*har;
run;
```

Figure 4: MIXED statements for three trials (first harvest years 2001, 2002 and 2003) conducted at location Kalteneber assuming homogeneity of variance between trials and a compound symmetry structure for all random within-trial terms.

Single-stage analysis: Several locations, several trials

```
proc mixed;
  class gen trl har yr loc blk plt;
  model yld=gen har gen*har;
  random yr yr*gen yr*har yr*gen*har;
  random yr yr*gen        yr*gen*har/subject=loc;
  random   har/subject=loc type=cs;
  random   har/subject=loc*gen type=cs;
  random   har/subject=loc*trl type=cs;
  random   har/subject=loc*trl*blk type=cs;
  repeated har/subject=loc*trl*blk*plt type=cs;
run;
```

Figure 5: MIXED statements for three trials (first harvest years 2001, 2002 and 2003) conducted at nine locations assuming homogeneity of variance between locations and a compound symmetry structure for all random within-trial terms.

Two-stage analysis: One location, one trial

a. Serial correlation estimated in stage two

Assuming homogeneity of variance between locations, the statements for this analysis are shown in Figure 6. If heterogeneity of variances and covariances is assumed between locations, the model may be modified accordingly using the GROUP statement in both the RANDOM and REPEATED statements (not shown) as in single-stage analysis, but we have refrained from doing so here, because the effects **LOC•*HAR*** and **LOC•GEN•*HAR*** in this analysis capture both within and between trial variance.

