Crown Rot Tolerance in Durum Wheat

Australasian Applied Statistics Conference — Rotorua 2018

Jess Meza[1], Ky Mathews[1], Gururaj Kadkol[2], Steven Simpfendorfer[2], Steve Harden[2]

[1]Centre for Bioinformatics and Biometrics
National Institute for Applied Statistics Research Australia
University of Wollongong
&
[2]NSW Department of Primary Industries

Crown Rot (CR)

Caused by the fungus *Fusarium pseudograminearum*, the presence of the pathogen within the plant stem limits water movement, reducing yield.

It is generally thought that all durum wheat varieties are equally susceptible to CR.

However, some varieties are more resistant/tolerant to CR and produce better yield under CR disease pressure.

Aim

Breeding for CR resistance/tolerance is a major breeding objective in the durum breeding program. Understanding and characterising the existing variation within the germplasm is a useful first step.

Here we investigate genetic variation for CR resistance/tolerance with an emphasis on current Durum Breeding Australia elite durum material.

- **Tolerance** is the ability of the plant to yield despite infection
- **Resistance** is the ability of the plant to restrict infection

Motivating Example

- 6 trials at Tamworth Agricultural Institute over 2012–2017

Motivating Example

- 6 trials at Tamworth Agricultural Institute over 2012–2017
- 49 genotypes including released varieties and advanced breeding lines, with the inclusion of tolerant bread wheat check varieties Wylie and Suntop, and susceptible durum check Bellaroi

Motivating Example

- 6 trials at Tamworth Agricultural Institute over 2012–2017
- 49 genotypes including released varieties and advanced breeding lines, with the inclusion of tolerant bread wheat check varieties Wylie and Suntop, and susceptible durum check Bellaroi
- 2 Crown Rot inoculation treatment regimes

Motivating Example

- 6 trials at Tamworth Agricultural Institute over 2012–2017
- 49 genotypes including released varieties and advanced breeding lines, with the inclusion of tolerant bread wheat check varieties Wylie and Suntop, and susceptible durum check Bellaroi
- 2 Crown Rot inoculation treatment regimes
- 2-level factorial design (2012–2013) or split-plot design (2014–2017)

Motivating Example

- 6 trials at Tamworth Agricultural Institute over 2012–2017
- 49 genotypes including released varieties and advanced breeding lines, with the inclusion of tolerant bread wheat check varieties Wylie and Suntop, and susceptible durum check Bellaroi
- 2 Crown Rot inoculation treatment regimes
- 2-level factorial design (2012–2013) or split-plot design (2014–2017)
- replication of 3 (2012–2014) or 4 (2015–2017)

Motivating Example

- 6 trials at Tamworth Agricultural Institute over 2012–2017
- 49 genotypes including released varieties and advanced breeding lines, with the inclusion of tolerant bread wheat check varieties Wylie and Suntop, and susceptible durum check Bellaroi
- 2 Crown Rot inoculation treatment regimes
- 2-level factorial design (2012–2013) or split-plot design (2014–2017)
- replication of 3 (2012–2014) or 4 (2015–2017)
- later sowing date adopted from 2015 onwards

Motivating Example

- 6 trials at Tamworth Agricultural Institute over 2012–2017
- 49 genotypes including released varieties and advanced breeding lines, with the inclusion of tolerant bread wheat check varieties Wylie and Suntop, and susceptible durum check Bellaroi
- 2 Crown Rot inoculation treatment regimes
- 2-level factorial design (2012–2013) or split-plot design (2014–2017)
- replication of 3 (2012–2014) or 4 (2015–2017)
- later sowing date adopted from 2015 onwards
- yield is the trait of interest

Crown Rot Inoculation Treatment Regime (CR)

All trials include a '+'/'−' Crown Rot inoculation treatment regime CR with two levels:

'−'

The '−' regime is the control.

'+'

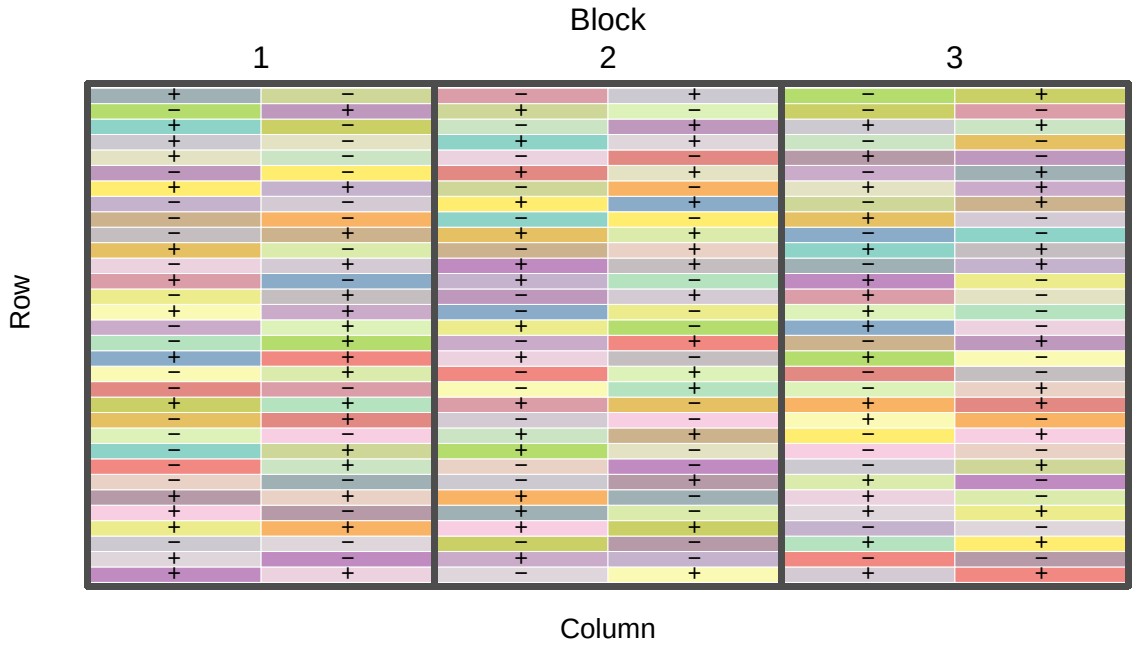
The '+' regime is the stress, and has sterilised grain infected with CR mixed with the seed for sowing at a rate of 2g per metre of row.

