2010

Combined analysis of 08/09 and 2010 experiments

Brian R. Cullis
University of Wollongong, bcullis@uow.edu.au

David Butler
Queensland DEEDI

Daryl Mares
University of Adelaide

Kolumbia Mrva
University of Adelaide

Hai Yunn Law
University of Adelaide

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

15-10

Combined analysis of 08/09 and 2010 experiments

Brian Cullis
Professor of Biometry
University of Wollongong

David Butler
Queensland DEEDI

Daryl Mares, Kolumbina Mrva and Hai Yunn Law
University of Adelaide

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

University of Wollongong

Combined analysis of 08/09 and 2010 experiments

Brian Cullis
Professor of Biometry
School of Mathematics and Applied Statistics
Faculty of Informatics
University of Wollongong

Mailing address: University of Wollongong
Wollongong, NSW 2522
Ph : +61 2 4221 5641
Mob : +61 407 338 825
Fax : +61 2 4221 4845
email: bcullis@uow.edu.au

David Butler
Queensland DEEDI
PO Box 102
Toowoomba Qld 4350
Australia
email: david.butler@deedi.qld.gov.au

Daryl Mares, Kolumbina Mrva and Hai Yunn Law
University of Adelaide
email: daryl.mares@adelaide.edu.au

October 1, 2010

Contents

1 Introduction

At a meeting held in Canberra in July 2010, it was agreed to undertake a study to investigate whether a combined analysis of the 08/09 series of LMA experiments and the single 2010 experiment was feasible. The current protocol for classification of lines relies on obtaining information from two valid LMA screening experiments. The analysis of the 08/09 series of experiments represented the first attempt to undertake a combined analysis of LMA screening experiments. It is well known (see for example, Mrode [1995]) that the optimum methodology to evaluate genetic material for eventual selection is to use a multivariate or multi-trait analysis. This multi-trait analysis takes account of the relationships which exist between the different traits. In our example, each experiment is considered as a trait and so for these data we have three traits, one for each experiment.

One of the major advantages of the multi-trait analysis is that it increases the accuracy of evaluation (hence the accuracy of selection). The gain in accuracy is a function of several factors. The most important factor which affects the gain in accuracy from conducting a multi-trait analysis, is the absolute difference between the (true) genetic and residual correlations between the traits. The larger the differences, the greater the gain in accuracy [Thompson and Meyer, 1986]. Since these traits are measured on different experimental units then the residual correlation is zero, and hence we would expect a significant increase in accuracy if there is a non-zero (hopefully positive) genetic correlation between traits. Even though there may be a (true) positive genetic correlation between traits, the realised increase in accuracy depends on the ability to reliably estimate this genetic correlation.

The reliability of the estimation of the genetic correlation between traits (experiments) is driven by the genetic connectivity between the traits. We define the genetic connectivity between two traits/experiments as the number of lines in common. Clearly if there are no lines in common then we cannot estimate the genetic correlation. We are unaware of any study which has examined what would be regarded as a "sufficient" degree of genetic connectivity between a series of experiments. Our empirical evidence from the analysis of multi-environment

trials using the approach of Smith et al. [2001] would suggest that between 10-20 would be a reasonable level of connectivity. We stress, however, that this estimate requires validation with a detailed and rigorous study.

In this report, we therefore will consider a multi-trait analysis of the three LMA screening experiments conducted in 2008, 2009 and 2010. The analyses of these experiments has been presented in two previous reports. The first report considered a re-analysis of two experiments (winter 2008 and summer 2008/2009) which attempted to take account of some features of the "experiment design" as well as incorporating other sources of variation which Butler et al. [2009b] identified as significant contributors to the variation in LMA expression. By applying a more comprehensive analysis it was shown that there was good agreement between the two experiments in terms of LMA expression giving more confidence in the information arising from these experiments, which were conducted prior to the implementation of the new protocol. The second report considered the analysis of the LMA screeing experiment which was conducted over the summer of 2009/2010. This experiment was the first successful experiment which used the new protocol based on the work of Butler et al. [2009b] in which appropriate replication and randomisation is used for the key phases.

