Lecture 31:
\textit{G x E: Selection response}

Bruce Walsh lecture notes
Synbreed course
version 12 July 2013
Key concept

- $G \times E$ is a correlated traits problem (Falconer)
- For a defined set of environments (1 through $k$), define $z_k$ as the trait value in that environment
  - $\text{Var}(z_k), \text{Var}(G_k), \text{Var}(A_k)$ are the phenotypic, genetic, and additive variances in environment $k$. $G \times E$ present if these variances differ over $k$
  - If $\text{corr}(A_k,A_k)$ or $\text{corr}(G_k,G_k)$ different from one, $G \times E$ present
- Direct response = change in selected environment
- Correlated response - change in the unselected environment
The Cost to Response from $G \times E$

As a benchmark for selection when $G \times E$ is present, if environmental structure is ignored and simple mass selection used (choosing the best performing individuals based solely on their phenotypic values), then the expected response becomes

$$R = \bar{r} \sigma_z h_z^2 = \bar{r} \frac{\sigma^2_A}{\sigma_z} = \bar{r} \frac{\sigma^2_A}{\sqrt{\sigma^2_G + \sigma^2_{G\times E} + \sigma^2_E}}$$  \hspace{1cm} (38.9a)$$

where $\sigma^2_G$ and $\sigma^2_E$ are the genetic and environmental variances. When $\sigma^2_{G\times E}$ is large relative to $\sigma^2_A$, the heritability is low and selection very inefficient, as an individual’s phenotypic value in one environment is a poor predictor of their average breeding value over all environments. If we are selecting among clones (or pure lines) then $\sigma^2_G$ replaces $\sigma^2_A$. Setting $\sigma^2_{G\times E}$ to zero, Matheson and Cotterill (1990) note that the “cost” (loss of potential gain) of genotype-environment interaction when using standard mass selection is

$$1 - \sqrt{\frac{\sigma^2_G + \sigma^2_E}{\sigma^2_G + \sigma^2_{G\times E} + \sigma^2_E}}$$  \hspace{1cm} (38.9b)$$

Hence, when Var($G \times E$) is a large fraction of Var($z$), ignoring $G \times E$ critical
Selecting under $G \times E$: Two possible genetic scenarios

- **Using fully inbred (pure) lines**
  - Selection target is the genotype value $G$
  - Can replicate genotypes over environments

- **Using segregating individuals**
  - Selection target is usually the breeding value $A$
  - Cannot replicate genotypes over environments
  - Can replicate by using relatives (typically family members) over environments. Here target is the family mean breeding value.
Estimating the cross-environmental genetic covariance

First, suppose we have a set of pure lines, and each point below represents the mean for that line in environment i and j. Here genotypic values are used.

\[
\text{Slope} = \frac{V(G_i, G_j)}{V(z_j)}
\]

\[
V(G_i, G_j) = \text{slope} \times V(z_j)
\]
Estimating the cross-environmental genetic covariance

Now suppose related, but not genetically identical, individuals are measured in the two environments. For half-sibs, $2\theta = 1/4$; for full-sibs $2\theta = 1/2$. Here breeding values are used.

Value in Environment $i$

Value in Environment $j$

Slope = $2\theta \frac{V(A_i, A_j)}{V(z_j)}$

$V(A_i, A_j) = \text{slope} \frac{V(z_j)}{2\theta}$
Two environments

• Suppose the goal of selection is to improve the index $H = a g_1 + g_2$, where $g_i$ is the genotype or breeding value in population $i$, depending whether one is selection clones or random-mating individuals.

• Suppose we select in environment 1 only. What is the expected response in the index?
  - Direct response in 1: $R_1 = \sigma_{A1} h_{1i}$
  - Correlated response in 2: $CR_2 = r_A \sigma_{A2} h_{1i}$
  - Response in index $R_H = a R_1 + CR_2$, or
  - $R_H(1) = h_{1i} \sigma_{A1} (a + r_A \sigma_{A2}/\sigma_{A1}) i$
  - If selecting using pure lines, genotypic values $G$ replace breeding values $A$ ($r_A \rightarrow r_G, \sigma_{Ai} \rightarrow \sigma_{Gi}$)
• Response in index when selecting in environment one,
  - \( R_{H}(1) = h_1 \sigma_{A1} \left( a + r_A \sigma_{A2}/\sigma_{A1} \right) \)
  - Key observation: Both components of \( G \times E \) enter: differences in genetic variances across environments and genetic correlation over environments.