Experimental Design

2-level factorial design

(2012–2013)

- 3 blocks in the column-wise direction
- treatments as defined to be every combination of `Genotype` level and `CR` treatment regime allocated to plots within blocks

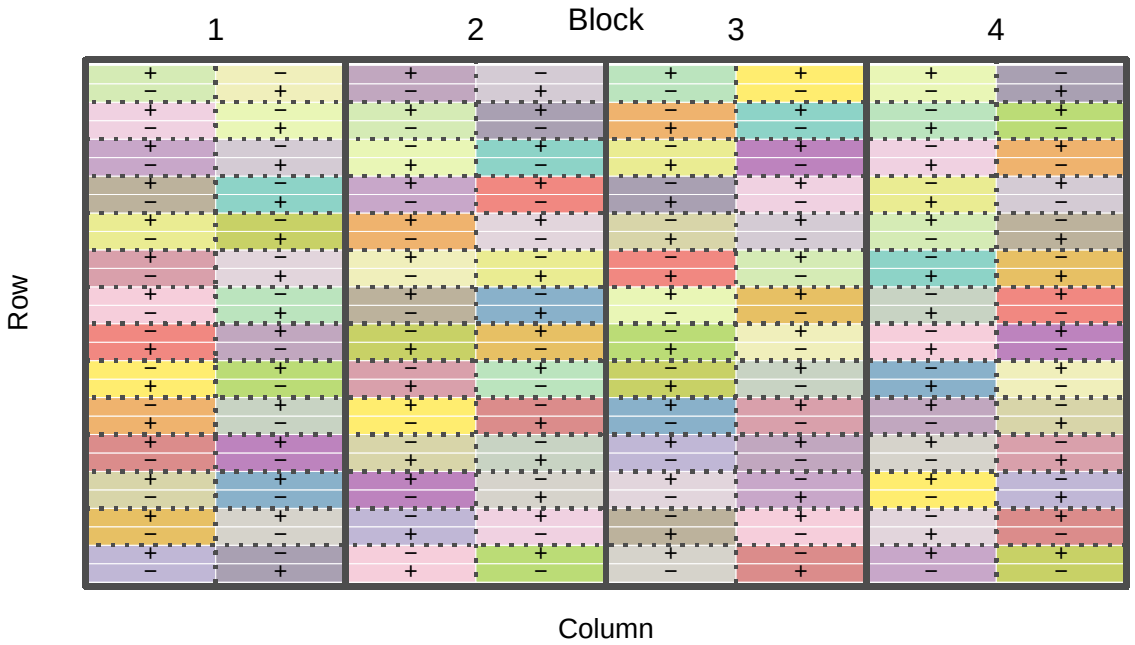


Experimental Design

Split-plot design

(2014–2017)

- 4 blocks in the column-wise direction
- `Genotype` allocated to main plots within blocks
- `CR` treatment regime applied to subplots within mainplots



Statistical Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

where
$\mathbf{y}$ is a n-vector of yield observations
$\boldsymbol{\tau}$ is a p-vector of fixed effects
$\mathbf{X}$ is the $n \times p$ design matrix associated with $\boldsymbol{\tau}$
$\mathbf{u}_g$ is a b-vector of genetic effects
$\mathbf{Z}_g$ is the $n \times b$ design matrix associated with $\mathbf{u}_g$
$\mathbf{u}_p$ is a q-vector of non-genetic (peripheral) effects
$\mathbf{Z}_p$ is the $n \times q$ design matrix associated with $\mathbf{u}_p$
$\mathbf{e}$ is a n-vector of residuals

and

$$\begin{bmatrix} \mathbf{u}_g \\ \mathbf{u}_p \\ \mathbf{e} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{G}_g & 0 & 0 \\ 0 & \mathbf{G}_p & 0 \\ 0 & 0 & \mathbf{R} \end{bmatrix} \right)$$

Linear Mixed Model (LMM)

Design Tableau – Smith & Cullis (2018)

Treatment Factors

{1, Genotype(27), CR(2)}

Plot Factors

{U, Block(4), Mainplot(27), Subplot(2)}

Treatment Structure

1/(Genotype∗CR) = 1 + Genotype + CR + Genotype:CR

Plot Structure

U/Block/Mainplot/Subplot = U + Block + Block:Mainplot
+ Block:Mainplot:Subplot

Linear Mixed Model (LMM)

Initial LMM for a single trial with split-plot design:

Source	Model Term	Fixed/Random	Variance Model
Mean	`1[U]`	F	–
`Genotype`	`Genotype`	F	–
`CR`	`CR`	F	–
`Genotype:CR`	`Genotype:CR`	F	–
`Block`	`Block`	R	$\sigma_b^2\mathbf{I}$
`Block:Mainplot`	`Block:Mainplot`	R	$\sigma_{bm}^2\mathbf{I}$
`Block:Mainplot:Subplot`	`units`	R	$\sigma^2\mathbf{I}$

Linear Mixed Model (LMM)

We extend this model to better suit the aims of the experiment, and to provide a more appropriate fit for the data.