2 Experimental details & statistical methods

2.1 Phenotypic Data & Experiment design

The details of the phenotypic data and experiment design for these trials has already been presented in previous reports. The first two experiments were not designed using a valid statistical experimental design. The 2009/2010 experiment was designed using the new protocol advocated by Butler et al. [2009b]. These experiments will be referred to by E1, E2 and E3. Table 2.1 presents a summary of genetic connectivity for the three LMA screening experiments. Numbers on the diagonals represent the number of lines tested in each experiment, while numbers on the off-diagonals represent the number of lines on common between pairs of experiments. So for example there were 364 lines in E1, 178 in E2 and 193 in E3. Similarly, there were 176 lines in common between E1 and E2, but only 23 and 19 lines in common between E1 and E2 and E3 respectively. Thus there was excellent genetic connectivity between E1 and E2 but somewhat less between E3 and the other two experiments, as these were part of another LMA screening series.

	E1	E2	E3
E1	364	176	34
E2	176	178	29
E3	34	29	193

Table 2.1: Summary of genetic connectivity for LMA screening experiments: diagonals represent lines tested, off-diagonals represent lines in common

There were 155 lines tested in two experiments and 28 lines tested in all three experiments. The latter included mainly check varieties such as Cranbrook, Hartog, Janz, Kennedy, RAC655 and Seri.

2.2 Statistical Methods

As before, we consider an analysis of the optical density data (`od`) based on a linear mixed model. Terms in the mixed model have been described in previous reports and involve both those to account for the restricted randomisation of `IDs` to pots and seeds within tillers, tillers within plants, plants within pots, pots within `IDs`, wells within a plate, and the natural blocking of the plate phase for E3 and the pseudo experimental design terms and factors created for the analysis of E1 and E2. The additional feature of the multi-experiment (MEX) analysis is the inclusion of the main effect of `Experiment` and the interaction of `ID:Experiment`, the former included as a fixed effect, the latter included as a random effect with a special variance structure. If we regard the `ID:Experiment` effects as a 535×3 matrix, then we assume that the effects in each of the 535 rows of this matrix are realisations of a multivariate normally distributed random variable, with zero mean and variance matrix given by

$$\begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix}$$

where $g_{ij}, i \neq j$ is the genetic covariance between experiments i and j, and g_{ii} is the genetic variance for experiment i. Note that $g_{ij} = g_{ji}$. The genetic correlation between experiments (i, j) is computed as

$$r_{g_{ij}} = \frac{g_{ij}}{\sqrt{g_{ii} g_{jj}}}$$

The analysis was conducted on the conducted using the linear mixed models software ASReml-R [Butler et al., 2009a].

3 Results

3.1 Estimation of genetic correlation between experiments

The original analysis of E1 and E2 was conducted on a log (base e) scale, however for consistency with the analysis of E3 we used the scaled empirical logit transformation defined in the following. For given (a_L, a_R), being determined for each experiment (from the optical density for the negative and positive controls respectively), we let

$$y = \frac{od - a_L}{a_R - a_L}$$

and define the variate for analysis (z) by

$$\begin{aligned} z &= logit(y) \\ &= log(y/(1-y)) \end{aligned}$$

The values of a_L for the 08/09 experiments were 0.046 and 0.040, while the values of a_R were 0.856 and 1.401 respectively.

	E1	E2	E3
E1	1.000		
E2	0.940	1.000	
E3	0.602	0.734	1.000

Table 3.1 REML estimate of the genetic correlation between experiments

Table 3.1 presents the REML estimate of the genetic correlation between experiments. The correlation between E1 and E2 is very high, while the correlation between E1 and E3 and E2 and E3 is still acceptable. The latter two estimates would be subject to much more sampling variation give the lower genetic connectivity for these pairs of experiments.

3.2 Prediction of LMA expression

Empirical Best Linear Unbiased Prediction

For conducting inferences on the true, but unknown, LMA expression of an `ID` for these experiments, we consider the conditional distribution of the true effects given the observed data. From this conditional distribution, which is approximately normal, with mean being the Empirical Best Linear Unbiased Prediction (e-BLUP) and variance given by the "plug-in" prediction error variance, we can obtain probabilities of the true LMA expression being say lower than an expressing control variety and higher than a non-expressing control variety for any given experiment. Since there is a strong positive association between experiments at the genetic level, we would expect the multi-trait analysis to have improved the accuracy of our predictions, compared to separate univariate analyses.