• Similarly, response when selecting in \( E_2 \) is
  - \( R_{H}(2) = h_2 \sigma_{A2} \left( a + r_A \sigma_{A1}/\sigma_{A2} \right) \)
  - Setting \( v = \sigma_{A2}/\sigma_{A1} \) gives the ratio of responses as

\[
\frac{R_{H}(1)}{R_{H}(2)} = \frac{h_1}{h_2} \left( \frac{a + r_A v}{ar_A + v} \right)
\]
General indirect response

- More generally, the response in environment k from selection in j is
  \[ CR_{k|j} = r_A(k,j) \sigma_{Ak} h_j i_j \]

- Assuming no covariance in environmental values, then the phenotype correlation equals the genetic correlation
  \[ r_P(k,j) = h_j h_k r_A(j,k), \text{ with again } G \text{ replacing } A \text{ for clines. Hence, } r_A(j,k) = r_P(k,j) / h_j h_k \]

\[ CR_{i|j} = \frac{\rho_P(i,j)}{h_i h_j} \sigma_{A_i} h_j \bar{i}_j = \frac{\rho_P(i,j)}{h_i} \frac{\sigma_{A_i}}{h_i} \bar{i}_j = \rho_P(i,j) \sigma_{z_i} \bar{i}_j \]

Entirely a function of phenotypic measures
Selecting in Multiple environments

• Thus far, we have been assuming selection in only one environment, with change in others as a correlated response.

• Now suppose that we can replicate genotypes (selecting using clone/pure lines or sibs) over environments.

• If using members of some genetic group (e.g., full or half sibs) that are measured over different environments, then the correlation of two individuals across environments $i$ and $j$ is given by $2\theta r_A(i,j)$, where $\theta$ is the coefficient of coancestry between the two individuals. This is just a standard result for correlated traits.
• What is the expected response when replicated genotypes (or related genetic group members) are measured over a series of defined environments with selection decision based on the average value of the line or genetic group (i.e., family selection)?

• First consider where replicates are clones (members from the same inbred line).
Suppose inbred lines (typically regarded as being sufficiently inbred to be considered clones, or nearly so) are measured in \( n_e \) environments, often with replication (\( n_r \) individuals from each group measured within each environment). The basic model for the value of the \( k \)-th replicate of line \( i \) in environment \( j \) is

\[
z_{ijk} = \mu + G_i + E_j + GE_{ij} + \epsilon_{ijk}
\]

where \( G \) and \( E \) are the line and macro-environmental effects and \( \epsilon \) the residual (the micro-environmental value that the \( k \)-th replicate experiences, which are assumed to be uncorrelated and homoscedastic with constant variance \( \sigma^2_e \)). If the lines are still segregating, then \( e \) also includes the deviation of the genotypic value from the line mean \( G \). The line mean for genotype \( i \) becomes

\[
\overline{z}_i = \frac{1}{n_e n_r} \sum_{j=1}^{n_e} \sum_{k=1}^{n_r} (\mu + G_i + E_j + GE_{ij} + \epsilon_{ijk})
\]

\[
= \mu + G_i + \frac{1}{n_e} \sum_{j=1}^{n_e} (E_j + GE_{ij}) + \frac{1}{n_e n_r} \sum_{j=1}^{n_e} \sum_{k=1}^{n_r} \epsilon_{ijk}
\]

Assuming that \( G_i, GE_{ij}, \) and \( E_j \) are all uncorrelated, the variance of the line means becomes

\[
\sigma^2(\overline{z}) = \sigma^2_G + \frac{\sigma^2_E + \sigma^2_{G \times E}}{n_e} + \frac{\sigma^2_e}{n_e n_r}
\]
Assuming that $G_{ij}$, $GE_{ij}$, and $E_j$ are all uncorrelated, the variance of the line means becomes

$$\sigma^2(\bar{z}) = \sigma_G^2 + \frac{\sigma_E^2 + \sigma_{GE}^2}{n_e} + \frac{\sigma_e^2}{n_en_r}$$  \hspace{1cm} (38.10c)

Selecting clones with the greatest mean over environments, the expected response becomes

$$R = \overline{\bar{z}} \frac{\sigma_G^2}{\sigma^2_{\bar{z}}} = \overline{\bar{z}} \frac{\sigma_G^2}{\sqrt{\sigma_G^2 + (\sigma_E^2 + \sigma_{GE}^2)/n_e + \sigma_e^2/(n_r n_e)}}$$  \hspace{1cm} (38.10d)

Replication of group members reduces the contributions from $\sigma_{GE}^2$, $\sigma_E^2$, and $\sigma_e^2$ to the variance of the line mean, which results in a higher heritability, increasing response. A common modification of Equation 38.10 is that the environmental effect is often treated as a fixed effect, and hence the data are adjusted to account for this, and the $\sigma_E^2$ term disappears. One version of this is stratified mass selection, when contrasts are made within a given block. Treating $E$ as a fixed effect is a more general way to accomplish this same goal of removing the effects of $E$ (but not $G \times E$!).