```
proc mixed;
  class gen har loc;
  model estimate=gen har gen*har;
  random   har/subject=loc type=CS;
  repeated har/subject=loc*gen type=CS;
run;
```

Figure 6: Stage-two MIXED statements for a single trial (first harvest year 2003) and six locations assuming homogeneity of variance between locations and a compound symmetry structure for both random within-trial terms. Genotype means are computed separately for each harvest year.

b1. Serial correlation estimated in stage one, carried forward in blocks

We here supply the variance-covariance matrix of the six locations via a single LIN(6) structure of the MIXED procedure of SAS, i.e., as a linear structure with six components corresponding to the six locations. For each location, only one of the components is required, so we need a way to "switch on" the appropriate component for an environment and to keep all other components switched off. This is achieved with the GROUP=loc option, which fits a separate set of six variance components ϕ_{ij} $(j = 1,...,6)$ for each of the six locations $(i = 1,...,6)$, i.e., there is a total of 6×6 variance parameters for this approach. Using the PARMS statement, these variance components are fixed at $\phi_{ij} = 1$ when $i = j$ (i.e., the *j*-th component is switched on in *i*-th location) and at $\phi_{ij} = 0$ when $i \neq j$ (i.e., the other components are switched off in the *i*-th location). Note that the matrices supplied in the LDATA= option are multiplied with these variance components. The dataset supplied with LDATA= must have the following variables: PARM (a running number for variance components; levels 1-6 in this example), COL1-COL*n*, where *n* is the total number of genotype-harvest year combinations in the trial (columns of the variance-covariance matrix) and ROW (a running number from 1 to *n* within each location). The dataset containing the adjusted means must also have the variable ROW corresponding to that same variable in the LDATA= dataset. In fact, it may be most convenient to use one and the same dataset for both purposes. The linear structure supplied with the LDATA= option has the form $\phi_{i1}\tilde{\boldsymbol{R}}_{\mathbf{1}} + \phi_{i2}\tilde{\boldsymbol{R}}_{\mathbf{2}} + \phi_{i3}\tilde{\boldsymbol{R}}_{\mathbf{3}} + \phi_{i4}\tilde{\boldsymbol{R}}_{\mathbf{4}} + \phi_{i5}\tilde{\boldsymbol{R}}_{\mathbf{5}} + \phi_{i6}\tilde{\boldsymbol{R}}_{\mathbf{6}}$, where $\tilde{\boldsymbol{R}}_{\boldsymbol{i}}$ is an $n \times n$ variance-covariance matrix corresponding to the *i*-th location. If the complete set of *n* genotypes-harvest year combinations was tested at the *i*-th location and all harvest years are available, then $\tilde{\boldsymbol{R}}_{\boldsymbol{i}} = \boldsymbol{R}_{\boldsymbol{i}}$, where $\boldsymbol{R}_{\boldsymbol{i}}$ is the variance-covariance matrix of the adjusted genotype-harvest year means in the *i*-th location. If only a subset of $m < n$ genotype-harvest year combinations was tested in the *i*-th location, then the upper left $m \times m$ block of $\tilde{\boldsymbol{R}}_{\boldsymbol{i}}$ equals $\boldsymbol{R}_{\boldsymbol{i}}$, while the remaining elements are set to zero. Thus, if a location has $m < n$ genotype-by-harvest year combinations, then the variance-covariance matrix $\boldsymbol{R}_{\boldsymbol{i}}$ is stored in the first *m* rows and the variables COL1 to COL*m* of the LDATA= dataset. The remaining rows and columns are set to zero. The variable ROW aligns adjusted means in the *i*-th location with the appropriate rows in the variance-covariance matrix $\tilde{\boldsymbol{R}}_{\boldsymbol{i}}$. An illustration of the required data structure is given in Figure 7 using a small toy example, which assumes that there are eight genotype-harvest year means (four genotypes, two harvest years for each genotype) in the first location with variance-covariance matrix

$$\boldsymbol{R}_1 = \begin{pmatrix} 41.7 & 21.2 & 20.6 & 11.3 & 20.6 & 11.3 & 20.6 & 11.3 \\ 21.2 & 41.7 & 11.3 & 20.6 & 11.3 & 20.6 & 11.3 & 20.6 \\ 20.6 & 11.3 & 41.7 & 21.2 & 20.6 & 11.3 & 20.6 & 11.3 \\ 11.3 & 20.6 & 21.2 & 41.7 & 11.3 & 20.6 & 11.3 & 20.6 \\ 20.6 & 11.3 & 20.6 & 11.3 & 41.7 & 21.2 & 20.6 & 11.3 \\ 11.3 & 20.6 & 11.3 & 20.6 & 21.2 & 41.7 & 11.3 & 20.6 \\ 20.6 & 11.3 & 20.6 & 11.3 & 20.6 & 11.3 & 41.7 & 21.2 \\ 11.3 & 20.6 & 11.3 & 20.6 & 11.3 & 20.6 & 21.2 & 41.7 \end{pmatrix}$$

and four means in the second location (the first two genotype are missing) with variance-covariance matrix

$$\boldsymbol{R}_2 = \begin{pmatrix} 31.9 & 19.9 & 21.9 & 16.0 \\ 19.9 & 31.9 & 16.0 & 21.9 \\ 21.9 & 16.0 & 31.9 & 19.9 \\ 16.0 & 21.9 & 19.9 & 31.9 \end{pmatrix}.$$

gen	har	loc	row	parm	col1	col2	col3	col4	col5	col6	col7	col8	yld
1	1	1	1	1	41.7	21.2	20.6	11.3	20.6	11.3	20.6	11.3	109.3
1	2	1	2	1	21.2	41.7	11.3	20.6	11.3	20.6	11.3	20.6	133.3
2	1	1	3	1	20.6	11.3	41.7	21.2	20.6	11.3	20.6	11.3	108.3
2	2	1	4	1	11.3	20.6	21.2	41.7	11.3	20.6	11.3	20.6	120.0
3	1	1	5	1	20.6	11.3	20.6	11.3	41.7	21.2	20.6	11.3	124.3
3	2	1	6	1	11.3	20.6	11.3	20.6	21.2	41.7	11.3	20.6	112.4
4	1	1	7	1	20.6	11.3	20.6	11.3	20.6	11.3	41.7	21.2	98.7
4	2	1	8	1	11.3	20.6	11.3	20.6	11.3	20.6	21.2	41.7	99.2
3	1	2	1	2	31.9	19.9	21.9	16.0	0	0	0	0	86.3
3	2	2	2	2	19.9	31.9	16.0	21.9	0	0	0	0	82.1
4	1	2	3	2	21.9	16.0	31.9	19.9	0	0	0	0	79.4
4	2	2	4	2	16.0	21.9	19.9	31.9	0	0	0	0	79.9
.	.	.	5	2	0	0	0	0	0	0	0	0	.
.	.	.	6	2	0	0	0	0	0	0	0	0	.
.	.	.	7	2	0	0	0	0	0	0	0	0	.
.	.	.	8	2	0	0	0	0	0	0	0	0	.

$\widetilde{\boldsymbol{R}}_i$

Figure 7: Structure of dataset to be supplied to the repeated statement using the LDATA= option as well as in the DATA= option to the call line of MIXED procedure, when the variance-covariance matrix of adjusted means is to be transferred in blocks by subject. The toy example comprises four genotypes, two harvest years, and two locations. The two first genotypes are not tested in the second location.

In case of our *Lolium* example, a single dataset g3 contains all relevant variables. With this dataset, the code given in Figure 8 may be used in stage two. Variance estimates are quite close to those from single-stage analysis (Table 3), but analysis was much faster. Note that the method implies heterogeneity of error variances between locations, because locations are analysed separately, as opposed to method (a), which assumes homogeneity.