Figure 3.1 presents a pairs plot of the e-BLUPs, expressed as deviations from the experiment means, between experiments, but only using those lines which are in common for each pair. There is excellent agreement, as expected, given the above REML estimate of the genetic correlation matrix. Figure 3.2 presents a similar pairs plot but only for the check varieties. The slightly reduced genetic correlation between E1 and E3 seems most likely to be influenced by the inconsistent result for RAC655 which stands out on both of the pairs plots. The expression of RAC655 for E3 was relatively lower than for the previous two experiments.

Controlling False Discovery Rate

The current LMA classification protocol relies on the assessment of two valid so-called LMA results. These results were based on whether a line attained a certain percentage of "positive" LMA values. This approach has some serious deficencies, none less serious than the inability to control the problem of multiplicity. Benjamini and Hochberg [1995] noted that

> When pursuing multiple inferences, researchers tend to select the (statistically) significant ones for emphasis, discussion and support of conclusions. An unguarded use of single-inference procedures results in a greatly increased false positive (significance) rate.

In our setting we have the problem of testing say m null hypotheses within an experiment and the null hypothesis is

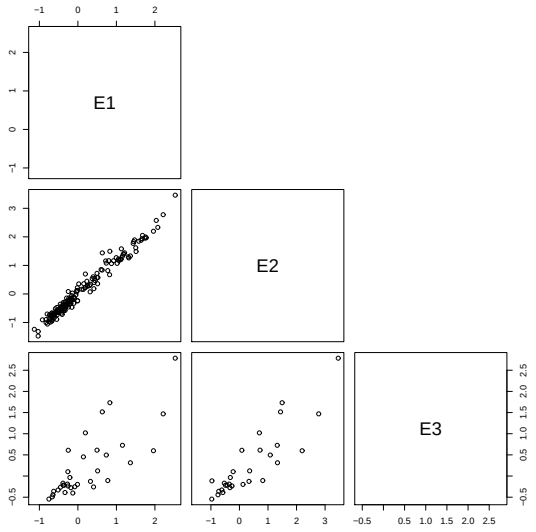

Figure 3.1: Pairs plots of the E-BLUPS between experiments for those lines in common

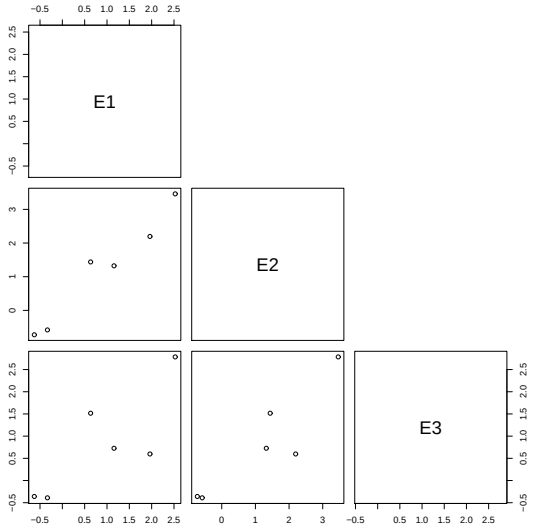

Figure 3.2: Pairs plots of the E-BLUPS between experiments for the check varieties in common

H_0**:** The true LMA expression of line A for experiment K is higher or equal to than that of a nominated set of check varieties for experiment K.

Our aim would therefore be to control (the expected value of) the proportion of the rejected null hypotheses that are erroneously rejected. This statistic is what Benjamini and Hochberg [1995] call the false discovery rate (FDR). Benjamini and Hochberg [1995] develop an approach which controls the FDR at a nominated value (say $q^* = 0.01$) and is simple to implement. Table 3.2 is taken from Benjamini and Hochberg [1995] to illustrate this concept. That is we test m (null hypotheses), of which m_0 are true. R is the number of hypotheses which are rejected and is an unobservable random variable, as are U, V, S and T. The FDR is simply the expected value of the ratio $V/(V + S)$.

