Bioinformatics and Biometrics for the Australian Grains Industry

CAIGE Barley

2015-2020 Multi-year Analysis Report

GRDC Project No: UOS1901-002RTX

Ky Mathews[1], Robin Wilson[2], Mark Dieters[3], Julie M Nicol[4], Richard Trethowan[4]
email:kmathews@uow.edu.au

email:julie.nicol@sydney.edu.au

[1]Centre for Bioinformatics and Biometrics
National Institute for Applied Statistics and Research Australia
School of Mathematics and Applied Statistics
University of Wollongong

[2]Integrated Breeding Platform

[3]School of Agriculture and Food Sciences
University of Queensland

[4]Plant Breeding Institute
Faculty of Science
University of Sydney

August 2, 2021

Contents

1 Background **3**

2 Genetic material **4**
2.1 Entry list 4
2.2 Pedigree information 5

3 Phenotypic data **9**
3.1 Trial information 9
3.2 Agronomic traits 11
3.3 Experimental design 12

4 Statistical Methods **13**

5 Results **14**
5.1 Yield: individual trials 2015-2020 14
5.2 Yield: Multi-environment trial 2015-2020 17
5.3 Agronomic traits: 2020 24

6 Resources **26**

CONTENTS

Confidentiality and IP Statement

Confidentiality and IP Statement *This report was produced as part of the Biometrics and Bioinformatics for the Australian Grain Industry (BBAGI) project funded by the Grains Research Development Corporation. Copyright in this report is owned by UOW and the GRDC. Project stakeholders have permission solely to reproduce this report for the purposes of the BBAGI project. All other uses of the report, including reproducing or adapting all or parts of this report for other contexts or purposes, must first be discussed with the authors and will require the written permission of the copyright owners.*

Acknowledgements: Thanks to Brian Cullis for his guidance in this analysis.

1 Background

The CIMMYT Australian ICARDA Germplasm Evaluation project (CAIGE) is a Grains Research and Development Corporation (GRDC) funded project (UOS1901-002RTX) which commenced in 2005 and is managed by the University of Sydney in collaboration with the University of Queensland and Australian wheat and barley breeders. From 2015 onwards statistical support has been provided by the University of Wollongong (UOW) employed on the GRDC projects: Statistics for the Australian Grains Industry (SAIG2) and Bioinformatics and Biometrics for the Australian Grains Industry (BBAGI) (UW00009). Prior to 2019 the UOW support was for providing experimental designs only. The key objective of CAIGE is to evaluate germplasm developed by CIMMYT (International Maize and Wheat Improvement Centre) and ICARDA (International Centre for Agricultural Research in the Dry Areas) for potential introgression into Australian wheat and barley breeding programs.

Entries (varieties) are evaluated at different sites across the Australian grain belt and selected by breeding companies for introgression into their breeding programs to ultimately release improved varieties to Australian wheat and barley growers. Each year a new cohort of entries is evaluated in a multi-environment trial (MET) series. The relationship between entries, pedigree information, is available in the CAIGE Breeding Management System (BMS) database, (http://www.caigeproject.org.au/breeding-management-system). The inclusion of pedigree information enhances the analysis of the CAIGE dataset in two ways. The CAIGE barley dataset has reasonable connectivity with approximately 50% of top yielding lines retained for testing in a subsequent year, however, until now the CAIGE MET analyses have been within year only. Seasonal variation in Australia is large and selecting entries for use as parents or direct release based on a multi-year analysis is recommended. Pedigree information enables an across year analysis through the ancestral relationships across years as the incoming germplasm are from ICARDA's breeding program which naturally has a high level of co-ancestry from year to year. Further, the use of pedigree information in the analysis allows for the partitioning of the total genetic (variety) effects into additive and non-additive effects. The additive effects are equivalent to estimating the general combining ability of a line and are the most appropriate effects to use when selecting entries to use as parents in a breeding program. This meets a key aim of the CAIGE project - the identification of lines for use as parents for introgression into Australian breeding programs. Furthermore, the total genetic effects can be estimated from this analysis and can be used to identify entries that could potentially be directly released to Australian growers - subject to meeting disease and quality criteria.

A factor analytic linear mixed model (FA-LMM) framework is used for the analysis. This one-stage approach, following Smith et al. (2001); Gogel et al. (2018), allows for the inherent imbalance in the dataset, individual trial spatial modelling and appropriate modelling of the variety by environment interaction (VEI). In addition, the genetic variance-covariance matrix is modelled using the numerator relationship matrix derived

from the pedigree information (Oakey et al., 2006, 2007; Burgueno et al., 2007; Cullis et al., 2010).

The purpose of this report is to describe the 2015-2020 dataset, outline the statistical methods and present the results of the yield dataset and the agronomic traits for 2020. This report is organised as follows: the genetic material is described in Section 2, the phenotypic data and experimental designs are described in Section 3, the statistical methods used are described in Section 4 and the results presented in Section 5.

2 Genetic material

The genetic material consists of ICARDA late stage breeding lines and Australian checks. The non-Australian lines are generally those ready to be deployed in international nurseries or were selected by Australian breeders at ICARDA's breeding locations. The material is imported into Australia under strict quarantine arrangements.