```
proc mixed data=g3;
  class gen har loc row;
  model estimate=gen har gen*har;
  random   har/subject=loc type=CS;
  random   har/subject=loc*gen type=CS;
  repeated row/subject=loc group=loc type=lin(6) ldata=g3;
  parms (1)(1)(1)(1)
        (1)(0)(0)(0)(0)(0)
        (0)(1)(0)(0)(0)(0)
        (0)(0)(1)(0)(0)(0)
        (0)(0)(0)(1)(0)(0)
        (0)(0)(0)(0)(1)(0)
        (0)(0)(0)(0)(0)(1)
        /hold=5 to 40;
run;
```

Figure 8: Stage-two MIXED statements for a single trial (first harvest year 2003) and six locations. Genotype-harvest year means computed along with variance-covariance matrix in stage one. The variance-covariance matrix of adjusted means is transmitted by subject (location). Adjusted means and variance-covariance matrix are contained in the same file (g3) in a format analogous to that shown in Figure 7. The first four entries in the PARMS statement refer to the CS structures in the first two RANDOM statements.

b2. Serial correlation estimated in stage one, carried forward as a single R matrix

In the previous section, the variance-covariance matrix of adjusted means ($\boldsymbol{R}$) was carried forward in blocks by subject (location in this case). Alternatively, the blocks may be joined to form a single block-diagonal variance-covariance matrix that is then transmitted via the REPEATED statement using a LIN(1) structure. The great advantage of this approach is that only a single variance parameter is needed in the model specification, which reduces computing time compared to the approach described in the previous section, where the number of parameters grows as a quadratic function of the number of subjects (locations). Figure 8 shows how the data must be coded. The variable ROW here also aligns the adjusted means with the corresponding entries in the variance-covariance matrix, but is coded differently than with method (b1) because only a single matrix is transmitted (see Figure 10). Using the SUBJECT option in the REPEATED statement, MIXED recognizes the block-diagonal nature of the $\boldsymbol{R}$ matrix. In the toy example of Figure 9, this matrix takes the form

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_2 \end{pmatrix} = \boldsymbol{R}_1 \oplus \boldsymbol{R}_2 ,$$

where $\oplus$ denotes the direct sum operator (Searle et al., 1992). More generally, the block-diagonal matrix $\boldsymbol{R}$ can be written as $\boldsymbol{R} = \bigoplus_{i=1}^{s} \boldsymbol{R}_i$, where s is the number of subjects (locations, trials, etc.).

gen	har	loc	row	parm	col1	col2	col3	col4	col5	col6	col7	col8	col9	col10	col11	col12	yld
1	1	1	1	1	41.7	21.2	20.6	11.3	20.6	11.3	20.6	11.3	0	0	0	0	109.3
1	2	1	2	1	21.2	41.7	11.3	20.6	11.3	20.6	11.3	20.6	0	0	0	0	133.3
2	1	1	3	1	20.6	11.3	41.7	21.2	20.6	11.3	20.6	11.3	0	0	0	0	108.3
2	2	1	4	1	11.3	20.6	21.2	41.7	11.3	20.6	11.3	20.6	0	0	0	0	120.0
3	1	1	5	1	20.6	11.3	20.6	11.3	41.7	21.2	20.6	11.3	0	0	0	0	124.3
3	2	1	6	1	11.3	20.6	11.3	20.6	21.2	41.7	11.3	20.6	0	0	0	0	112.4
4	1	1	7	1	20.6	11.3	20.6	11.3	20.6	11.3	41.7	21.2	0	0	0	0	98.7
4	2	1	8	1	11.3	20.6	11.3	20.6	11.3	20.6	21.2	41.7	0	0	0	0	99.2
3	1	2	9	2	0	0	0	0	0	0	0	0	31.9	19.9	21.9	16.0	86.3
3	2	2	10	2	0	0	0	0	0	0	0	0	19.9	31.9	16.0	21.9	82.1
4	1	2	11	2	0	0	0	0	0	0	0	0	21.9	16.0	31.9	19.9	79.4
4	2	2	12	2	0	0	0	0	0	0	0	0	16.0	21.9	19.9	31.9	79.9

R

Figure 9: Structure of dataset to be supplied to the repeated statement using the LDATA= option as well as in the DATA= option to the call line of MIXED procedure, when the variance-covariance matrix of adjusted means is to be transferred as a single block-diagonal matrix. The toy example comprises four genotypes, two harvest years, and two locations. The first two genotypes are not tested in the second location. The first four entries in the PARMS statement refer to the CS structures in the first two RANDOM statements.