- first-order autoregressive model (AR1 $\times$ AR1)
- appropriate linear and random row and column terms
- to predict the '+'/'−' contrast and obtain Best Linear Unbiased Predictors (BLUPs) fit `Genotype` and `Genotype:CR` as random effects

Modelling Tolerance and the Factor Analytic Model

Lemerle et al. (2006) and Smith et al. (2001)

For a single trial with '+'/'−' treatments

$$\begin{bmatrix} \mathbf{u}_{g-} \\ \mathbf{u}_{g+} \end{bmatrix} = \begin{bmatrix} \lambda_{g-}\,\mathbf{f}_g \\ \lambda_{g+}\,\mathbf{f}_g \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\delta}_g \end{bmatrix}$$
$$= (\boldsymbol{\lambda}_g \otimes \mathbf{I}_m)\,\mathbf{f}_g + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\delta}_g \end{bmatrix}$$

with covariance matrix

$$\text{var}(\mathbf{u}_g) = \left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}\right) \otimes \mathbf{I}_m$$

Modelling Tolerance and the Factor Analytic Model

The main effect of `Genotype` is dropped from the model.

The control ('−') nested within `Genotype` is considered to be the baseline and the variance component of the stress ('+') is set to zero.

Following Cullis & Smith (2016), we consider `Genotype` by `CR` as a genetic regression of $\mathbf{u}_{g_+}$ on $\mathbf{u}_{g_-}$ in which $\boldsymbol{\delta}_g$ is the deviation from the regression.

$$\mathbf{u}_{g_+} = \left(\frac{\lambda_{g_+}}{\lambda_{g_-}}\right)\mathbf{u}_{g_-} + \boldsymbol{\delta}_g$$

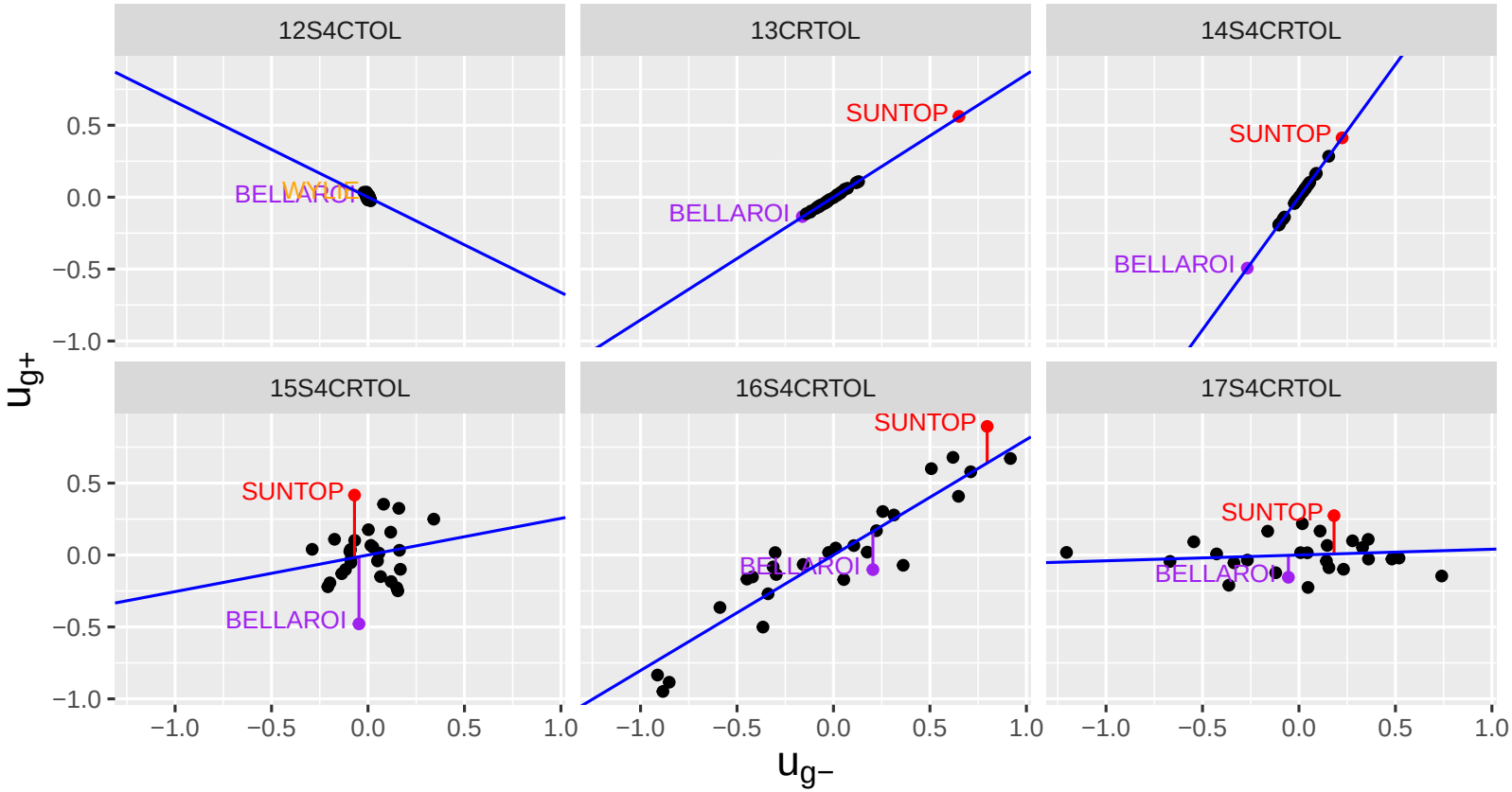
Linear Mixed Model (LMM)