2.1 Entry list

After each new entry cohort is released from quarantine and seed increase it is evaluated in yield trials in Australian environments. There are between 10 and 17 Australian released varieties allocated to usually two plots in each trial each year, 4 of these (COMMANDER, COMPASS, FATHOM and LA TROBE) have been evaluated in all 6 years. There are two ICARDA checks, NAWAIR-1 and VMORALES, which have been evaluated in 5 of the 6 years. NAWAIR-1 was not evaluated in 2015 and VMORALES was not evaluated in 2017. The ICARDA check FURAT-3 was evaluated in 4 years, 2016 and 2018-2020. There were 767 unique entries (excluding fillers) evaluated between 2016 and 2020, their distribution by year and material source is presented in Table 1.

Each entry has passport information which help to uniquely identify it. This passport information includes the Quarantine Code (QCode) for the international nursery and Quarantine Number (QNo), the entry number within a nursery; Genotype Identifier (GID) from either the CAIGE or ICARDA Breeding Management System (BMS); accession identifier (acc_id) from the Australian Grain Genebank; the germplasm source (Source = ICARDA, Australia), the pedigree string and selection history. Occasionally, an entry is imported a second time and is assigned a different QNo:QCode - this can lead to duplicate entries. However, in the barley datasets this has been well managed and the only instances of duplicates are with named checks such as NAWAIR-1 which has the concatenated name NAWAIR-1_22:ZIB15 indicating that it was originally evaluated as 22:ZIB15. Fortunately, the GID for these are the same. GID assures that duplicate entries are identified, it is then a discussion between the biometrician and CAIGE staff to determine that they are true duplicates (see Mathews (2021) for further discussion).

Entry names in CAIGE are usually a concatenation of the QCode and QNo, such that QNo:QCode which is allocated to the EntryCD field in the BMS database. It is important to understand that QNo is nested within QCode, so an entry with QNo=1 with

QCode=ZIB15 is different from an entry with QNo=1 and QCode=ZIC14. If the line is a check line from ICARDA then the QNo:QCode designation is replaced with the check name, e.g. the ICARDA check, VMORALES, was introduced as both 24:ZIC14 and 24:ZIC15. They both had the same GID. Since there were some duplicate entries in the CAIGE barley dataset, GID was used as the unique identifier and the term used in the analysis to link the pedigree and phenotypic data together.

Table 1: Number of entries (excluding fillers) in the CAIGE barley 2015-2020 dataset by year and material source (Australia and ICARDA). The source totals are the unique number of entries across years.

Year	Australian	ICARDA	Total
2015	16	120	136
2016	17	133	150
2017	12	111	123
2018	14	139	153
2019	11	89	100
2020	10	276	286
Total	24	743	767

There are 17 ICARDA entries which were evaluated in only 1 plot across all years due to limited seed availability. They are retained as test lines in the dataset because they are linked to other lines by pedigree. There is 1 filler in the dataset and it was not included in the estimation of genetic variance but is retained in the dataset and modelled following the process described in Appendix C of Tolhurst et al. (2019).

2.2 Pedigree information

Pedigree information describes the relationships between all entries in the dataset. Following Henderson (1976) this information is converted to a numerator relationship matrix, $\boldsymbol{A}$, which can be used in a linear mixed model to partition the total genetic variance into additive and non-additive effects (Oakey et al., 2006, 2007; Burgueno et al., 2007; Cullis et al., 2010). Estimated breeding values for parental selection can then be obtained from this analysis.

The pedigree information was extracted by CAIGE staff from the CAIGE BMS database. Whilst there is some uncertainty about how far back in the ancestral tree it is required for accurate estimation of the coefficient of parentage it is reasonable to assume that the greater the depth the more accurate the estimate. This required a significant curation process of the entry selection histories and pedigree strings, and the parent and grandparent GID and pedigree strings which is documented in Mathews (2021). Drs. Miguel Sanchez Garcia (ICARDA breeder) and Mark Dieters's (UQ) collaborative assistance with this process was invaluable.

The result of this data curation is a data frame in the so-called "Me Mum Dad" format which is described in the `R pedicure` library (Butler, 2019) as a *data frame with (at least) three columns that correspond to the individual, female parent and male parent,*

respectively. The row giving the pedigree of an individual must appear before any row where that individual appears as a parent. Founders use 0 (zero) or NA in the parental columns. This information is provided to demonstrate that the pedigree file requires more than just the parents of test lines. For example, the"Me Mum Dad" data frame used in the generation of the numerator relationship matrix for this dataset had 2264 records. An example "Me Mum Dad" file is presented in Table 2. The individual in the last row, 600103992 which has QNo:QCode 87:ZBT19, has a 'Mum' (female parent) with pedigree 4-1MBN11 and a 'Dad' (male parent) of CANELA//E.ACACIA/DEFRA. These parents are in rows 5 and 6, respectively, in this table, *above* the individual that will be tested in the field.

Table 2: Example of a *Me Mum Dad* file for the CAIGE barley dataset. Me is the individual being tested or an ancestor of an entry being tested, Mum and Dad contain the name of the parents. A value of zero for Mum or Dad indicates unknown pedigree or base individuals. Fn is the filial generation number and is a single number for single plant selections (SPP) or of the form $n_b : n_f$ indicating the level of SPP and number of bulked generations. $\mathcal{F}$ is the inbreeding coefficient.