	Declared non-significant	Declared significant	Total
True H_0	U	V	m_0
Non-true H_0	T	S	$m - m_0$
	$m - R$	R	m

Table 3.2 Number of errors committed when testing m null hypotheses

The choice of the set of check varieties has been based on previous experience and knowledge of the genetic control of LMA. The varieties Kennedy, RAC655 and Cranbrook are all known to possess at least one of the known QTL which control LMA expression and have been used as so-called "positive" controls for many LMA experiments.

Tables have been prepared for each of the three experiments. The number of null hypotheses is restricted to only consider those lines which were tested in the experiment. These tables are not presented here due to confidentiality reasons, but a sanitized and reduced version of one of the tables is presented in in the following.

Table 3.3 presents a summary of the e-BLUPs, the prediction standard error, expressed as an accuracy, the number of cells tested and the probability value for the test-statistic of H_0 described above. Implementation of the FDR procedure is as follows. Consider testing m null hypotheses for which we have m corresponding p-values, say $P_1, P_2, \ldots, P_m$. Let $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$ be the ordered p-values, and denote by $H_{(i)}$ the null hypothesis corresponding to $P_{(i)}$. These p-values have been computed from a simple one-sided t-test comparing each of m lines with a nominated check variety for each of the three LMA experiments. We consider

the multiplicity problem within experiment and for each check variety. Define the following Bonferroni-type multiple-testing procedure:

Let k be the largest i for which $P_{(i)} \leq p_{ci} = \frac{i}{m}q^*$; then reject all $H_{(i)}, i = 1, 2, \ldots, k$.

The column labelled `Pcrit` is the value $p_{ci} = \frac{i}{m}q^*$ and the final column denotes which null hypotheses are rejected (ie significant).

ID	e-BLUP	acc	cells tested	PKen	Pcrit	significant
181	-1.132	0.851	8	0.00000	0.00003	YES
199	-1.033	0.853	8	0.00000	0.00006	YES
140	-1.030	0.853	8	0.00000	0.00008	YES
180	-0.922	0.851	8	0.00000	0.00011	YES
⋮	⋮	⋮	⋮	⋮	⋮	⋮
66	0.307	0.805	8	0.00446	0.00760	YES
203	0.419	0.853	8	0.00462	0.00763	YES
162	0.415	0.851	8	0.00466	0.00765	YES
279	0.446	0.855	8	0.00564	0.00768	YES
58	0.407	0.823	8	0.00750	0.00771	YES
246	0.371	0.802	8	0.00781	0.00774	NO
36	0.386	0.809	8	0.00784	0.00777	NO
243	0.373	0.802	8	0.00791	0.00779	NO
131	0.433	0.828	8	0.00809	0.00782	NO
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 3.3: Summary of e-BLUPS, the accuracy of the e-BLUPs, the number of wells tested, the probability of comparison with the positive check varieties and FDR statistics as described in the text for a selected number of lines for E1

This procedure provides a clear decision as to whether the null hypothesis is rejected. however, with all statistical tests there still remains uncertainty. There is conjecture regarding the choice of q^* and there will be also some null hypotheses which are "closer" to rejection or acceptance than others. Figures 3.3, 3.4 and 3.5 present plots of the p-values for each of the $m_k, k = 1, 2, 3$ null hypotheses for experiments 1, 2 and 3. We have superimposed the FDR critical value. Null hypotheses which occur before the intersection of the two "lines" are deemed significant. It is clear that some null hypotheses are more strongly rejected than others. To put this into context the p-values for Hartog and Janz for each of the three experiments is less than -12 on the $\log 10$ scale.