Final LMM for a single trial with split-plot design:

Source	Model Term	Fixed/Random	Variance Model
Mean	`1[U]`	F	–
lin(Column)	*lin(Column)*	F	–
lin(Row)	*lin(Row)*	F	–
`Genotype`	`Genotype`	–	–
`CR`	`CR`	F	–
`Genotype:CR`	`rr(CR,1):Genotype` + `at(CR, "+"):Genotype`	R	$\left(\mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Psi}\right) \otimes \mathbf{I}_m$
Column	*Column*	R	$\sigma_c^2\mathbf{I}$
Row	*Row*	R	$\sigma_r^2\mathbf{I}$
`Block`	`Block`	R	$\sigma_b^2\mathbf{I}$
`Block:Mainplot`	`Block:Mainplot`	R	$\sigma_{bm}^2\mathbf{I}$
`Block:Mainplot:Subplot`	`ar1(Column):ar1(Row)`	R	$\sigma^2\mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r$

Single Site '+' vs '−' BLUPs

Trial-specific genetic regression lines $= \frac{\lambda_{g+}}{\lambda_{g-}}$.



MET Models

In the MET we need to account for `Genotype` by `CR` by `Trial` interaction (G $\times$ CR $\times$ T).

Create `et` = `CR`$\wedge$`Trial`, with 6 levels (2 treatment regimes in each of 3 trials).

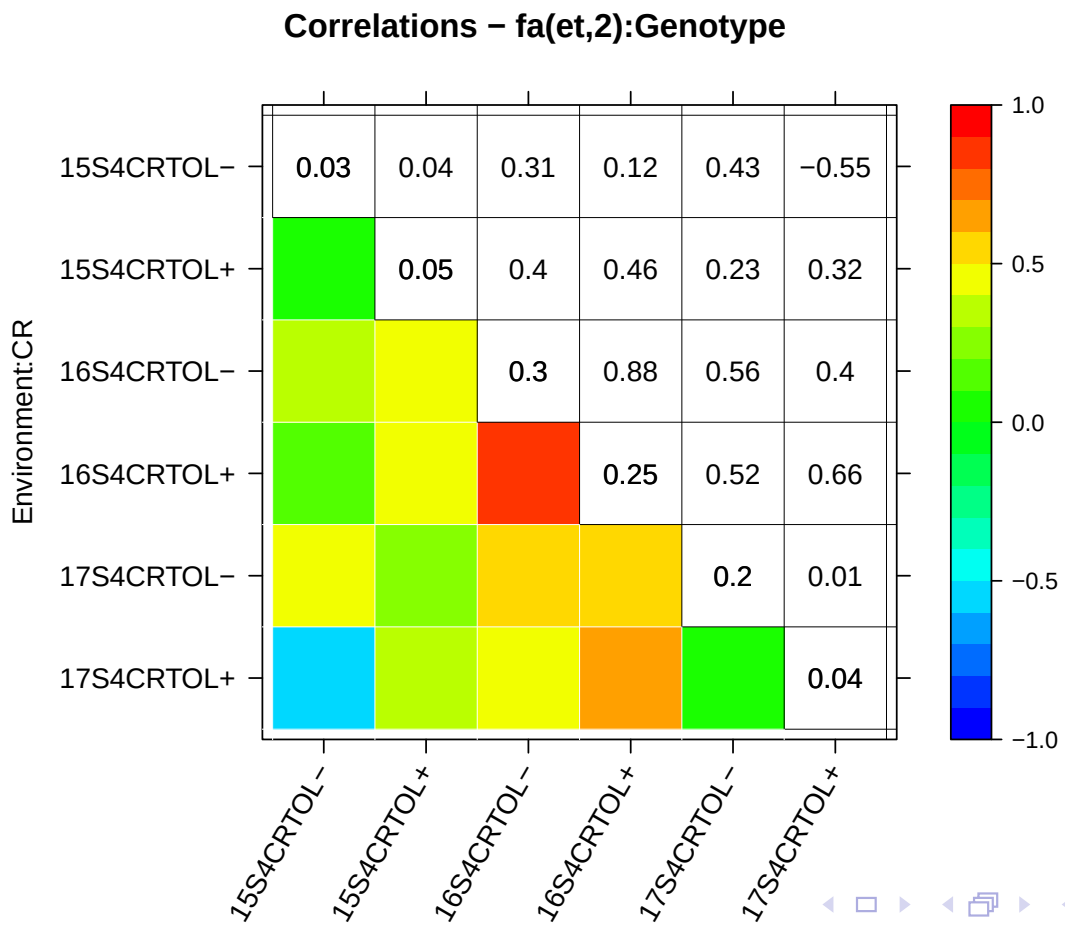
Models to investigate:

- Compound symmetry
- Unstructured
- FA(k), $k = 1, 2, 3$

Model	loglik	t	AIC
CS	291.98	4	-575.96
US	325.38	21	-608.80
FA(2)	**322.74**	**17**	**-611.50**
FA(3)	325.51	21	-609.01