Me	Mum	Dad	Fn	$\mathcal{F}$
CANELA	MARIS CANON/LAUREL	ALELI	7	0.98438
E.ACACIA	0	0	6	0.96875
DEFRA	0	0	7	0.98438
E.ACACIA/DEFRA	E.ACACIA	DEFRA	7	0.98438
4-1MBN11	0	0	7	0.98438
CANELA//E.ACACIA/DEFRA	CANELA	E.ACACIA/DEFRA	7	0.98438
600103992	4-1MBN11	CANELA//E.ACACIA/DEFRA	5:7	0.99219

Functions `chkPed` and `trimPed` from the `pedicure` (Butler, 2019) library on the `R` statistical computing platform (R Core Team, 2020) are used to curate the pedigree (i.e. "Me Mum Dad") file. `Pedicure` is freely available from www.mmade.org/pedicure. A function specifically written for this project, `pedigree.cut` was used to cut the pedigree strings back to founders and include them in the "Me Mum Dad" file. The pedigree data frame for the 2015-2020 dataset contained 2264 records; the 767 entries evaluated in 2015-2020 and 1497 ancestors.

The fourth column, Fn, in the pedigree file (Table 2) is the filial generation number of the line. This can take the form of an integer (1 = simple cross, 2 indicates the progeny of this cross selfed once, 3 is selfed twice, and so on). If the breeding program uses bulk selections before a single plant selection then Fn takes the form $n_b : n_f$, where n_b is the Fn of the single plant selection and n_f indicates the final generation of bulking. The difference between n_f and n_b is the number of generations of bulking. For example, an $F_{2:5}$ line, F_2 derived F_5 line, would have an Fn of 2:5, indicating a single plant selection at F2 followed by 3 generations of bulking. The fifth column self represents the level of selfing where

$$self = \begin{cases} Fn - 1, & \text{where } Fn \text{ is an integer} \\ n_f - 1, & \text{otherwise} \end{cases} \tag{1}$$

Jordan & Cullis (2020) have developed an improved algorithm for calculating the inbreeding coefficient of bulk selected lines. However, this was not successfully implemented for this dataset. The ICARDA breeders believe the inbreeding values to be higher than the values obtained from this algorithm and hence an approximation of $n_f - 1$ was used for these lines.

The selection histories of the CAIGE test entries were used to define the Fn values for these lines. This is described in Mathews (2021). Australian checks are considered fixed and assigned a Fn=8, whilst all ancestors are assigned a Fn=6.

The inbreeding coefficient $\mathcal{F}$ (Henderson, 1976) for each record in the pedigree file is then calculated as

$$\mathcal{F} = 1 - \frac{1}{2}^{self}, \text{ where } self \text{ is defined in equation (1).} \quad (2)$$

The diagonals of the numerator relationship matrix ($\boldsymbol{A}$) are equal to one plus the inbreeding coefficient for an individual (i.e $1 + \mathcal{F}$), and the off-diagonals are equal twice the coefficient of parentage (ancestry) between any two individuals (Wright, 1922). In practice, the inverse of $\boldsymbol{A}$ is used in the design and analysis and this is calculated using the `ainverse` function in the linear mixed model software `asreml` (Butler et al., in prep.) following the process outlined in Gilmour et al. (2004) and Meuwissen & Luo (1992). `asreml` is licensed software available from www.vsni.co.uk.

Table 3: Numerator relationship (**A**) matrix for 600103992 (87:ZBT19) and ancestors in the CAIGE barley pedigree file.

GID	CANELA	E.ACACIA	DEFRA	E.ACACIA/DEFRA	4-1MBN11	CANELA/ E.ACACIA/DEFRA	600103992
CANELA	1.984	0.000	0.000	0.000	0.000	0.992	0.496
E.ACACIA	0.000	1.969	0.000	0.984	0.000	0.492	0.246
DEFRA	0.000	0.000	1.984	0.992	0.000	0.496	0.248
E.ACACIA/DEFRA	0.000	0.984	0.992	1.984	0.000	0.992	0.496
4-1MBN11	0.000	0.000	0.000	0.000	1.984	0.000	0.992
CANELA/ E.ACACIA/DEFRA	0.992	0.492	0.496	0.992	0.000	1.984	0.992
600103992	0.496	0.246	0.248	0.496	0.992	0.992	1.984

A summary of the resulting inbreeding coefficients by genetic source (Table 4) shows that the ICARDA breeding program has a range of heterozygous and homozygous lines. Many lines are bulked from F_3 onwards with single plant selections occurring between 1 and 8 bulking generations later.

The variety connectivity matrix (Figure 1) illustrates a number of interesting points about the CAIGE barley dataset. Firstly, the practice of retaining approximately 50% of the highest yield lines in subsequent years was upheld in 2015-2016, less so in 2016-2017,

2 Genetic material

Table 4: Summary of the inbreeding coefficients for the 767 entries by genetic source (Australian and ICARDA) in the CAIGE barley 2015-20 dataset.