13
Sibs are replicated over environments

The idea is essentially the same when using half- or full-sib families, but with a little more bookkeeping (Chapter 17). Selection is based on the family means, with representative members from the chosen families randomly crossed to form the next generation. The resulting response is given by

$$R = \frac{\sigma^2_{AF}}{\sigma_z}$$  \hspace{1cm} (38.11a)

where the between-family additive genetic variance $\sigma^2_{AF}$ is given below and the variance in family means $\sigma^2_z$ is given by Equation 17.39a (using the definitions offered by Equations 17.11a and 17.11b). If $n_r$ family members are measured in each of $n_e$ environments, then setting $N = n_r n_e$,

$$\sigma^2_z = \sigma^2_{GF} + \sigma^2_{E_c} + \frac{\sigma^2_{F \times E}}{n_e} + \frac{\sigma^2_{Gw} + \sigma^2_e}{N}$$  \hspace{1cm} (38.11b)

where $E_c$ is the common family environmental effect, $E$ the remainder of the environmental effects and $GF$ the total genetic variation across families and $F \times E$ the family-by-environment interaction. Ignoring epistasis, the total ($\sigma^2_{GF}$) and additive ($\sigma^2_{AF}$) genetic variation across families is

$$\sigma^2_{GF} = \begin{cases} \frac{(1 - \frac{1}{N})}{4} \frac{1}{4} \sigma^2_A & \text{half-sibs} \\ \frac{(1 - \frac{1}{N})}{4} \frac{1}{2} \sigma^2_A + \frac{1}{4} \sigma^2_D & \text{full-sibs} \end{cases} \hspace{1cm} \sigma^2_{AF} = \begin{cases} \frac{(1 - \frac{1}{N})}{4} \frac{1}{4} \sigma^2_A & \text{half-sibs} \\ \frac{(1 - \frac{1}{N})}{4} \frac{1}{2} \sigma^2_A & \text{full-sibs} \end{cases}$$  \hspace{1cm} (38.11c)
\[
\sigma^2_{GF} = \begin{cases} 
(1 - \frac{1}{N}) \frac{1}{4} \sigma^2_A & \text{half-sibs} \\
(1 - \frac{1}{N}) \left( \frac{1}{2} \sigma^2_A + \frac{1}{4} \sigma^2_D \right) & \text{full-sibs}
\end{cases}
, \quad \sigma^2_{AF} = \begin{cases} 
(1 - \frac{1}{N}) \frac{1}{4} \sigma^2_A & \text{half-sibs} \\
(1 - \frac{1}{N}) \frac{1}{2} \sigma^2_A & \text{full-sibs}
\end{cases}
\] (38.11c)

while the genetic variation within each family \((\sigma^2_{Gw})\) and the family by environment interaction variance \((\sigma^2_{F \times E})\) are

\[
\sigma^2_{Gw} = \begin{cases} 
\frac{3}{4} \sigma^2_A + \sigma^2_D & \text{half-sibs} \\
\frac{1}{2} \sigma^2_A + \frac{3}{4} \sigma^2_D & \text{full-sibs}
\end{cases}
, \quad \sigma^2_{F \times E} = \begin{cases} 
\frac{1}{4} \sigma^2_{A \times E} & \text{half-sibs} \\
\frac{1}{2} \sigma^2_{A \times E} + \frac{1}{4} \sigma^2_{D \times E} & \text{full-sibs}
\end{cases}
\] (38.11d)

Similar expressions can be developed for other types of families, such as \(S_1\) and \(S_2\) (first and second-generation selfing). While these formulae seem a bit busy, the key point to notice is that while the additive-genetic covariance is less with family selection than individual selection, so is the phenotypic variance. In particular, if \(G \times E\) is significant, only part (the family \(\times\) environmental component) appears in the family mean variance and this part is weighted by \(1/n_e\). While the above expressions are typically not directly used, due to the difficulty in estimating the component variances, they provide important insight into the expected response when this sort of selection scheme of evaluating the performance of the genetic group in \(n_e\) random environments is used.
It is important to point out a critical assumption that leads to Equation 38.10d. Starting with Equation 38.10a, we assumed that the genotypic and G × E effects are uncorrelated and homoscedastic (variances are constant, being independent of the subscript on G and GE), namely $G_i \sim (0, \sigma_G^2)$ and $GE_{ij} \sim (0, \sigma_{G\times E}^2)$. This implies that the genetic variances are the same in each environment, as (ignoring the environment random factor E),

$$\sigma(z_{ij}, z_{ij}) = \sigma(G_i + GE_{ij}, G_i + GE_{ij}) = \sigma_G^2 + \sigma_{G\times E}^2$$

Likewise, the genetic covariance between the same genotype (i) measured in two environments (j, k) is

$$\sigma(z_{ij}, z_{ik}) = \sigma(G_i + GE_{ij}, G_i + GE_{ik}) = \sigma_G^2$$

This particular covariance structure wherein the genetic variances are the same across all environments and the genetic covariances are the same across all pairs of environments is called **compound symmetry**. Obviously, this is only a very narrow view of G × E, as in general the genetic variances can change across environments and different pairs of environments can display different correlations (e.g., Equation 38.1b).