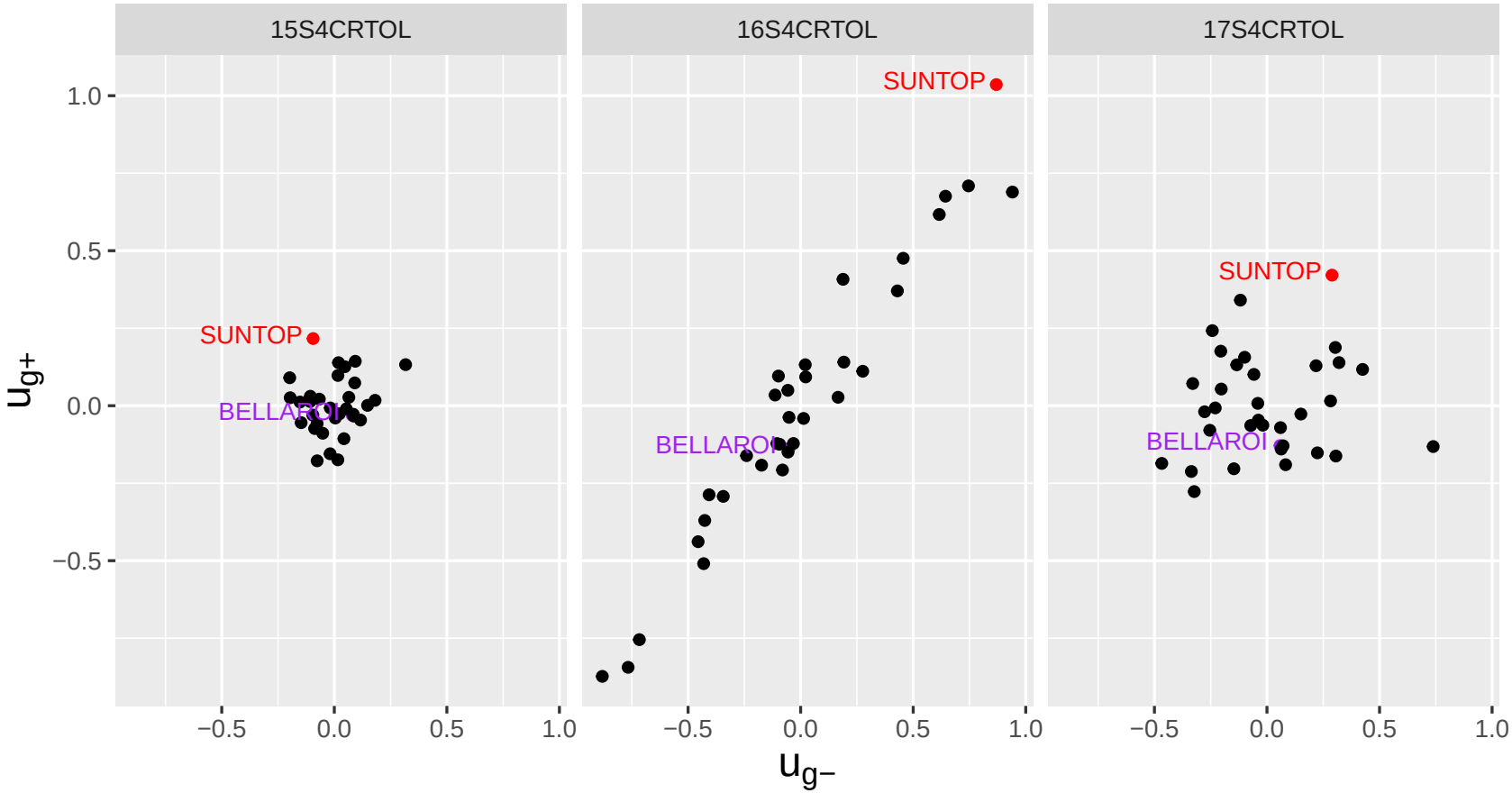
FA2 Model: between et correlation matrix

2015–2017 MET



MET '+' vs '−' BLUPs

Trial-specific genetic regression lines = ???



How do we get the genetic regression for the MET?

- determine the conditional distribution
- generalised inverse needed as variance matrix is singular (some elements of Ψ are zero)
- calculate the lack-of-fit effects
- marginal distribution required to estimate genetic regression for individual years

Conditional distribution of $\mathbf{u}_{g_+}|\mathbf{u}_{g_-}$

General Case:

Theorem 3.2.4 (Mardia et al. (1979)).

For

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1^\top & \mathbf{x}_2^\top \end{pmatrix}^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

such that $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1^\top & \boldsymbol{\mu}_2^\top \end{pmatrix}^\top$ *and* $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, *the conditional distribution of* $\mathbf{x}_2$ *given* $\mathbf{x}_1$ *is*

$$\mathbf{x}_2|\mathbf{x}_1 \sim \mathcal{N}\left(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)$$

where $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$ *is the matrix of regression coefficients.*

Conditional distribution of $\mathbf{u}_{g_+}|\mathbf{u}_{g_-}$

Specific Case:

Applying these results and recalling that the vector of random genetic effects $\mathbf{u}_g$ is distributed as multivariate Gaussian, with zero means and separable variance structure $\text{var}(\mathbf{u}_g) = \boldsymbol{\Sigma} \otimes \mathbf{I}_m$

$$\left[\begin{array}{c} \mathbf{u}_{g_-} \\ \hline \mathbf{u}_{g_+} \end{array}\right] \sim \mathcal{N}\left(\left[\begin{array}{c} \mathbf{0} \\ \hline \mathbf{0} \end{array}\right], \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{--} & \boldsymbol{\Sigma}_{-+} \\ \hline \boldsymbol{\Sigma}_{+-} & \boldsymbol{\Sigma}_{++} \end{array}\right] \otimes \mathbf{I}_m\right)$$