Source	Min	Mean	Max
Australian	0.992	0.992	0.992
ICARDA	0.750	0.937	0.996

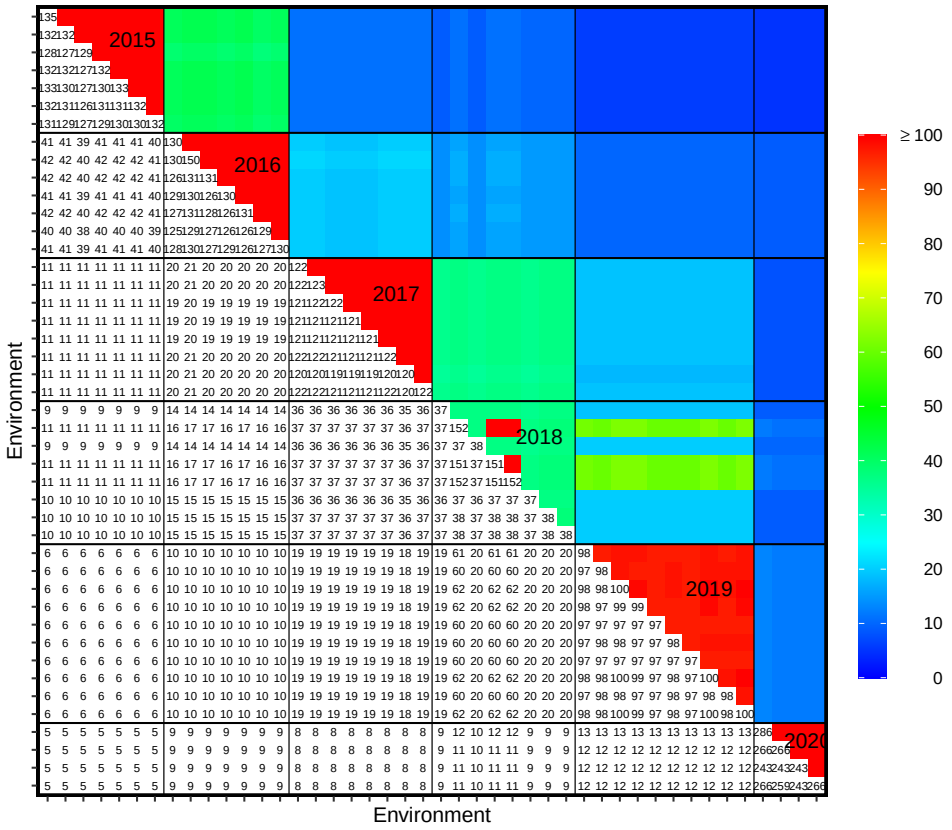

Figure 1: Heatmap of the variety connectivity matrix between all environments in the CAIGE barley 2015-2020 dataset

2017-2018 and 2018-2019 (minimum 6, maximum 62). The low number of entries in 2018 is a result of significant lodging in trials in Morocco which reduced the number of entries imported. There is a clear lack of connectivity between 2020 entries with previous years with only the checks in common with 2019 entries. This is a direct result of a change in direction in the current phase of the CAIGE project commencing in 2019. It was decided to focus on a different set of barley pedigrees which have a different agronomic type from previous years. Further, it was also decided not to carry through any entries from the previous phase of the project. The parental connectivity matrix (Figure 2) indicates that there is reasonable connectivity at the parent level across years, including 2020, with a minimum number of parents in common between environments of 26.