This can be treated more formally form a general genetic covariance matrix $\mathbf{G}$ for the trait over environments (return to this shortly). Under **Compound symmetry**, $\mathbf{G} = \sigma_G^2 \mathbf{H} + \sigma_{G\times E}^2 \mathbf{I}$ where $\mathbf{H}$ is a matrix of ones.
Joint selection for performance and sensitivity

- Selection over multiple environments involves at least two types of response:

  - Mean response over all (or some weighted index) of environments
  - Response in sensitivity (or tolerance), variation of response over the individual environments
  - Decreased sensitivity (or tolerance) is especially critical in subsistence agriculture, as farmers and their families simply cannot afford even a single bad year. Hence, lines with high mean performance, but high sensitivity to year-to-year environmental variation, are not appropriate in this setting.
  - A simple two-environmental model provides some insight into these tradeoffs.
Let $u_L = \text{mean in low environment}, \ u_H = \text{mean in high}$

Total production = $u_L + u_H = \text{twice mean performance}$

$m = (u_L + u_H)/2$

Sensitivity = $s = u_L - u_H < 0$. $R_S > 0$ means less sensitivity

Figure 38.2. Mean performance vs. sensitivity. Both populations have the same mean performance (average value over both populations), but rather different sensitivities. The population represented by the solid circles and the solid line has greater sensitivity, so that it has a greater performance in the high environment but significantly poorer performance in the low environment relative to the low sensitivity line with the same mean performance.
\[ \rho_{s,H} = \frac{\sigma(s, H)}{\sigma(s) \sigma(H)} = \frac{\sigma(L - H, H)}{\sqrt{\sigma^2(L - H) \sigma^2(H)}} = \frac{\sigma(L, H) - \sigma^2(H)}{\sqrt{\sigma^4(H) + \sigma^2(H)\sigma^2(L) - 2\sigma^2(H)\sigma^2(L)}} \]  \hspace{1cm} (38.13a)

All correlations, variances and covariances refer to additive genetic variation (if outbreeding populations are considered) or total genetic variance (if selection is among pure lines). Denoting the ratio of the genetic variances for the high versus low environment by

\[ \phi = \frac{\sigma^2(L)}{\sigma^2(H)}, \]  \hspace{1cm} (38.13b)

Rosielle and Hamblin simplify Equation 38.13a to obtain

\[ \rho_{s,H} = \frac{\phi \rho_{H,L} - 1}{\sqrt{1 + \phi^2 - 2\phi \rho_{H,L} \phi}}. \]  \hspace{1cm} (38.13c)

The resulting sign for the genetic correlation between sensitivity \( s \) and high performance \( H \) becomes

\[ \text{sign}(\rho_{s,H}) = \text{sign}(\phi \rho_{H,L} - 1), \]  \hspace{1cm} (38.13d)

which is negative unless \( \sigma^2(L) > \sigma^2(H) \) and \( \rho_{H,L} \) is sufficiently large (such that \( \phi \rho > 1 \)). When this correlation is negative, selection for decreased sensitivity (\( \Delta s > 0 \)) results in a correlated response to decrease the mean performance in the high environment. Thus, unless the genetic variance is larger in the low-performing environment (which is unusual, e.g., Allen et al. 1978), selection for increased tolerance/stability results in a decreased performance in the high environment. Equivalently, selection in just the high environment to increase the mean (\( \Delta H > 0 \)) generally results in increased sensitivity (\( \Delta s < 0 \)), which is a restatement of the Jinks-Connolly rule.
Similarly, the genetic correlation between sensitivity and performance in the low environment is

$$r_{L,s} = \frac{\phi - \rho_{H,L}}{\sqrt{1 + \phi^2 - 2\rho_{H,L}\phi}}$$  \hspace{1cm} (38.13e)$$

so that selection on $s$ increases the performance in the low environment when $\phi > \rho_{H,L}$, otherwise it decreases $\mu_L$. Proceeding in exactly the same fashion, the genetic correlations between mean performance on one hand and performance in the high and low environments on the other are, respectively,

$$r_{H,m} = \frac{\phi \rho_{H,L} + 1}{\sqrt{1 + \phi^2 + 2\rho_{H,L}\phi}}, \quad \text{and} \quad r_{L,m} = \frac{\phi + \rho_{H,L}}{\sqrt{1 + \phi^2 + 2\rho_{H,L}\phi}}$$  \hspace{1cm} (38.14)$$

Both of which are positive unless the genetic correlation between environments is negative.
Finally, the genetic correlation between mean performance $m$ and sensitivity $s$ is

$$r_{s,m} = \frac{\phi^2 - 1}{\sqrt{1 + 2\phi^2 + \phi^4 - 4\rho_{H,L}^2\phi^2}}$$

