

A whole genome approach for QTL detection using a linear mixed model with correlated marker effects

Alison Smith

`alismith@uow.edu.au`

Centre for Statistical and Survey Methodology
University of Wollongong

QTL detection with correlated marker effects

Collaborations and Acknowledgements

- This presentation is joint work with Brian Cullis (UOW)
- Thanks to Dave Butler (DEEDI) for ASReml-R help
- Thanks to Matthew Nelson (UWA) for use of data
- Thanks to Grains Research and Development Corporation for financial support



Motivating example

- Glasshouse experiment to investigate quantitative trait loci (QTL) controlling photoperiod sensitivity in a doubled haploid (DH) canola population
- 142 DH lines from M x L cross
- DH lines plus 9 other varieties (total of 151) grown in pots in glasshouse
- 2 treatments: long day-length (enabled with lamps) and short day-length

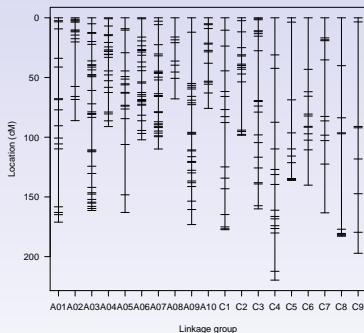
Motivating example



- Treatments randomised to benches (2 benches per treatment)
- Lines randomised to pots within benches using partially replicated design (49 lines with 2 pots each + 102 with single pot = 200 pots per bench)
- Trait of interest: days to first flowering
- For simplicity we will only analyse short day-length treatment

Motivating example

- Genotypic data for 126 DH lines
- 327 DArT markers classified into 19 linkage groups



- Marker covariate information for each DH line: +1 (L allele) or -1 (M allele)

- Mixed models provide flexible framework for QTL detection
- Verbyla, Cullis and Thompson (2007):
 - Whole genome approach (all marker covariates included simultaneously as random effects)
 - Alternative Outlier Model to select “large” random effects
 - Stress importance of accommodating non-genetic variation
- Malosetti, Ribaut, Vargas, Crossa and van Eeuwijk (2008)
 - Marker covariates fitted one at a time as fixed effects: Wald tests (Bonferroni) to select important covariates
 - Important covariates fitted simultaneously then backward elimination

QTL detection

- Most QTL approaches framed as variable selection problems
- Marker information included in analysis as a set of covariates: aim to select those with maximum influence on trait of interest
- “Model search is not a simple task and there is often no unique solution to this problem” (Malosetti et al, 2008)

QTL detection

- Much research on variable selection within framework of random effects (ridge regression, LASSO, Bayesian)
- Often applied for QTL detection but major difference: the covariates have a natural ordering on well-defined metric (genetic distance)
- Why not use tools associated with longitudinal and spatial (geostatistical) analysis?
- Model covariance as a function of distance
- Gianola et al (2003) suggested use of spatial associations between markers for predicting genetic merit

Mixed model with correlated marker effects

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_m(\mathbf{u}_m + \mathbf{u}_n) + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_o\mathbf{u}_o + \mathbf{e}$$

- \mathbf{y} is $n \times 1$ data vector
- \mathbf{u}_m is the $n_m \times 1$ vector of (correlated) marker effects and n_m is number of markers
- \mathbf{u}_n is the $n_m \times 1$ vector of nugget marker effects (ie. additional noise about the correlated effects)
- \mathbf{Z}_m is $n \times n_m$ matrix of marker covariate values
- \mathbf{u}_g is the $n_g \times 1$ vector of *residual* genetic effects (not explained by markers) and n_g is number of genotypes
- $\boldsymbol{\tau}$ is vector of fixed effects (including overall mean)
- \mathbf{u}_o is vector of other (non-genetic) random effects
- \mathbf{e} is $n \times 1$ vector of residuals

Mixed model with correlated marker effects

- Variance structures for marker effects:

$$\text{var}(\mathbf{u}_m) = \sigma_m^2 \boldsymbol{\Sigma}_m$$

$$\text{var}(\mathbf{u}_n) = \sigma_n^2 \mathbf{I}_{n_m}$$

- For correlation model we choose Matern with smoothness parameter ν fixed at 1.5 (ensures differentiability)
- Correlation between two marker effects for markers separated by d cM:

$$\rho = \exp(-d/\phi)(1 + d/\phi)$$

- So elements in $\boldsymbol{\Sigma}_m$ are of this form and we must estimate ϕ , the range parameter

Motivating example: base-line mixed model without marker effects

```
df1.asr <- asreml(df1 ~Gtype,  
random = ~Geno + Bench + Bench:Block,  
rcov = ~idv(Bench):ar1(Range):ar1(Row),data=flgSD.df)
```