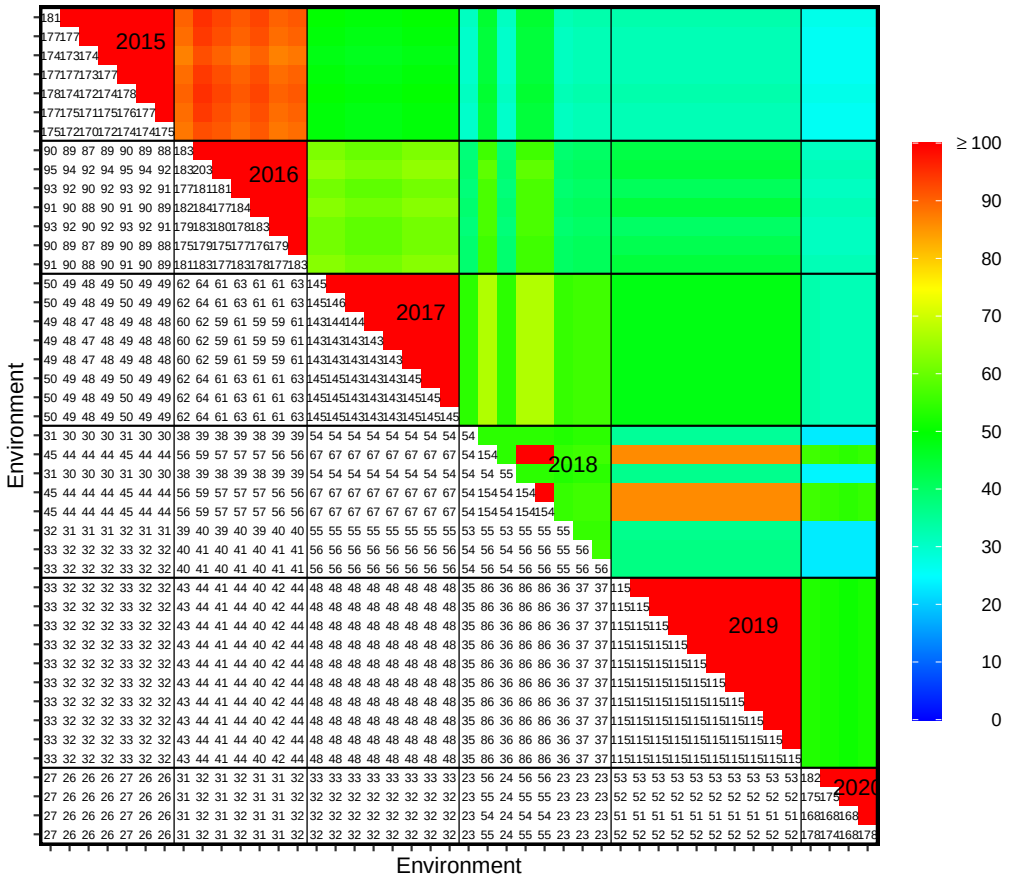

Figure 2: Heatmap of the parental connectivity matrix between all environments in the CAIGE barley 2015-2020 dataset

3 Phenotypic data

The key trait of interest for this analysis is yield (t/ha). Other traits such as flowering and maturity dates, plant height, have also been collated but only sparsely across the dataset. These secondary traits are analysed for 2020 only.

3.1 Trial information

There are 44 trials in the full 6 year dataset evaluated at 19 locations. Table 5 provides a summary of the CAIGE yield trials each year including the number of entries, number of trials and range of percentage replication, environment mean yield (t/ha) and individual trial reliabilities. Each trial is evaluated in a unique environment which is the year-location combination. The terms *trial* and *environment* are synonymous.

3 Phenotypic data

The specific details for all 44 trials is provided in Table 6. For each trial the experimental design plot factors, the number of entries (varieties) and their level of replication and the environment mean yield (t/ha) and reliability estimated from single environment analyses.

The number of entries in each trial ranges from 119 in Kalkee 2014 to 312 in Narrabri in 2019. Gatton is considered the home site and is irrigated. However, there was no trial at Gatton in 2015 or 2016. Trials have been conducted at (Kalkee (Victoria) and Toodyay (Western Australia)) in 5 of the 6 years. There is no location where a trial was grown each year. The environment mean yield ranges from 0.62-6.94t/ha which is representative of Australian barley production values. The reliability, equivalent to the line mean broad sense heritability, ranges from 0.476 to 0.925.

At the advent of the 3-year project (US00073) commencing in 2019, GRDC advised the CAIGE team that there was now a requirement to track and manage publicly and privately funded trials in the project. Ky Mathews, from BBAGI, made it very clear to GRDC and the CAIGE team that separate analyses for publicly and privately funded trials would not be entertained as this is a commercial, not scientific, decision. However, the tracking of public and private trials is required to ensure results from privately funded trials are not released publicly and is an extra burden on the statistical team as the CAIGE raw plot data is not stored in BMS and hence the management of this is not a simple filter as one would anticipate. There are 5 privately funded trials in the dataset, all evaluated in 2019, as indicated in italics in Table 6.

Table 5: Summary of the CAIGE barley 2015-20 dataset by year: information includes the number of locations, range of entries, percentage replication, environment mean yield (t/ha) and environment reliabilities in each year.

Year	Number of Locations	Number of Entries	Percentage Replication (%)	Environment Mean Yield (t/ha)	Environment Reliability
2015	7	129-135	73-82	0.99-3.52	0.65-0.87
2016	7	129-150	72-79	2.41-4.81	0.25-0.92
2017	8	120-123	93-98	0.74-3.51	0.52-0.89
2018	8	37-152	24-59	1.17-5.27	0.21-0.83
2019	10	97-100	68-73	0.37-4.34	0.45-0.91
2020	4	243-286	38-53	2.5-4.59	0.61-0.89

3 Phenotypic data

Table 6: Experimental design information for each trial in the 2015-2020 CAIGE barley dataset. The number of blocks, ranges, rows, plots and entries, environment mean yield (EMY, t/ha) and reliability (r^2), total genetic variance and additive variance expressed as a percentage of the total genetic variance (%Additive Variance) from single environment analyses are provided. The number of entries allocated to one, two and more that two plots is indicated and the resulting percentage partial replication (%-p-rep). Filler plots are not included in the entry number figures. *Italicised* environment names are privately funded trials.