(38.15)

which is negative unless $\phi^2 > 1$. Hence, selection on sensitivity decreases mean performance ($\Delta s > 0 \rightarrow \Delta m < 0$), and selection on mean performance increases sensitivity ($\Delta m > 0 \rightarrow \Delta s < 0$) unless $\sigma^2(L) > \sigma^2(H)$. Rosielle and Hamblin caution not to over-interpret these two-environment results when multiple environments are considered, but their point is still well made.
We can easily incorporate joint selection on mean performance and sensitivity into a selection index. To slightly simplify matters, consider an index selecting on total performance over both environments (i.e., $2m$ in place of $m$) and on the sensitivity $s$, where $a$ is the weight (relative to total performance) placed on sensitivity,

$$I = 2m + a \cdot s = (g_H + g_L) + a(g_L - g_H) = \begin{cases} g_H + \left( \frac{1 + a}{1 - a} \right) g_L & a \neq 1 \\ g_L & a = 1 \end{cases} \quad (38.16)$$

The last step follows by recalling we can always rescale one of the index weights to one (Chapter 33). If total performance and sensitivity are given equal weight ($a = 1$), the index reduces to the breeding (or genotypic) value $g_L$ of performance in the low environment. Very small $a$ corresponds to selection on total performance, while very large $a$ corresponds to selection on sensitivity. If the genetic and phenotypic variances and covariances between low and high performance are known, then index selection theory can be used to obtain the Smith-Hazel weights for this index (Equation 33.18a).

Likewise, we could construct a restricted index giving the weights on $m$ and $s$ to maximize response in $m$ while keeping $s$ unchanged.
From the index selection notes, the restricted index weights are given by

\[ I_r = m - \left( \frac{\sigma(A_m, A_s)}{\sigma^2(A_s)} \right) s \]

The variance components become

\[ \sigma(A_m, A_s) = \sigma(H + L, L - H) \]
\[ = \sigma(H, L) - \sigma(H, H) + \sigma(L, L) - \sigma(L, H) \]
\[ = \sigma^2(L) - \sigma^2(H) \]

\[ \sigma^2(s) = \sigma(L - H, L - H) \]
\[ = \sigma^2(L) + \sigma^2(H) - 2\rho_A \sigma(L) \sigma(L) \]

Giving the restricted index selection weights as

\[ I_r = 2m - \left( \frac{\sigma^2(L) - \sigma^2(H)}{\sigma^2(L) + \sigma^2(H) - 2\rho_A \sigma(L) \sigma(H)} \right) s \]
Multiple-environment trails

- **Multiple-environment trails** (MET) is the main design used by plant breeders to deal with $G \times E$ with a set of pure (or nearly pure) lines.
- Thus far, we have presented environments as being discrete and reproducible, such as average performance in a growing region ($G \times l$) -- genotype x location.
- Another source of $G \times E$ is $G \times y$ ($G \times year$). This component is unique each year and thus unreproducible.
- Finally, can have $G \times l \times y$ (location-year interactions). Again, not reproducible.
- Ideal line: High $G \times l$ (targets specific growing regions), low $G \times y$, $G \times l \times y$ (Stability across years)
$G \times E$ interaction can be decomposed into contributions for locations, years, and location $\times$ year interactions,

$$
\sigma^2_{G \times E} = \sigma^2_{G \times \ell} + \sigma^2_{G \times y} + \sigma^2_{G \times y \times \ell}
$$

(38.17a)

As mentioned, the breeder may wish to exploit $G \times E$ in predictable environments while trying to mitigate it (though selection for stability) in unpredictable environments. Note that $\frac{\sigma^2_{G \times \ell}}{\sigma^2_{G \times E}}$ essentially represents the potential fraction of interaction variance due to predictable environmental factors, while the remainder likely represents unpredictable features. The relative contribution of these two components informs the breeder of their options (breeding for location and/or stability).
Equation 38.17a provides the motivation for multiple-environment trails (or MET) where varieties are scored for several years over several locations (Chapter 20). The importance of Equation 38.17a is that replication can reduce the noise from $G \times E$ when trying to assess genotypes. So see this, suppose that $n_r$ individual from a line are scored in a single environment, and the mean performance of these individuals is reported. The resulting residual error variance becomes $\sigma^2_e / n_r$. Likewise, if such replication occurs over $n_l$ locations (environments) and for $n_y$ years per location, then the $G \times E$ and environmental variance associated with the mean performance of a line becomes (Lonnquist 1964, Comstock and Moll 1973, Patterson et al 1977, Brennan and Byth 1979, Thompson and Cunningham 1979)

$$\frac{\sigma^2_{G \times E}}{n_l} + \frac{\sigma^2_{G \times E \times Y}}{n_y} + \frac{\sigma^2_{G \times E \times Y \times L}}{n_sn_l} + \frac{\sigma^2_e}{n_en_sn_y}$$  \hspace{1cm} (38.17b)