```
proc mixed data=cc;
  class gen har loc row;
  model estimate=gen har gen*har;
  random   har/subject=loc type=CS;
  random   har/subject=loc*gen type=CS;
  repeated row/subject=loc type=lin(1) ldata=cc;
  parms (1)(1)(1)(1)
        (1)
        /hold=5;
run;
```

Figure 10: Stage-two MIXED statements for a single trial (first harvest year 2003) and six locations. Genotype-harvest year means computed along with variance-covariance matrix in stage one. The variance-covariance matrix of adjusted means is transmitted as a single block-diagonal matrix. Adjusted means and variance-covariance matrix are contained in the same file (cc) in a format analogous to that shown in Figure 9.

Apart from small numerical differences, the results are the same as with method (b1) in the previous section, so we do not report them separately (see Table 3). When there is a large number of subjects (s) in $\boldsymbol{R}$, the method described here may be considerably faster than that given in the previous section.

Two-stage analysis: One location, several trials

a. Serial correlation estimated in stage two

A stage-two model may be written as

GEN + HAR + GEN•HAR : YR + YR•GEN + YR•GEN•HAR
+ **TRL•*HAR*** + **TRL•GEN•*HAR*** .

The effect YR•GEN•HAR is confounded with **TRL•GEN•*HAR*** (see sub-section ‘Several locations, several trials’ in the section ’Single-stage analysis’) and so is dropped. The SAS

code is shown in Figure 11, the corresponding variance component estimates are given in Table 5.

```
proc mixed data=c;
  class gen trl har yr;
  model estimate=gen har gen*har;
  random yr yr*gen;
  random   har/sub=trl type=cs;
  repeated har/sub=trl*gen type=cs;
run;
```

Figure 11: Stage-two MIXED statements for three trials (first harvest years 2001, 2002 and 2003) conducted at location Kalteneber assuming homogeneity of variance between trials and a compound symmetry structure for both random within-trial terms. Genotype means are computed separately for each harvest year.

b1. Serial correlation estimated in stage one

The model for genotype-harvest year means may be written as

GEN + HAR + GEN•HAR : YR + YR•GEN + YR•GEN•HAR
+ **TRL•*HAR***

The SAS code is given in Figure 12, the corresponding variance component estimates are given in Table 5.

```
proc mixed data=g3;
  class gen yr har trl row;
  model estimate=gen har gen*har;
  random yr yr*gen yr*gen*har;
  random   har/subject=trl type=cs;
  repeated row/subject=trl group=trl type=lin(3) ldata=g3;
  parms (1)(1)(1)(1)(1)
        (1)(0)(0)
        (0)(1)(0)
        (0)(0)(1)
        /hold=6 to 14;
run;
```

Figure 12: Stage-two MIXED statements for three trials (first harvest years 2001, 2002 and 2003) conducted at location Kalteneber assuming homogeneity of variance between trials and a compound symmetry structure for the within-trial term **TRL•*HAR***. Genotype-harvest year means computed along with variance-covariance matrix in stage one. Variance-covariance matrix of adjusted means is transmitted by subject (trial). Adjusted means and variance-covariance matrix are contained in the same file (g3) in a format analogous to that shown in Figure 7.

Results are shown in Table 5.

b2. Serial correlation estimated in stage one, carried forward as a single R matrix

```
proc mixed data=cc;
  class gen yr har trl row;
  model estimate=gen har gen*har;
  random yr yr*gen yr*gen*har;
  random   har/subject=trl type=cs;
  repeated row/subject=trl type=lin(1) ldata=cc;
  parms (1)(1)(1)(1)(1)
        (1)
        /hold=6;
run;
```

Figure 13: Stage-two MIXED statements for three trials and one location (Kalteneber). Genotype-harvest year means computed along with variance-covariance matrix in stage one. Variance-covariance matrix of adjusted means is transmitted as a single matrix. Adjusted means and variance-covariance matrix are contained in the same file (cc) in a format analogous to that shown in Figure 9.

Results are shown in Table 5.

Two-stage analysis: Several locations, several trials

a. Serial correlation estimated in stage two

Our initial model is:

(GEN : YR + YR•LOC + YR•GEN + LOC•YR•GEN) × HAR
+ **LOC•*HAR*** + **LOC•GEN•*HAR*** + **LOC•TRL•*HAR*** + **LOC•TRL•GEN•*HAR***

= GEN + HAR + GEN•HAR : YR + YR•LOC + YR•GEN + LOC•YR•GEN
+ YR•HAR + YR•LOC•HAR + YR•GEN•HAR + LOC•YR•GEN•HAR
+ **LOC•*HAR*** + **LOC•GEN•*HAR*** + **LOC•TRL•*HAR*** + **LOC•TRL•GEN•*HAR***

The effects LOC•YR•GEN•HAR and **LOC•TRL•GEN•*HAR*** are confounded because of the contained effects LOC•YR•HAR and **LOC•TRL•*HAR***, which are also confounded (see sub-section ‘Several locations, several trials’ in the section ’Single-stage analysis’). For this reason, we drop the effects LOC•YR•HAR and LOC•YR•GEN•HAR.