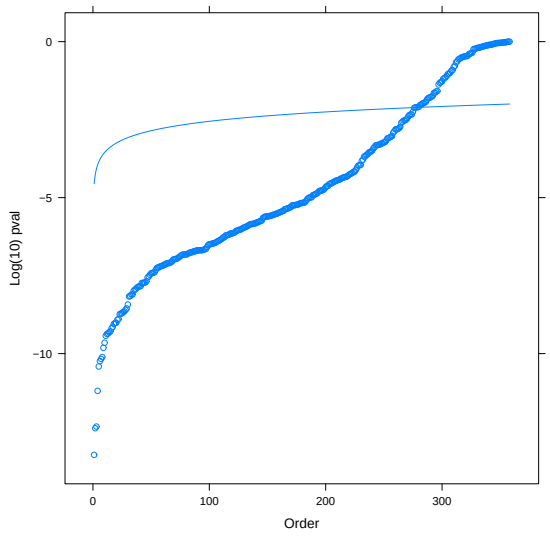

Figure 3.3: p-value plot for experiment 1. The points are the individual p-values for each null hypothesis while the line is the FDR critical value

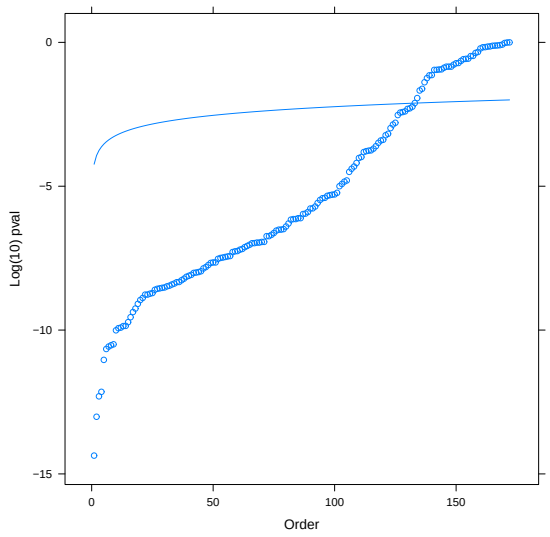

Figure 3.4: p-value plot for experiment 2. The points are the individual p-values for each null hypothesis while the line is the FDR critical value

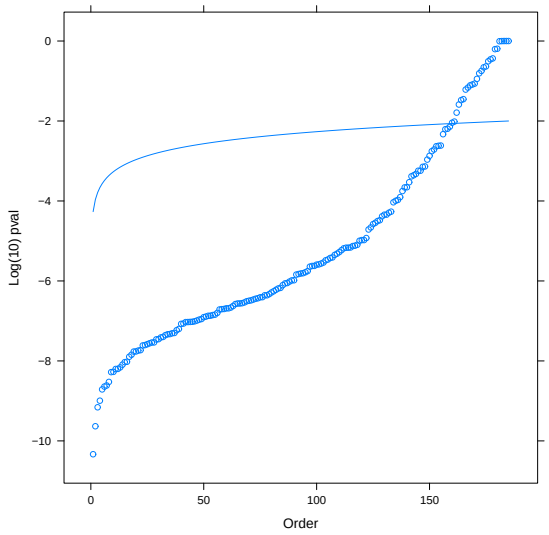

Figure 3.5: p-value plot for experiment 3. The points are the individual p-values for each null hypothesis while the line is the FDR critical value

4 Conclusion

This report illustrates the robustness of the LMA screening protocols. There is strong agreement between experiments in terms of ranking of lines. The analysis of `od` using a mixed models approach followed by a formal assessment of LMA expression on this quantitative trait using an FDR approach offers a valid alternative to the current classification protocol.

Bibliography

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57:289–300, 1995.

D. G. Butler, B. R. Cullis, A. R. Gilmour, and B. J. Gogel. ASReml-R reference manual, release 3. Technical report, Queensland Department of Primary Industries, 2009a.

D. G. Butler, M. K. Tan, and B. R Cullis. Improving the accuracy of selection for late maturity α-amylase in wheat using multi-phase designs. *Crop and Pasture Science*, 60:1202–1208, 2009b.

R.A. Mrode. *Linear models for the prediction of animal breeding values*. CABI Publishing, 1995.

A.B. Smith, B. R. Cullis, and R. Thompson. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, 57:1138–1147, 2001.

R. Thompson and K. Meyer. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science*, 15: 299–313, 1986.