The key feature of Equation 38.17b is that suitable replication can reduce the contribution of any particular component of $\sigma^2_{G \times E}$ to the variance of a line mean, thereby increasing the heritability of the line mean (Equation 38.4). The idea of METs is to find those lines that perform well over some target populations of environments, recognizing that while locational correlations might be reasonably stable, the unpredictability in year-to-year (and hence location-by-year) interactions implies that the breeder must select for lines that perform well over some (largely unpredictable) distribution of environments.
Example 38.4. Atlin et al. (2001) use data from six different crops to show the benefits of replication across years and locations. Estimates of line, G × E components, and residual variance were obtained from (1) Atlin and McRae (1994), (2) Cullis et al. (1996), (3) Talbot (1984), (4) Cooper and Somrith (1997), and (5) Copper et al. (1999), and were as follows:

<table>
<thead>
<tr>
<th>Crop</th>
<th>Region</th>
<th>$\sigma^2_G$</th>
<th>$\sigma^2_{G \times \ell}$</th>
<th>$\sigma^2_{G \times y}$</th>
<th>$\sigma^2_{G \times \ell \times y}$</th>
<th>$\sigma^2_e$</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring barley</td>
<td>Canada</td>
<td>62</td>
<td>29</td>
<td>18</td>
<td>63</td>
<td>174</td>
<td>1</td>
</tr>
<tr>
<td>Spring Oat</td>
<td>Canada</td>
<td>122</td>
<td>58</td>
<td>21</td>
<td>53</td>
<td>178</td>
<td>1</td>
</tr>
<tr>
<td>Wheat</td>
<td>Australia</td>
<td>23</td>
<td>8</td>
<td>9</td>
<td>53</td>
<td>87</td>
<td>2</td>
</tr>
<tr>
<td>Winter wheat</td>
<td>UK</td>
<td>99</td>
<td>7</td>
<td>22</td>
<td>113</td>
<td>128</td>
<td>3</td>
</tr>
<tr>
<td>Potatoes</td>
<td>UK</td>
<td>9780</td>
<td>2980</td>
<td>2630</td>
<td>14960</td>
<td>18790</td>
<td>3</td>
</tr>
<tr>
<td>Lowland rice</td>
<td>Thailand</td>
<td>198</td>
<td>82</td>
<td>18</td>
<td>199</td>
<td>178</td>
<td>4</td>
</tr>
<tr>
<td>Lowland rice</td>
<td>Thailand</td>
<td>60</td>
<td>3</td>
<td>49</td>
<td>259</td>
<td>440</td>
<td>5</td>
</tr>
</tbody>
</table>
Recalling Equation 38.17b, the heritability of the line means is given by

\[ h^2_z = \frac{\sigma^2_G}{\sigma^2_z}, \quad \text{with} \quad \sigma^2_z = \sigma^2_G + \frac{\sigma^2_{G \times \ell}}{n_\ell} + \frac{\sigma^2_{G \times y}}{n_y} + \frac{\sigma^2_{G \times \ell \times y}}{n_\ell n_y} + \frac{\sigma^2_e}{N} \]

where \( N = n_r n_\ell n_y \). With increased replication, \( \sigma^2_z \) approaches \( \sigma^2_G \), and hence the heritability of line means can be made to approach one by using sufficient replication. Using the above values, Atlin et al. (2001) calculated the estimated heritabilities under different designs (different allocation of lines over locations \( n_\ell \), years \( n_y \), and replications per site \( n_r \)). The last column gives the ratio of the single-replication heritability with that for the most complete design considered here (5,5,2 = 5 replicates per location, 5 locations, 2 years per location),

<table>
<thead>
<tr>
<th>Crop</th>
<th>1,1,1</th>
<th>1,2,1</th>
<th>4,1,1</th>
<th>5,5,2</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring barley</td>
<td>0.18</td>
<td>0.29</td>
<td>0.29</td>
<td>0.71</td>
<td>3.9</td>
</tr>
<tr>
<td>Spring Oat</td>
<td>0.28</td>
<td>0.42</td>
<td>0.31</td>
<td>0.79</td>
<td>2.8</td>
</tr>
<tr>
<td>Wheat</td>
<td>0.13</td>
<td>0.22</td>
<td>0.20</td>
<td>0.63</td>
<td>4.8</td>
</tr>
<tr>
<td>Winter wheat</td>
<td>0.27</td>
<td>0.40</td>
<td>0.36</td>
<td>0.79</td>
<td>2.9</td>
</tr>
<tr>
<td>Potatoes</td>
<td>0.20</td>
<td>0.32</td>
<td>0.28</td>
<td>0.72</td>
<td>3.6</td>
</tr>
<tr>
<td>Lowland rice</td>
<td>0.29</td>
<td>0.44</td>
<td>0.37</td>
<td>0.80</td>
<td>2.8</td>
</tr>
<tr>
<td>Lowland rice</td>
<td>0.07</td>
<td>0.13</td>
<td>0.13</td>
<td>0.49</td>
<td>7.0</td>
</tr>
</tbody>
</table>
Balancing rate of response with accuracy in METs

Adding extra years decreases $G \times y$ and $G \times y \times l$ interaction variances, increasing precision. However, it occurs at the potential cost of reducing the rate of response.