Env	Block	Range	Row	Plot	Entry	Entries replicated			%p-rep	EMY	r^2	Genetic	%Additive
						once	twice	> 2		(t/ha)		Variance	Variance
B15BROO4	2	12	20	240	135	30	105	0	77.8	3.52	0.866	0.195	20
B15GERA5	2	12	20	240	132	24	108	0	81.8	0.99	0.757	0.077	46
B15KALK3	2	12	19	228	129	30	99	0	76.7	2.04	0.838	0.285	46
B15PINE5	2	12	20	240	132	24	108	0	81.8	3.46	0.803	0.172	23
B15TOOD6	2	12	20	240	133	26	107	0	80.5	2.91	0.870	0.306	25
B15TUCK5	2	12	20	240	132	25	106	1	81.1	2.88	0.647	0.268	51
B15WONG6	2	12	19	228	132	36	96	0	72.7	3.36	0.737	0.261	51
B16ANGV5	2	24	10	240	130	27	98	5	79.2	3.60	0.887	0.508	46
B16BROO4	2	12	22	264	150	40	108	2	73.3	4.30	0.857	0.478	48
B16CHAR5	2	12	20	240	131	28	99	4	78.6	2.86	0.798	0.385	49
B16KALK3	2	12	19	228	130	36	92	2	72.3	2.41	0.916	1.196	47
B16ROSE5	2	12	20	240	131	28	99	4	78.6	3.23	0.916	0.782	50
B16TOOD6	2	12	20	240	129	27	96	6	79.1	4.19	0.245	0.271	69
B16WONG6	2	12	19	228	130	36	92	2	72.3	4.81	0.872	1.036	33
B17DULA4	2	12	20	240	122	7	113	2	94.3	3.16	0.885	0.161	43
B17GATT4	2	20	12	240	123	9	112	2	92.7	3.11	0.771	0.348	53
B17MERR6	2	12	20	240	122	4	118	0	96.7	0.74	0.517	0.022	50
B17NARR2	2	20	12	240	121	2	119	0	98.3	3.47	0.866	0.514	43
B17PINN5	2	24	10	240	121	2	119	0	98.3	1.98	0.688	0.271	48
B17ROSE5	2	24	10	240	122	4	118	0	96.7	3.39	0.820	0.554	48
B17TOOD6	2	12	20	240	120	2	116	2	98.3	3.51	0.796	0.757	49
B17WONG6	2	12	20	240	122	4	118	0	96.7	2.11	0.772	0.118	45
B18CORR2	2	4	12	48	37	28	9	0	24.3	3.67	0.623	0.139	56
B18GATT4	2	10	24	240	152	64	88	0	57.9	4.12	0.656	0.479	52
B18KALK3	2	4	12	48	38	29	9	0	23.7	3.58	0.455	0.080	62
B18MERR6	2	24	10	240	151	62	89	0	58.9	1.50	0.684	0.209	51
B18NARR2	2	10	24	240	152	64	88	0	57.9	2.61	0.826	0.221	41
B18PINN5	2	2	24	48	37	26	11	0	29.7	1.17	0.613	0.058	50
B18ROSE5	2	2	24	48	38	28	10	0	26.3	3.40	0.210	0.040	92
B18TOOD6	2	4	12	48	38	28	10	0	26.3	5.27	0.764	0.658	45
B19BREE2	2	12	14	168	98	28	70	0	71.4	0.37	0.464	0.016	48
B19CORR2	2	12	14	168	98	28	70	0	71.4	1.68	0.562	0.078	41
B19GATT4	2	24	7	168	100	32	68	0	68.0	4.34	0.448	0.215	56
B19KALK3	2	12	14	168	99	30	69	0	69.7	2.49	0.901	0.678	42
B19MERR6	2	12	14	168	97	26	71	0	73.2	1.75	0.891	0.195	40
B19NARR2	2	10	17	170	98	28	69	1	71.4	2.92	0.797	0.317	53
B19PINN5	2	24	7	168	97	26	71	0	73.2	2.04	0.792	0.198	53
B19ROSE5	2	24	7	168	100	32	68	0	68.0	2.38	0.710	0.249	52
B19SWAN3	2	12	14	168	98	28	70	0	71.4	1.44	0.910	0.318	54
B19TOOD6	2	12	14	168	100	32	68	0	68.0	2.57	0.677	0.280	52
B20GATT4	2	12	33	396	286	176	110	0	38.5	2.50	0.613	0.332	39
B20KALK3	2	12	32	384	266	149	116	1	44.0	4.20	0.891	1.720	46
B20NARR2	2	31	12	372	243	114	129	0	53.1	4.59	0.853	0.524	49
B20YORK6	2	12	32	384	266	148	118	0	44.4	2.67	0.849	0.126	46

3.2 Agronomic traits

Phenology and height data were measured on the four 2020 barley trials. Not all traits were measured in all environments and phenology was measured as either days to heading, days to maturity, Zadoks or a relative maturity score (scale of 1-5), depending on the

location. The distribution of traits across environments is presented in Table 7.

Table 7: Mean (standard deviation) for days to heading and maturity, maturity zadoks, relative maturity (1-5) and height (cm) data for CAIGE barley trials evaluated in 2020.

Env	Days to Heading	Days to Maturity	Maturity Zadoks	Relative Maturity	Height (cm)
B20GATT4				2.8 (1.08)	65.4 (10.64)
B20KALK3			52.1 (5.64)		
B20NARR2	99.8 (5.99)	144.1 (3.08)			
B20YORK6	99.1 (5.03)				78.2 (10.01)

3.3 Experimental design

The experimental designs for the CAIGE barley dataset within and across years are implicitly unbalanced. Firstly, the number of entries varied between years from 37 to 286 with 4 checks evaluated in all years (Table 1) and the number of Australian checks in a year ranged from 10 to 17. Secondly, the amount of seed available per entry post-quarantine and multiplication varied such that some entries had enough seed for allocation to 2 plots per location and others did not. Further, in order to accommodate the known large variety by environment interaction in Australian environments the entries were evaluated in as many locations as possible within a year. The partially replicated designs proposed by Cullis et al. (2006) are well suited to this scenario. Cullis et al. (2020) extended the partially-replicated model based designs to include pedigree information thereby ensuring appropriate allocation of entries at the ancestral level across a location.