```
proc mixed data=d;
  class gen trl har yr loc;
  model estimate=gen har gen*har;
  random yr yr*gen yr*har yr*gen*har;
  random yr yr*gen                  /subject=loc;
  random   har/subject=loc type=CS;
  random   har/subject=loc*gen type=CS;
  random   har/subject=loc*trl type=cs;
  repeated har/subject=loc*trl*gen type=cs;
run;
```

Figure 14: Stage-two MIXED statements for three trials (first harvest years 2001, 2002 and 2003) conducted at two locations (Osterseeon and Kalteneber) assuming homogeneity of variance between trial-location combinations and a compound symmetry structure for both random within-trial terms. Genotype means computed separately for each harvest year.

b1. Serial correlation estimated in stage one

(GEN : YR + YR•LOC + YR•GEN + LOC•YR•GEN) × HAR
+ **LOC•*HAR*** + **LOC•GEN•*HAR*** + **LOC•TRL•*HAR***

The confounded effect YR•LOC•HAR is dropped with method (b) (see sub-section ‘Several locations, several trials’ in the section ’Single-stage analysis’).

```
proc mixed data=g3;
  class gen loc trl har yr row;
  model estimate=gen har gen*har;
  random yr yr*gen yr*har yr*gen*har;
  random yr yr*gen        yr*gen*har/subject=loc;
  random   har/subject=loc type=CS;
  random   har/subject=loc*gen type=CS;
  random   har/subject=loc*trl type=CS;
  repeated row/subject=loc*trl group=loc*trl type=lin(5) ldata=g3;
  parms (1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)
        (1)(0)(0)(0)(0)
        (0)(1)(0)(0)(0)
        (0)(0)(1)(0)(0)
        (0)(0)(0)(1)(0)
        (0)(0)(0)(0)(1)
        /hold=14 to 38
         lowerb=0,0,0,0,0,0,0,0,0,0,0,0,0;
run;
```

Figure 15: Stage-two MIXED statements for three trials (first harvest years 2001, 2002 and 2003) conducted at two locations (Osterseeon and Kalteneber) (five trial-by-location combinations) assuming homogeneity of variance between trial-location combinations and a compound symmetry structure for both random within-trial terms. Genotype-harvest year means computed along with variance-covariance matrix in stage one. Variance-covariance matrix of adjusted means is transmitted by subject (trial-location combination). Adjusted means and variance-covariance matrix are contained in the same file (g3) in a format analogous to that shown in Figure 7.

b2. Serial correlation estimated in stage one, carried forward as a single R matrix

```
proc mixed data=cc;
  class gen loc trl har yr row;
  model estimate=gen har gen*har;
  random yr yr*gen yr*har yr*gen*har;
  random yr yr*gen        yr*gen*har/subject=loc;
  random   har/subject=loc     type=cs;
  random   har/subject=loc*gen type=cs;
  random   har/subject=loc*trl type=cs;
  repeated row/subject=loc*trl type=lin(1) ldata=cc;
  parms (1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)
        (1)
        /hold=14 lowerb=0,0,0,0,0,0,0,0,0,0,0,0,0;
run;
```

Figure 16: Stage-two MIXED statements for three trials and two locations (Osterseeon and Kalteneber) (five trial-by-location combinations). Genotype-harvest year means computed along with variance-covariance matrix in stage one. Variance-covariance matrix of adjusted means is transmitted as a single matrix. Adjusted means and variance-covariance matrix are contained in the same file (cc) in a format analogous to that shown in Figure 9.

Dataset and complete SAS code

The Lolium dataset is made available at the Journal's website (file lolium.xls). The full SAS code that was used to perform the analyses for the Lolium data is provided in the file lolium.sas.

References

PIEPHO H.P. and MÖHRING J. (2011) On estimation of genotypic correlations and their standard errors by multivariate REML using the MIXED procedure of the SAS System. *Crop Science*, **51**, 2449-2454.