As we have seen, many family-based breeding schemes take $c$ generations/years per cycle. If $k$ extra years of testing are added, the yearly rate of response becomes

$$
\Delta R = \bar{i} \frac{\sigma_G^2}{\sigma(\bar{z}_k)} \left( \frac{1}{c + k} \right)
$$

Increasing $k$ increases cycle time, decreasing rate

Increasing $k$ results in a decrease in the family variance.
\[ \Delta R = 7 \frac{\sigma^2_G}{\sigma(\bar{x}_k)} \left( \frac{1}{c + k} \right) \]

where

\[ \sigma^2(\bar{x}_k) = \sigma^2_G + \frac{\sigma^2_{G \times \ell}}{n_\ell} + \frac{\sigma^2_{G \times y}}{(k + 1)} + \frac{\sigma^2_{G \times \ell \times y}}{(k + 1)n_\ell} + \frac{\sigma^2_e}{(k + 1)n_\ell n_r} \] (38.18a)

Equation 38.18a assumes that the number of locations per year and the number of replicates per location-year combination remains constant, with only the number of years of testing \((n_y = k + 1)\) changing. The ratio of rate of gain for one versus \(k\) additional years of testing becomes

\[ \frac{\Delta R_k}{\Delta R_1} = \left( \frac{c}{c + k} \right) \frac{\sigma(\bar{x}_1)}{\sigma(\bar{x}_k)} \] (38.18b)

Comstock and Moll (1963) consider the most extreme case where \(\sigma^2_{G \times y}\) dominates all other interaction terms. In this case, for one and 2 years of testing, we have

\[ \sigma^2(\bar{x}_1) \simeq \sigma^2_G + \sigma^2_{G \times y}, \quad \sigma^2(\bar{x}_2) \simeq \sigma^2_G + \frac{\sigma^2_{G \times y}}{2} \]

Even if \(\sigma^2_{G \times y}\) accounts for 95% of \(\sigma^2_G + \sigma^2_{G \times y}\) for two years of replication \(\sigma(\bar{x}_1)/\sigma(\bar{x}_2) = 1.3\). Substituting into Equation 38.18b shows that for the rate of response to be increased by replication requires a cycle time of \(c \geq 3\) years. For four years of replication \(\sigma(\bar{x}_1)/\sigma(\bar{x}_4) = 1.9\), requiring a cycle of at least four years for replication to increase the rate of response. Hence while replication over many years increases precision, this is often more than offset by the longer cycle time.
Example 38.6. Using an extensive dataset on soybean yield in Ontario, Yan and Rajcan (2003) examined the effects using one, two, three, and five years worth of data in predicting yield. A partial set of their data is shown in the table below, which compares the actual performance (measured by the variance-scaled BLUP estimate of the genotype value, or t-BLUP, Yan et al. 2002) in a given year with the predicted performance based on the previous one, two, three, and five years worth of data. Prediction performance was measured by the correlation between estimates of the genotypic effects of lines in the focal year and the predicted value based on using results from previous year(s).

<table>
<thead>
<tr>
<th>Year</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.57</td>
<td>0.63</td>
<td>0.61</td>
<td>0.61</td>
</tr>
<tr>
<td>1999</td>
<td>0.57</td>
<td>0.57</td>
<td>0.57</td>
<td>0.65</td>
</tr>
<tr>
<td>1998</td>
<td>0.56</td>
<td>0.68</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>1997</td>
<td>0.51</td>
<td>0.56</td>
<td>0.52</td>
<td>0.53</td>
</tr>
<tr>
<td>1996</td>
<td>0.47</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td>Average</td>
<td>0.54</td>
<td>0.59</td>
<td>0.58</td>
<td>0.59</td>
</tr>
</tbody>
</table>
As the above table shows, the conclusion is that using two years of data does slightly better than using a single year, but that adding additional years results in no further improvement. Yan and Rajcan also examined another measure, the number of genotypes that can be decisively evaluated (allowing them to be judged as significantly inferior or superior, as indicated by their t-BLUP values being greater than 2 in absolute value, Yan et al. 2002). As shown below, the number of decisively evaluated genotypes increases by adding additional years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of genotypes</th>
<th>Number of years of testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>112</td>
<td>1</td>
</tr>
<tr>
<td>1999</td>
<td>112</td>
<td>2</td>
</tr>
<tr>
<td>1998</td>
<td>112</td>
<td>3</td>
</tr>
<tr>
<td>1997</td>
<td>104</td>
<td>4</td>
</tr>
<tr>
<td>1996</td>
<td>90</td>
<td>5</td>
</tr>
</tbody>
</table>

While replication over several years does not significantly improve prediction ability for any given line in subsequent years, it does help the breeder in culling the least desirable, and choosing the most desirable, lines.
Participatory Breeding and $G \times E$