In partially-replicated designs, entries with sufficient seed (including test and check entries) are replicated across the field trial whilst the remaining entries are allocated to one plot only. The allocation of entries across locations within years can ensure that across the MET dataset the entries are equally replicated, depending on seed and plot availability. The extension of the pedigree partially-replicated designs of Cullis et al. (2020) to include allocation of entries across environments is presented in Cullis et al. (2018); Ganesalingam et al. (2019). These so-called incomplete MET designs using pedigree were implemented in the design of the 2021 CAIGE barley trials, Pirathiban et al. (2021) but not implemented in the design of any trials in the 2015-2020 dataset.

Crop variety trials are usually laid out in a rectangular array of rows and column. The experimental designs for trials in 2015 are partially replicated row-column designs generated by Dr. Mark Dieters using CycDesign v1.0. There are two key steps in the experimental design of METs. The first is the allocation of entries to trials and the second is the allocation of entries to plots within trials. From 2016 onwards, Dr. Mark Dieters (UQ) allocated the entries to trials and the allocation of entries to plots within trials were performed by University of Wollongong staff employed on the GRDC-funded BBAG project. For the allocation of entries to plots a model-based design approach was implemented using the Optimal Design (`od`) library (Butler & Cullis, 2019) on the `R` statistical platform (R Core

Team, 2020). The designs were partially replicated block designs which take into account the row and column (range) arrangement. Replicated entries were allocated across the field layout such that the blocks at each location were near-resolvable. The percentage replication ranged from 24% for B18KALK3 to 98.3% for B17NARR2, B17PINN5, B17TOOD6, with a median of 72.5%; 8 of the 44 trials had a %p-rep less than 50%.

4 Statistical Methods

In this section the statistical methods applied to the 2015-2020 yield dataset and the 2020 phenology and height data are described. For the purposes of the analysis we defined the term Variety to describe entries and Environment defined a year by location combination.

A one-stage analysis of trials from a MET dataset combines the individual plot data across trials. This allows appropriate modelling of the individual trial variation and is achieved in the linear mixed model framework (Smith et al., 2001; Gogel et al., 2018).

The base model for each environment was a random block effect (where available, Table 8) and the residual model was a separable auto-regressive correlation model fitted in both the column and row directions. Spatial models were then determined following Gilmour et al. (1997). Table 8 provides a summary of the models fitted by year.

Table 8: Model summary by year indicating the number of environments where a term was fitted for the 2015-2020 yield CAIGE barley analysis. Vertical lines delineate between plot structure and residual terms. ar1 refers to the autoregressive correlation model of order 1 fitted at the residual level; an identity model is fitted if an ar1 is not.

Year	Environment	Block	Column	Row	ar1Column	ar1Row
2015	7	7	3	2	4	6
2016	7	7	3	3	3	4
2017	8	8	4	1	6	6
2018	8	8	1	3	3	3
2019	10	10	1	0	10	9
2020	4	4	2	2	3	4

The first step in the analysis of a MET dataset is to analyse each environment individually, this can be performed in one step where the genetic variance-covariance matrix is modelled with a diagonal structure, that is, independence between environments and different genetic variances for each environment. This model was fitted with and without the numerator relationship matrix, $\boldsymbol{A}$, to demonstrate to the CAIGE collaborators that the ranks of the total variety effects within each environment with and without pedigree are comparable. The agronomic trait data were analysed as individual trials as there were no more than 2 environments measured for any one trait (Table 7). They were analysed with pedigree information in order to estimate both the additive and total variety effects.

A key feature of MET datasets is the re-ranking of variety performance in different environments, known as variety by environment interaction (VEI). The factor analytic lin-

ear mixed model (FA-LMM) parsimoniously models the between environment variance-covariance matrix and usually results in a good fit to the data (Smith et al., 2001; Gogel et al., 2018). FA-LMMs account for the covariances of the VE effects between environments using a small number, k, of (unknown) common factors, which are estimated from the data. An FAk model has order k.

Further, the genetic variance-covariance matrix can also be incorporated into a FA-LMM using the numerator relationship matrix derived from pedigree information (Oakey et al., 2006, 2007). This enables the total VE effects to be partitioned into additive and non-additive VE effects, and their respective between environment genetic variances matrices can be modelled with separate FA models (Oakey et al., 2007; Smith & Cullis, 2018). The order, k_a or k_e, of the respective additive and non-additive FA models are increased independently until a good fit to the data is found. Logliklihood ratio tests for nested models, Akaike Information Criterion (AIC) for non-nested models and descriptive statistics, such as the percentage genetic variance accounted for (%vaf) (Smith et al., 2015), are utilised in this decision making process.

Recently, factor analytic selection tools (FAST) were derived to summarise the large amount of information generated from FA-LMM (Smith & Cullis, 2018). They consist of measures of overall performance (OP), stability (root mean square deviation, RMSD) and responsiveness and summarise variety performance across the dataset. OP and RMSD are based on the first factor, which by design, accounts for the most genetic variance and responsiveness is based on the remaining k factors. Importantly, all measures are based on the same scale as the data, i.e. t/ha for yield in this dataset.

