Analyzing Multi-Environment Variety Trials

Using Randomization-Derived Mixed Models

T. Caliński, S. Czajka, Z. Kaczmarek, P. Krajewski, and W. Pilarczyk

Appendix

To describe the randomization procedure underlying the model considered in Section 2.1, suppose that a large number of experimental fields at different environments within an agricultural region of interest are conceptually available, that each of them can be divided into several subfields, called superblocks, each of which can further be divided into several parts, called blocks, and that, in turn, each of them can be divided into a suitable number of plots considered as experimental units. Let the original labels for the experimental fields (synonymous with environments) be $\epsilon = 1, 2, ..., N_E$, for the superblocks within field $\epsilon$ be $\kappa(\epsilon) = 1, 2, ..., A_\epsilon$, for the blocks within superblock $\kappa(\epsilon)$ be $\xi\{\kappa(\epsilon)\} = 1, 2, ..., B_{\kappa(\epsilon)}$, and for the plots within block $\xi\{\kappa(\epsilon)\}$ be $\pi[\xi\{\kappa(\epsilon)\}] = 1, 2, ..., K_{\xi\{\kappa(\epsilon)\}}$. Before conducting trials under these options, the experimental fields, the superblocks within fields, the blocks within superblocks and the units within blocks are all to be randomized using independent randomizations. The purpose of these randomizations is not only to “homogenize” the within-field variation among the superblocks, the within-superblock variation among the blocks, and the within-block variation among the experimental units (plots), but also to “average out” these possible heterogeneous variances and covariances from different environments, different superblocks, and different blocks to some common values (cf., White, 1975, Sections 2 and 7). For more understanding of this randomization procedure and its probabilistic consequences see Caliński and Kageyama (2000, Sections 3.1.1 and 5.2.1).
Adopting the concept of a “null” multi-environment experiment (as in Nelder, 1965, p. 147), i.e., assuming that on all possible $N = \sum_{\epsilon=1}^{N_E} \sum_{\xi:\{\kappa(\epsilon)\}=1}^{A_{\epsilon}} \sum_{\xi:\{\kappa(\epsilon)\}=1}^{B_{\kappa(\epsilon)}} K_{\xi:\{\kappa(\epsilon)\}}$ plots the same variety is grown, let the true yield of the plot originally labeled $\pi[\xi:\{\kappa(\epsilon)\}]$ be denoted by $\mu_{\pi[\xi:\{\kappa(\epsilon)\}]}$. Also, let $e_{\pi[\xi:\{\kappa(\epsilon)\}]}$ denote the technical error by which the variable observed on the plot $\pi[\xi:\{\kappa(\epsilon)\}]$ differs from $\mu_{\pi[\xi:\{\kappa(\epsilon)\}]}$. As usual, assume that $\{e_{\pi[\xi:\{\kappa(\epsilon)\}]}\}$ are uncorrelated random variables, each with expectation zero and a finite variance, and that they are distributed independently of the randomizations applied. Now, introducing the identity

$$
\mu_{\pi[\xi:\{\kappa(\epsilon)\}]} = \mu_{\{\cdot(\epsilon)\}} + (\mu_{\{\cdot(\epsilon)\}} - \mu_{\{\cdot(\epsilon)\}}) + (\mu_{\{\cdot(\epsilon)\}} - \mu_{\{\cdot(\epsilon)\}}) + (\mu_{\{\cdot(\epsilon)\}} - \mu_{\{\cdot(\epsilon)\}}),
$$

with the usual dot notation for means, and the variance components

$$
\sigma_E^2 = (N_E - 1)^{-1} \sum_{\epsilon=1}^{N_E} (\mu_{\{\cdot(\epsilon)\}} - \mu_{\{\cdot(\epsilon)\}})^2,
$$

$$
\sigma_A^2 = N_E^{-1} \sum_{\epsilon=1}^{N_E} \sigma_{A_{\epsilon}}^2, \quad \text{where} \quad \sigma_{A_{\epsilon}}^2 = (A_{\epsilon} - 1)^{-1} \sum_{\kappa(\epsilon)=1}^{A_{\epsilon}} (\mu_{\{\cdot(\epsilon)\}} - \mu_{\{\cdot(\epsilon)\}})^2,
$$