- The notion of regional testing directly relates to a recent movement in plant breeding --- *participatory breeding*, with local farmers actively involved in the selection of new lines.
- The motivation for this approach was the concern that *formal plant breeding* (FPB), namely multiple-environment traits of new lines conducted through national and international centers, was not producing products for low-input farmers working in marginal environments.
- The success of FPB is clearly undeniable, but the concern was that much of its focus was for stable crops widely-adaptive to high yield environments (where most of the profit occurs), potentially at the expense of marginal farmers whose yield, while critical to them and their families, is economically quite small.
- In *participatory plant breeding* (PPB), local farmers help in the breeding, either through input on selectively desirable traits and/or evaluation of performance in their own fields.
From a strictly breeding standpoint, there are three issues when comparing PPB and FPB. The first is that the targets of selection under formal plant breeding may be different from the targets of selection desired by low-input farmers. From this standpoint, there is clearly a significant benefit from seeking input from the ultimate end-users, the farmers themselves (e.g., Ceccarelli et al. 2000). The remaining two issues can (again) be rephrased in a direct versus correlated response framework. The target population of environments (TPE) are low-input growing situations, often under considerable stress (relative to high-input systems). Thus, PPB occurs within the TPE, while the fields at research stations typically do not, although this not be the case (for example, field trials can be made under conditions of enforced stress). Balancing this are the higher heritabilities for line means that can be obtained under FPB by using highly replicated experimental designs (e.g., Example 38.4) and greater access to more diverse genotypes (and hence a larger $\sigma_G^2$). These various tradeoffs can be placed in terms of Equation 30.22a, giving the ratio of the correlated response (FPB) to the direct response (PPB) as

$$\frac{\text{Response under FPB}}{\text{Response under PPB}} = \frac{CR_X}{R_X} = \left(\frac{\bar{t}_{FPB}}{\bar{t}_{PPB}}\right) \rho_G \left(\frac{h_{FPB}}{h_{PPB}}\right)$$

(38.19)

What is not included in Equation 38.19 is accounting for differences in the targets of selection desired by the farmer versus those selected for the breeder.
Full multivariate response

• Finally, treating traits measured in k discrete environments as a set of correlated traits, selection response (on either individual components = environments or on some index) fully from the theory of multivariate response.
• Here, the matrix $G$ are the genetic variances and covariances of a trait over the k (reproducible) environments (such as locations)
• As shown, the elements of $G$ can be estimated using either clones or relatives (such as sibs) using standard methods for correlated traits.
• Response is given by the multivariate breeder’s equation.
Structured Covariance Matrices and Selection Response

We started our introduction of selection under \( G \times E \) in Chapter 38 by placing it within the framework of selection on correlated traits. Under this framework, the expected vector of responses over a series of environments given selection within each (or some subset) is given by \( GP^{-1}S \). As mentioned in Chapter 38, the problem with widely applying the multivariate breeder’s equation is the difficulty of estimating \( G \) with any precision for more than two environments. The concern is that \( G \) matrix is completely unstructured, requiring a large number of covariances to be estimated with some precision. However, when \( G \) has some structure, we now have a potential solution, namely using factor-analytic estimates of \( G \) as obtained from mixed-models (Table 39.3). In particular, treating environments as fixed effects (as the concern is response over a specified set of environments, for example locations), then AMMI-type mixed models (taking \( G \) random) offer estimates of the covariance matrix \( G \).

If the concern is predicting selection response over a set of environments (such as particular locations), then if \( z \) is a vector of line means over this set of environments and \( \mu \) is the vector of means over this environment, the vector \( S \) of selection differentials is given by

\[
S = E_s(z - \mu)
\]

(39.39a)

where the expectation is taken over those lines selected. Thus, the \( j \)th entry in \( S \) is just

\[
s_j = \mu_j^* - \mu_j
\]

(39.39b)

where \( \mu_j^* \) is the mean of the selected lines in that environment.
The resulting selection response follows from the multivariate breeder’s equation,

\[ R = GP^{-1}S = G (G + V_e)^{-1} S \]  

(39.39c)

where \( G \) is estimated using one of the models in Table 39.3 and \( V_e \) is the covariance matrix of the residuals. Standard results index selection theory also apply. Suppose the goal is the expected response over \( n \) environments, where the desired gains are weighted by

\[ I = a_1 g_1 + \cdots + a_n g_n = \sum_{i=1}^{n} a_i g_i = a^T g \]

The response in \( I \) is simply

\[ a^T R = a^T G (G + V_e)^{-1} S \]  

(39.40a)

Likewise, a factor-analytic approximation of a structured \( G \) matrix could also be used in a Smith-Hazel index to obtain the weights \( b \) to maximinze the response in \( I \) (Chapter 33), namely

\[ b = P^{-1} Ga = (G + V_e)^{-1} Ga \]  

(39.40b)