Source	REML estimate
Genetic	826.5
Bench	24.0
Bench:Block	0
Residual	181.7
Range autocorrelation	0.13
Row autocorrelation	0.14

Motivating example: mixed model with correlated marker effects

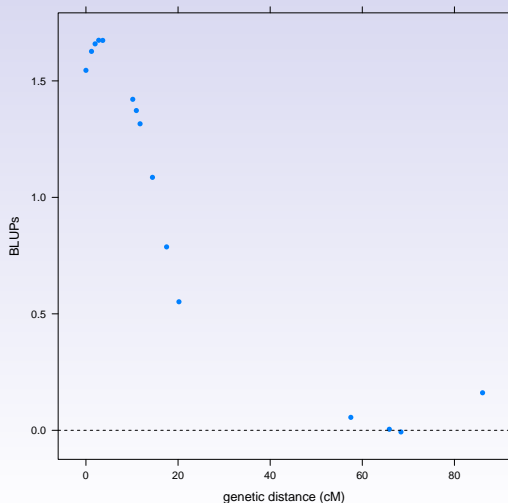
```
dfIQTL.asr <- asreml(dfl ~Gtype,  
random = ~mtrnv(grp('marker'),0,phi=9,nu="1.5F",delta="1F",  
alpha="0F",lambda="2F") +  
grp('resmarker') + Geno + Bench + Bench:Block,  
rcov = ~idv(Bench):ar1(Range):ar1(Row),  
group=list('marker'=1:327,'resmarker'=1:327),data=flgSD.dfm,  
na.method.X='include',pwrpoints=list('marker'=mdist),...)
```

Motivating example: mixed model with correlated marker effects

Source	Base-line model	Marker model
Matern range		5.48
Matern variance		0.46
Nugget		0
Residual genetic	826.5	167.3
Bench	24.0	23.7
Bench:Block	0	0
Residual	181.7	181.9
Range autocorrelation	0.13	0.10
Row autocorrelation	0.14	0.15

Marker BLUPs for Linkage group 2

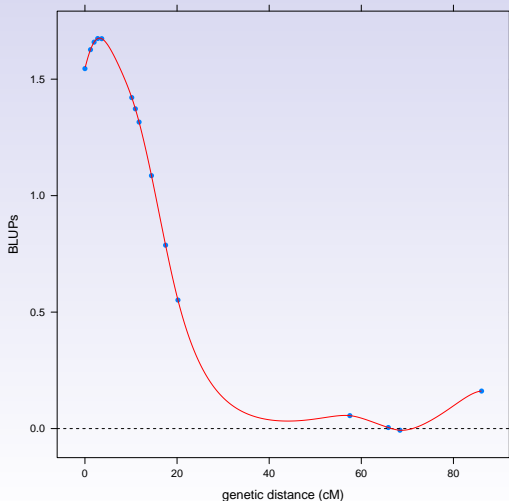
- Best Linear Unbiased Predictions (BLUPs) of marker effects: $\tilde{u}_m = \sigma_m^2 \Sigma_m Z_m' P y$



Predicted marker profile for Linkage group 2

- BLUPs at intermediate distances (“kriging”):

$$\tilde{u}_p = \Sigma_{pm} \Sigma_m^{-1} \tilde{u}_m$$

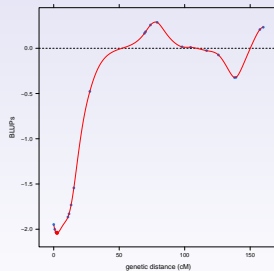
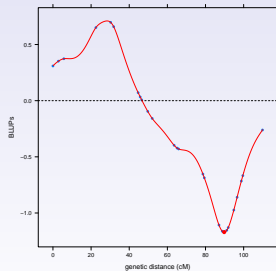
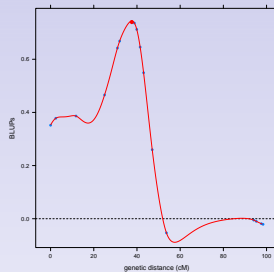
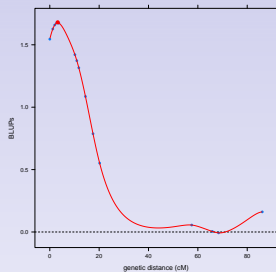


Putative QTL for Linkage group 2

- Compute turning points of profile (numerical procedure)
- If a maximum, compute probability that true effect is greater than zero. Here $p > 0.99$



Putative QTL found on 4 Linkage groups



Ongoing work

- Simulations to assess performance of approach
- Extension for multi-trait (have done bivariate for long/short day-length analysis)
- Extension for multi-environment trials
- Extension for multi-population data
- Application to genomic selection