Thus the conditional distribution is Multivariate Normal

$$\mathbf{u}_{g_+}|\mathbf{u}_{g_-} \sim \mathcal{N}\left(\boldsymbol{\Sigma}_{+-}\boldsymbol{\Sigma}_{--}^{-1}\mathbf{u}_{g_-}, \boldsymbol{\Sigma}_{++} - \boldsymbol{\Sigma}_{+-}\boldsymbol{\Sigma}_{--}^{-1}\boldsymbol{\Sigma}_{-+}\right)$$

where $\boldsymbol{\Sigma}_{+-}\boldsymbol{\Sigma}_{--}^{-1}$ is the matrix of regression coefficients.

Generalised Inverse

The variance matrix $\boldsymbol{\Sigma}$ is singular so a generalised inverse will be required. Similar results hold using generalised inverses for the case of singular distributions

$$\mathbf{u}_{g_+}|\mathbf{u}_{g_-} \sim \mathcal{N}\left(\boldsymbol{\Sigma}_{+-}\boldsymbol{\Sigma}_{--}^{+}\mathbf{u}_{g_-}, \boldsymbol{\Sigma}_{++} - \boldsymbol{\Sigma}_{+-}\boldsymbol{\Sigma}_{--}^{+}\boldsymbol{\Sigma}_{-+}\right)$$

where $\boldsymbol{\Sigma}_{--}^{+}$ is the Moore-Penrose generalised inverse of $\boldsymbol{\Sigma}_{--}$.

As $\boldsymbol{\Sigma}$ is a partitioned matrix, we calculate the Moore-Penrose generalised inverse according to Hung & Markham (1975).

Lack-of-Fit Effects

We obtain the system of equations for the genetic regression

$$E\begin{bmatrix}\mathbf{u}_{g_{1+}}|\mathbf{u}_{g_{1-}},\mathbf{u}_{g_{2-}},\mathbf{u}_{g_{3-}}\\ \mathbf{u}_{g_{2+}}|\mathbf{u}_{g_{1-}},\mathbf{u}_{g_{2-}},\mathbf{u}_{g_{3-}}\\ \mathbf{u}_{g_{3+}}|\mathbf{u}_{g_{1-}},\mathbf{u}_{g_{2-}},\mathbf{u}_{g_{3-}}\end{bmatrix}=\begin{bmatrix}-0.13 & -0.07 & -0.05\\ -1.43 & -0.85 & 0.89\\ -0.77 & -0.45 & 0.34\end{bmatrix}\begin{bmatrix}\mathbf{u}_{g_{1-}}\\ \mathbf{u}_{g_{2-}}\\ \mathbf{u}_{g_{3-}}\end{bmatrix}.$$

and calculate the predicted value of $\mathbf{u}_{g_{j+}}|\mathbf{u}_{g_{1-}},\mathbf{u}_{g_{2-}},\mathbf{u}_{g_{3-}}$ for each of the three trials (j =1, 2, and 3 corresponding to 2015, 2016 and 2017), and thus we calculate $\boldsymbol{\eta}_{g_j}$, the deviation from the regression for each geneotype in each trial

$$\boldsymbol{\eta}_{g_j}=\mathbf{u}_{g_{j+}}-\left(\mathbf{u}_{g_{j+}}|\mathbf{u}_{g_{1-}},\mathbf{u}_{g_{2-}},\mathbf{u}_{g_{3-}}\right)$$

which are also referred to as the lack-of-fit effects.

Marginal Distribution

The genetic regression lines unique to each trial obtained using the marginal distribution for that trial. We obtain the marginal distribution over a subset of multivariate normal random variables by dropping the irrelevant variables from the mean vector ($\boldsymbol{\mu}$) and the covariance matrix ($\boldsymbol{\Sigma}$).

Thus, we partition the vector of random genetic effects by `Trial` as follows

$$\begin{bmatrix} \mathbf{u}_{g_1} \\ \hline \mathbf{u}_{g_2} \\ \hline \mathbf{u}_{g_3} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \hline \mathbf{0} \\ \hline \mathbf{0} \end{bmatrix}, \left[\begin{array}{c|c|c} \boldsymbol{\Sigma}_1 & & \\ \hline & \boldsymbol{\Sigma}_2 & \\ \hline & & \boldsymbol{\Sigma}_3 \end{array}\right] \otimes \mathbf{I}_m \right).$$

Marginal Distribution

The general form of the marginal distribution for the jth trial is thus given by

$$\begin{bmatrix} \mathbf{u}_{g_{j+}} \\ \mathbf{u}_{g_{j-}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{j--} & \sigma_{j-+} \\ \sigma_{j+-} & \sigma_{j++} \end{bmatrix} \otimes \mathbf{I}_m \right)$$

where the 2×2 covariance matrix for the jth trial is expanded to

$$\mathbf{\Sigma}_j = \begin{bmatrix} \sigma_{j--} & \sigma_{j-+} \\ \sigma_{j+-} & \sigma_{j++} \end{bmatrix}.$$

And Finally, the Regression Line!