$$
\sigma_B^2 = N_E^{-1} \sum_{\epsilon=1}^{N_E} A_{\epsilon}^{-1} \sum_{\kappa(\epsilon)=1}^{A_{\epsilon}} \sigma_{B_{\kappa(\epsilon)}}^2, \quad \text{where} \quad \sigma_{B_{\kappa(\epsilon)}}^2 = (B_{\kappa(\epsilon)} - 1)^{-1} \sum_{\xi:\{\kappa(\epsilon)\}=1}^{B_{\kappa(\epsilon)}} (\mu_{\{\cdot(\epsilon)\}} - \mu_{\{\cdot(\epsilon)\}})^2,
$$

$$
\sigma_U^2 = N_E^{-1} \sum_{\epsilon=1}^{N_E} A_{\epsilon}^{-1} \sum_{\kappa(\epsilon)=1}^{A_{\epsilon}} B_{\kappa(\epsilon)}^{-1} \sum_{\xi:\{\kappa(\epsilon)\}=1}^{B_{\kappa(\epsilon)}} \sigma_{U,\xi:\{\kappa(\epsilon)\}}^2,
$$

where

$$
\sigma_{U,\xi:\{\kappa(\epsilon)\}}^2 = (K_{\xi:\{\kappa(\epsilon)\}} - 1)^{-1} \sum_{\pi:\{\xi:\{\kappa(\epsilon)\}\}=1}^{K_{\xi:\{\kappa(\epsilon)\}}} (\mu_{\pi[\xi:\{\kappa(\epsilon)\}]} - \mu_{\{\cdot(\epsilon)\}})^2,
$$

$$
\sigma_T^2 = N_E^{-1} \sum_{\epsilon=1}^{N_E} A_{\epsilon}^{-1} \sum_{\kappa(\epsilon)=1}^{A_{\epsilon}} B_{\kappa(\epsilon)}^{-1} \sum_{\xi:\{\kappa(\epsilon)\}=1}^{B_{\kappa(\epsilon)}} \text{Var}(e_{\pi[\xi:\{\kappa(\epsilon)\}]}),
$$

and using the weighted harmonic averages $A_H$, $B_H$ and $K_H$, defined by the equalities

$$
A_H^{-1} = N_E^{-1} \sum_{\epsilon=1}^{N_E} A_{\epsilon}^{-1} \sigma_{A_{\epsilon}}^2 / \sigma_A^2, \quad B_H^{-1} = N_E^{-1} \sum_{\epsilon=1}^{N_E} A_{\epsilon}^{-1} \sum_{\kappa(\epsilon)=1}^{A_{\epsilon}} B_{\kappa(\epsilon)}^{-1} \sigma_{B_{\kappa(\epsilon)}}^2 / \sigma_B^2 \quad \text{and} \quad K_H^{-1} = N_E^{-1} \sum_{\epsilon=1}^{N_E} A_{\epsilon}^{-1} \sum_{\kappa(\epsilon)=1}^{A_{\epsilon}} B_{\kappa(\epsilon)}^{-1} \sum_{\xi:\{\kappa(\epsilon)\}=1}^{B_{\kappa(\epsilon)}} K_{\xi:\{\kappa(\epsilon)\}} \sigma_{U,\xi:\{\kappa(\epsilon)\}}^2 / \sigma_U^2,
$$

respectively, one can write the model of the variable (yield) observed on the plot $\pi[\xi:\{\kappa(\epsilon)\}]$, which after the randomizations receives the label $\ell[h\{g(j)\}]$, as

$$
y_{\ell[h\{g(j)\}]} = \mu + u_{E,j} + u_{A,g(j)} + u_{B,h\{g(j)\}} + e_{U,\ell[h\{g(j)\}]} + e_{T,\ell[h\{g(j)\}]}, \quad (A.1)
$$
where $\mu = \mu_{\cdot \{\cdot\}}$ is a constant parameter, whereas $u_{E,j}$, $u_{A,g(j)}$, $u_{B,h\{g(j)\}}$, $e_{U,\ell\{h\{g(j)\}\}}$ and $e_{T,\ell\{h\{g(j)\}\}}$ are random variables, representing an environment random effect, a superblock random effect, a block random effect, a unit error and a technical error, respectively. Their expectations are $E(u_{E,j}) = E(u_{A,g(j)}) = E(u_{B,h\{g(j)\}}) = E(e_{U,\ell\{h\{g(j)\}\}}) = E(e_{T,\ell\{h\{g(j)\}\}}) = 0$ and their variances and covariances involve the above variance components and weighted harmonic averages in a similar manner to that described in Caliński and Kageyama (2000, Section 5.2.1). In particular,

