The development of a bivariate mixed model approach for plant survival data

http://agritourismaustralia.com.au
Aanandini Ganesalingam
School of Plant Biology, 
University of Western Australia

Adjunct Assoc. Prof. Alison Smith
School of Plant Biology, 
University of Western Australia

Department of Industry & Investment

Dr. Cameron Beeck
The UWA Institute of Agriculture, 
University of Western Australia

Winthrop Prof. Wallace Cowling
The UWA Institute of Agriculture, 
University of Western Australia

Prof. Brian Cullis
School of Mathematics and Applied Statistics, Faculty of Informatics 
University of Wollongong
Outline

* Data set background
  * Bivariate model
  * Model Results
  * Findings
  * Conclusion

Photo supplied by Wayne Burton, Victoria DPI
Plant survival is the main measure of disease resistance

* Often requires multiple measurements before & after infection
* Counts are used to form an index
To develop a bivariate mixed model approach for plant survival data

Motivating example

2009 blackleg disease resistance trials for canola
Caused by *Leptosphaeria maculans*

* Crown cankers are the main cause of plant death
* In W.A. yield losses in 1998 and 1999 were $20M and $50M
  
  (Khangura & Barbetti 2001)
Blackleg disease resistance

* Determined by counting the number of plants at emergence & maturity for each plot
* Historically we convert to:

$$\% \text{ survival} = \frac{\text{maturity}}{\text{emergence}} \times 100$$

* The bivariate approach uses both plant counts as two ‘traits’
Disease resistance at maturity

Disease susceptible

Disease resistant

Photo supplied by Canola Breeders Western Australia Pty Ltd
**Motivation towards bivariate analysis**

<table>
<thead>
<tr>
<th>Historical</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Single derived variable: %survival</td>
<td>* Two traits: emergence &amp; maturity counts</td>
</tr>
<tr>
<td>* Assumes counts at emergence are without error</td>
<td>* Model error - spatial field trend for each trait</td>
</tr>
<tr>
<td>* Confounds errors in emergence and maturity traits</td>
<td>* Identification trait based outliers</td>
</tr>
<tr>
<td></td>
<td>* Examine individual trait genetic effects</td>
</tr>
</tbody>
</table>
York, Western Australia - 2009 disease nursery site
  * Plots of approx. 100 plants per variety
  * 3 columns x 79 rows
* Log of emergence & maturity counts
  * Spatial mixed model approach of Gilmour et al. (1997)
* Modelling and analysis undertaken in ASReml – R (Butler et al. 2009)
Spatial modelling of individual traits

\[ y_j = X\tau_j + Z_v u_{vj} + Z_b u_{bj} + e_j \]

* \( y_j \) is the \( n \times 1 \) vector of data (\( j=1 \) for emergence, \( j=2 \) for maturity)
* \( \tau_j \) is the vector of fixed effects (overall site mean)
* \( u_{vj} \) is the \( m \times 1 \) vector of random variety effects
* \( u_{bj} \) is the \( b \times 1 \) vector of random block effects
* \( e_j \) is the vector of residuals ordered as per the data vector
Spatial effects for the emergence mixed model

Initial trial variogram

Resulting trial variogram

+ Random Row
Bivariate analysis

* Combines individual emergence and maturity mixed models
* Retains spatial terms from base modeling
Bivariate analysis

\[ y = X^*\tau + Z_v^*u_v + Z_b^*u_b + Z_o^*u_o + e \]

- \( y = (y_1', y_2') \) is the combined vector of data across sampling times
- \( \tau \) is the vector of fixed effects
- \( u_v = (u_{v1}', u_{v2}') \) is the \( 2m \times 1 \) vector of random variety effects
- \( u_b = (u_{b1}', u_{b2}') \) is the \( 2b \times 1 \) vector of random block effect
- \( u_o \) includes any random effects determined in the spatial modeling
- \( e = (e_1', e_2') \) is the vector of errors ordered as for the data vector
Bivariate model - assumptions

* Variety effects

\[
\begin{pmatrix}
u_{v1} \\ u_{v2}
\end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_{v11} & \sigma_{v21} \\ \sigma_{v21} & \sigma_{v22} \end{bmatrix} \otimes I_m \right)
\]

Covariance between emergence and maturity

* Errors

\[
\begin{pmatrix}e_1 \\ e_2\end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \otimes \Sigma_c \otimes \Sigma_r \right)
\]

Covariance between emergence and maturity
Model Comparisons

* Spatial model:
  REML estimates of genetic, error & row autocorrelation variance parameters

<table>
<thead>
<tr>
<th>Trait</th>
<th>Variety</th>
<th>Error</th>
<th>ρ_r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log eme</td>
<td>0.400</td>
<td>0.382</td>
<td>0.723</td>
</tr>
<tr>
<td>Log mat</td>
<td>0.511</td>
<td>0.299</td>
<td>0.218</td>
</tr>
</tbody>
</table>

* Bivariate model:
  REML estimates of genetic, error & row autocorrelation variance parameters

<table>
<thead>
<tr>
<th>Trait</th>
<th>Variety</th>
<th>Error</th>
<th>ρ_r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log eme</td>
<td>0.354</td>
<td>0.329</td>
<td>0.362</td>
</tr>
<tr>
<td>Log mat</td>
<td>0.493</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.71</td>
<td>0.59</td>
<td></td>
</tr>
</tbody>
</table>

* Correlation between genetic effects and errors are also shown
Variety predictions

- Best linear unbiased predictors (BLUPS) of variety means
  - At emergence
  - At maturity
- Difference between predicted maturity & emergence means for variety $k$ on the analyzed scale can be back transformed to the % survival scale

\[
\exp(\hat{\pi}_{2k} - \hat{\pi}_{1k}) = \frac{\exp(\hat{\pi}_{2k})}{\exp(\hat{\pi}_{1k})}
\]
Findings

- Individual analysis of 9 blackleg disease nurseries from the 2009 growing season
1. Trait spatial modeling

- Spatial components Gilmour et al. (1997) differed for each trait
  - Local trend, global trend (extraneous variation)
- The number of outliers also differed between traits
  - Emergence trait having the largest number
- Such differences not obvious from the historical approach
2. Variation for emergence

* Demonstrated there is large variation between varieties for emergence
  * Variation in seed source
    * Seed lot factors: age of seed, storage (Ellis and Roberts 1980)
    * ‘Juvenile’ blackleg (Li et al., 2007)
  * Emergence & maturity have different causes of variation but are strongly correlated at most sites
    * averaged 0.57 (range 0.22 to 0.94)
3. Selection for disease resistance

* Emergence & maturity BLUPs provide 3 indices for selection
  * Emergence counts
  * Maturity counts
  * % Survival values (analogous to historical method)
* Method of selection is up to the breeder’s discretion
The bivariate analysis is statistically more accurate than the percentage survival approach.

- Model error individually for each trait
- Identification trait based outliers
- Examine individual trait genetic effects
Future work……..

* A bivariate approach within a Multi Environment Trial (MET) framework
References

Acknowledgements

* National Blackleg Group for the use of the 2009 blackleg disease resistance data set
* Financial support from the Grains Research and Development Corporation of Australia (GRDC)
* Bayer Crop Science PhD Research Scholarship