The corresponding conditional distribution is

$$\mathbf{u}_{g_{j+}}|\mathbf{u}_{g_{j-}} \sim \mathcal{N}\left(\sigma_{j+-}\sigma_{j--}^{-1}\,\mathbf{u}_{g_{j-}}, \sigma_{j++} - \sigma_{j+-}\sigma_{j--}^{-1}\sigma_{j-+}\right)$$

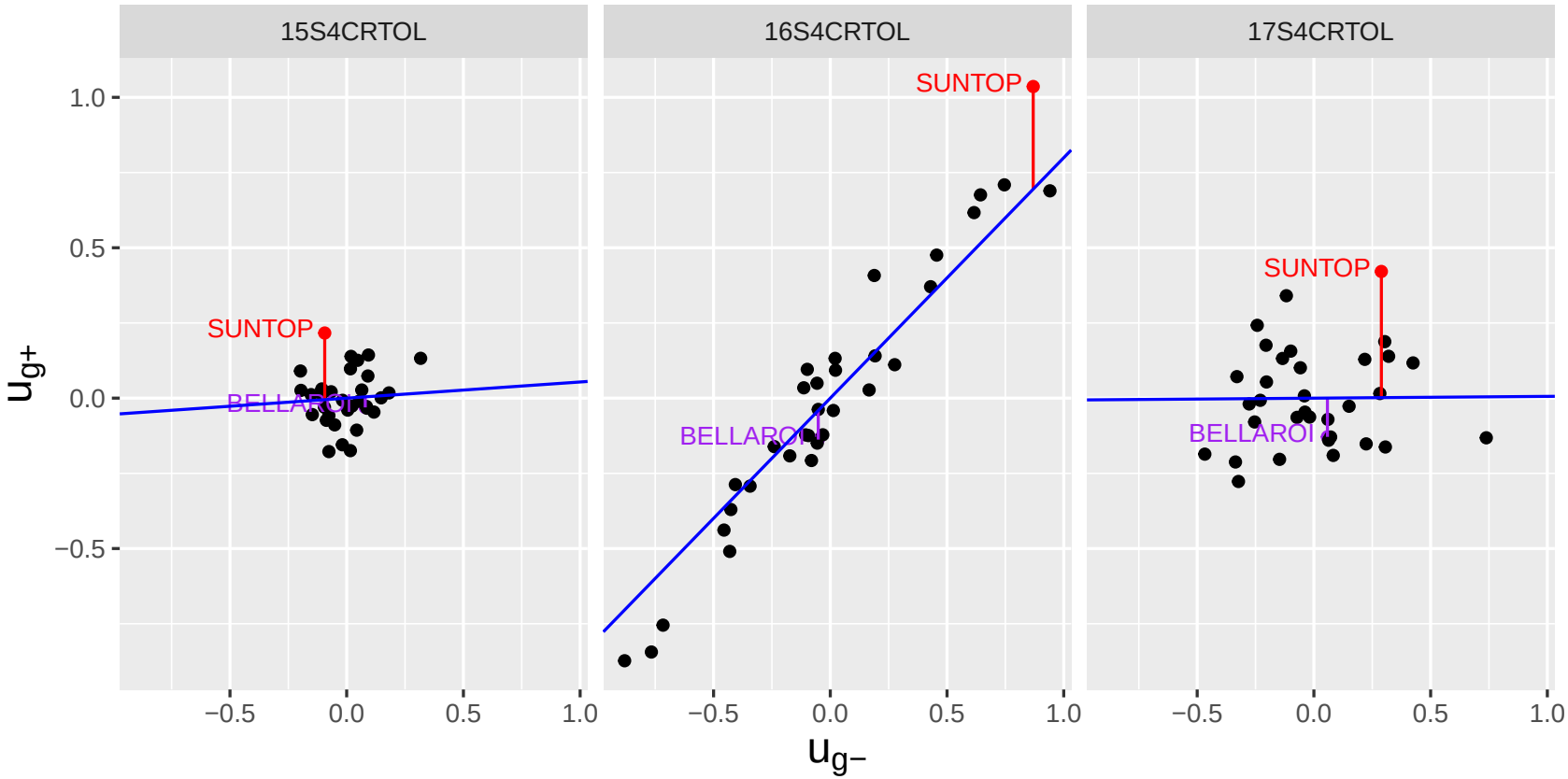
and thus the general form of the genetic regression for the jth trial is given by

$$\mathbf{u}_{g_{j+}}|\mathbf{u}_{g_{j-}} = \sigma_{j+-}\sigma_{j--}^{-1}\,\mathbf{u}_{g_{j-}}$$

where j =1, 2, and 3 corresponding to 2015, 2016 and 2017.

MET '+' vs '−' BLUPs

Trial-specific genetic regression lines $= \sigma_{j+-}\sigma_{j--}^{-1}$.



Acknowledgements

- GRDC for funding the research.

- Professor Brian Cullis for ideas and discussion.