$$\text{Cov}(u_{E,j}, u_{E,j'}) = \begin{cases} N^{-1}_E (N_E - 1) \sigma_E^2 & \text{if } j = j', \\ -N^{-1}_E \sigma_E^2 & \text{if } j \neq j'. \end{cases}$$

(A.2)

Let this preliminary model be now adjusted to a real situation of comparing $m$ distinct varieties in $p$ experiments conducted at different environments, each of the experiments being designed according to a GL design, the same for each experiment or possibly different.

Although the choice of one such design for all $p$ experiments seems to be preferable, this will not be considered as a necessary condition here. However, it will be assumed that any such design has the same number of superblocks, $a$, that each superblock is composed of the same number of blocks, $b_0$, and that each block is composed of the same number of experimental units (plots), $k$, so that the number of blocks for each design is $b = ab_0$. Moreover, it will be assumed that $b_0k = m$. These numbers will allow the experiments to be designed so that each variety appears in any superblock exactly once, as required for a GL design with $r = a$ replications. For any such design the allocation of varieties to the blocks and, hence, to the superblocks can be described by an $m \times b$ incidence matrix

$$N_{(j)} = [N_{1(j)} : N_{2(j)} : \cdots : N_{a(j)}], \quad j = 1, 2, ..., p,$$

(A.3)

where $N_{g(j)} = [n_{ih\{g(j)\}}]$ is an $m \times b_0$ submatrix for superblock $g(j)$, with rows corresponding to the varieties and columns corresponding to the blocks, the element $n_{ih\{g(j)\}}$ being 1 if the
variety \( i \) is to be allocated in the block \( h \), and being 0 otherwise. Assuming that \( p \) suitable designs describable by \((A.3)\) have been chosen, not necessarily all equal, the principle of randomization can be implemented by randomizing the \( p \) incidence matrices \( \{ N_{(j)} \} \) to \( p \) out of the \( N_E \) available environments, the \( a \) submatrices \( \{ {N}_{(g(j))} \} \) of \( N_{(j)} \) to \( a \) of the \( A_j \) superblocks available at the \( j \)th environment, and the \( b_0 \) columns of \( {N}_{g(j)} \) to \( b_0 \) out of the \( B_{g(j)} \) blocks available in the \( g(j) \)th superblock. This assignment is to be accomplished for \( h \{ g(j) \} = 1, 2, ..., b_0, g(j) = 1, 2, ..., a \) and \( j = 1, 2, ..., p \). Finally, the varieties indicated (through 1s) by the \( h \{ g(j) \} \)th column of \( {N}_{g(j)} \) are assigned to the plots of the block labeled \( h \{ g(j) \} \), in the order determined by the labels \( \ell [h \{ g(j) \}] \), for \( \ell = 1, 2, ..., k \), which the plots of the block have received after the randomization. Certainly, this rule implies that \( p \leq N_E \), that \( a \leq A_j \) for \( j = 1, 2, ..., p \), that \( b_0 \leq B_{g(j)} \) for any \( g(j) \), and that the plots in all available blocks are in numbers not smaller than \( k \).

To obtain a model that incorporates the inclusion of different varieties under the described randomization procedure, it is usual (see White, 1975, Section 5) to adopt the strong assumption of unit—variety and technical error—variety additivities holding over the whole series of experiments, i.e., to use the model

\[
y_{\ell \{ h \{ g(j) \} \}}(i) = \mu(i) + u_{E,j} + u_{A,g(j)} + u_{B,h \{ g(j) \}} + \epsilon_U,\ell [h \{ g(j) \}] + \epsilon_T,\ell [h \{ g(j) \}],
\]