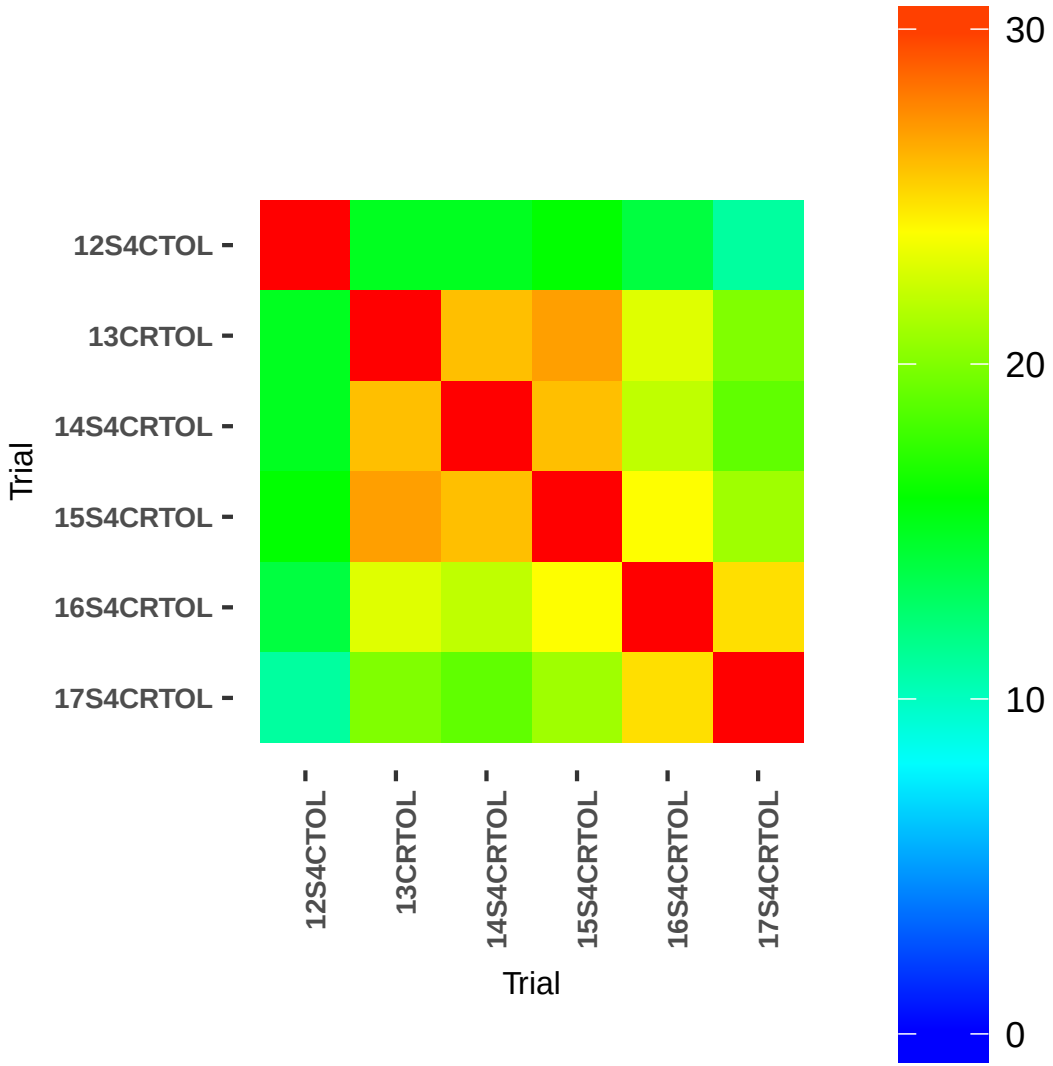
Thank you for listening

Any questions?

References

- Bibby, J. M., Kent, J., Mardia, K. (1979). *Multivariate analysis*.
- Burgess, L. W., Backhouse, D., Summerell, B. A., Swan, L. J. (2001). Crown rot of wheat. In *Fusarium: Paul E. Nelson Memorial Symposium* (pp. 271–294). American Phytopathological Society, St. Paul, MN. field trend. *Biometrics* 57, (pp. 1138–1147).
- Cullis, B. & Smith, A. (2016). The analysis of QTL and QTL $\times$ treatment experiments using spatial models for marker effect. Technical report, Statistics for the Australian Grains Industry.
- Hung, C.-h. Markham, T. (1975). The Moore-Penrose inverse of a partitioned matrix. *Linear Algebra and its Applications* 11, (pp. 73–86).
- Lemerle, D., Smith, A., Verbeek, B., Koetz, E., Lockley, P., Martin, P. (2006). Incremental crop tolerance to weeds: A measure for selecting competitive ability in Australian wheats. *Euphytica*, 149, (pp. 85–95).
- Smith, Cullis, Thompson (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial
- Smith, A. & Cullis, B. (2018). Design Tableau : An aid to specifying the linear mixed model for a comparative experiment. NIASRA working paper.
- Thompson, Cullis, Smith and Gilmour (2003). A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Australian New Zealand Journal of Statistics*, 45(4), (pp. 445–459).
- Wilkinson, G. N., Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, (pp. 392–399).

Genotype Connectivity



SS BLUPs

From the the randomisation-based model the interaction between `CR` treatment regime and `Genotype` are shown to be statistically non-significant based on the Wald test for fixed effects for the 2012, 2013 and 2014 trials.

Trial	p-value		
	CR	Genotype	CR:Genotype
12S4CTOL	0.000	0.528	0.579
13CRTOL	0.000	0.000	0.442
14S4CRTOL	0.000	0.331	0.725
15S4CRTOL	0.000	0.000	0.000
16S4CRTOL	0.000	0.000	0.006
17S4CRTOL	0.000	0.023	0.000

SS δ_g Variance Components

- 2012 and 2013 have very low REML estimate
- 2014 is bound
- no difference between the CR treatment regimes for these trials
- proceed to a MET with 2015–2017

Trial	Variance component for δ (at(CR, +):Genotype)	bound
12S4CTOL	0.00095	P
13CRTOL	0.00024	P
14S4CRTOL	1.41e-08	B
15S4CRTOL	0.04635	P
16S4CRTOL	0.04980	P
17S4CRTOL	0.03284	P