(A.4)

with \( \mathbb{E}\{y_{\ell \{ h \{ g(j) \} \}}(i)\} \) equal to

\[
\mu(i) = N_E^{-1} \sum_{\epsilon=1}^{N_E} \sum_{\kappa(e)=1}^{A_e} \frac{A_e}{A} \sum_{\kappa(e)=1}^{B_{\kappa(e)}} \sum_{K_{\xi(\kappa(e))}=1}^{K_{\xi(\kappa(e))}} \sum_{\pi(\xi(\kappa(e)))=1}^{\pi(\xi(\kappa(e)))} \mu_{\pi(\xi(\kappa(e)))}(i),
\]

(A.5)

where \( \mu_{\pi(\xi(\kappa(e)))}(i) \) is the true response (yield) of plot \( \pi \) in block \( \xi \) of superblock \( \kappa \) at environment \( \epsilon \) to variety \( i \), and with \( \text{Cov}\{y_{\ell \{ h \{ g(j) \} \}}(i)\}, y_{\ell' \{ h' \{ g'\{j' \} \} \}}(i')\} \) equal to

\[
\text{Cov}\{y_{\ell \{ h \{ g(j) \} \}}, y_{\ell' \{ h' \{ g'\{j' \} \}}\} = (\delta_{jj'} - N_E^{-1}) \sigma_E^2 + \delta_{jj'} (\delta_{gg'} - A_{H}^{-1}) \sigma_A^2 + \delta_{jj'} \delta_{gg'} (\delta_{hh'} - B_{H}^{-1}) \sigma_B^2 + \delta_{jj'} \delta_{gg'} \delta_{hh'} \delta_{tt'} \sigma_T^2
\]

(A.6)
for all $\ell[h\{g(j)\}]$ and all $\ell'[h'\{g'(j')\}]$, the $\delta$'s being the usual Kronecker delmas, denoting 1 if the indices coincide, and 0 otherwise. Another possibility, more realistic for the analysis of MET data, is to restrict the unit-variety additivity assumption to conditions within the experimental fields only, i.e., to use, instead of (A.4), the model

$$y_{\ell[h\{g(j)\}]}(i) = \mu(i) + u_{E,j}(i) + u_{A,g(j)} + u_{B,h\{g(j)\}} + e_{U,\ell[h\{g(j)\}]} + e_{T,\ell[h\{g(j)\}]}.$$  \hfill (A.7)

This, however, requires the replacement of (A.2) by

$$\text{Cov}\{u_{E,j}(i), u_{E,j'}(i')\} = \begin{cases} 
N_{E}^{-1}(N_{E} - 1)\sigma_{E(i')} & \text{if } j = j', \\
-N_{E}^{-1}\sigma_{E(i')} & \text{if } j \neq j',
\end{cases}$$  \hfill (A.8)

where, for any $i, i'$,

$$\sigma_{E(i')} = (N_{E} - 1)^{-1}\sum_{\epsilon=1}^{N_{E}} \{\mu_{[\{\epsilon\}]}(i) - \mu_{[\{\epsilon\}]}(i)\}\{\mu_{[\{\epsilon\}]}(i') - \mu_{[\{\epsilon\}]}(i')\}.$$  \hfill (A.9)

Hence, $\text{Cov}\{y_{\ell[h\{g(j)\}]}(i), y_{\ell'[h'(j')]}(i')\}$ becomes equal to (A.6), but with $\sigma_{E}^{2}$ replaced by $\sigma_{E(i')}$. Now, writing (A.7) in a more familiar way as

$$y_{h\{g(j)\}} = \tau_{i} + u_{E,ij} + u_{A,g(j)} + u_{B,h\{g(j)\}} + e_{U,i[h\{g(j)\}]} + e_{T,i[h\{g(j)\}]};$$  \hfill (A.10)

where $y_{h\{g(j)\}} = y_{\ell[h\{g(j)\}]}(i), e_{U,i[h\{g(j)\}]} = e_{U,\ell[h\{g(j)\}]}$, $e_{T,i[h\{g(j)\}]} = e_{T,\ell[h\{g(j)\}]}$ for $i = \ell$, and where $\tau_{i} = \mu(i)$, $u_{E,ij} = u_{E,j}(i)$, one can express the model, using the vector and matrix notation, as in (1), with the expectation vector (2) and the covariance matrix (3).