In practice, the agronomic traits can be analysed before the yield data as they can be used as a data curation step to ensure that the process of naming entries is appropriate, see Mathews et al. (2021) for details.

5 Results

In this section, we present a summary of the models fitted for each trait: yield, days to heading, maturity zadoks and height. For the key trait, yield, the results from an individual analysis (i.e. no between environment covariances considered) with and without pedigree information is presented to demonstrate the similarity in total variety effects predicted from these two models (Section 5.1). Next, the results from the MET analysis, where the between environment variance-covariance matrix is modelled using a factor analytic linear mixed model (Section 5.2). The results for the agronomic traits are presented in Section 5.3.

5.1 Yield: individual trials 2015-2020

An independent genetic variance model, where no relationship between environments (i.e. a diagonal variance structure) was fitted, without and with considering the relationship between varieties, to the 2015-2020 dataset. The pedigree (DIAG-DIAG) model had

a significantly larger REML log-likelihood (522) than the non-pedigree (DIAG) model (133.6), Table 9. The percentage of additive and non-additive to total genetic variances were 92% and 8%, respectively. This ratio is typically higher than observed for yield in other crops.

Variety by environment empirical Best Linear Unbiased Predictions (VE-eBLUPs) were formed from both models. We compare the VE-BLUPS from each model to demonstrate that the more complex pedigree model provides predictions that are highly correlated with those from the simplest model where between environment correlations and between entry relationships are not considered. From the pedigree model both the additive and non-additive predictions were formed and these were summed to produce the total VE-eBLUPs. The correlations between the total VE-eBLUPS from both models and the additive VE-eBLUPS from the pedigree model were calculated overall and at the environment level. The correlation between the total VE-eBLUPs from both models was 0.973, the correlation between the total VE-eBLUPs from the identity model and the additive VE-eBLUPs from the pedigree model is 0.956 (Table 11). The high correlation of the additive VE-eBLUPS is consistent with the additive genetic variance explaining 92% of the total genetic variance. The correlations between the non-pedigree and pedigree VE-eBLUPs for individual environments ranged from 0.973 to 1 across the full dataset. The correlations between the non-pedigree and pedigree VE for the 2020 environments is presented graphically in Figure 3. B20GATT4 has the lowest % additive genetic variance, 30% (Table 6) which explains the greater variation observed in that plot compared to B20KALK3 (46%), B20NARR2 (49%) and B20YORK6 (46%). Figure 3 demonstrates that the predictions from the simple model are highly correlated with those from the more complex, informative model.

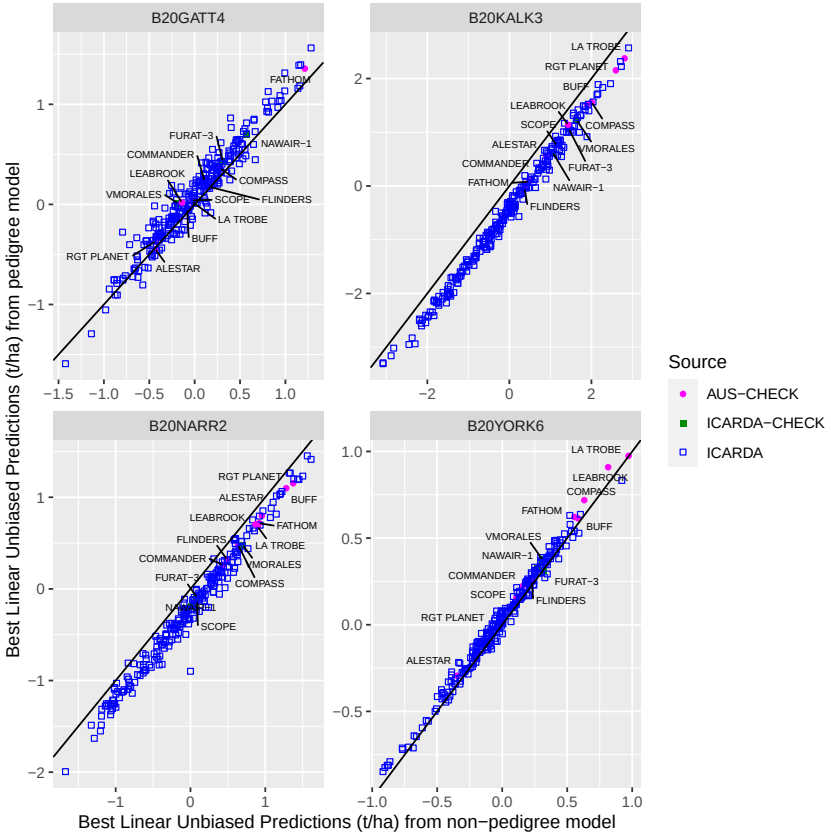

Figure 3: Total Variety by Environment Best Linear Unbiased Predictions for the pedigree versus the non-pedigree model for trials conducted in 2020 where each environment is considered independently. The legend indicates the germplasm source. Australian and ICARDA checks are labelled.

5.2 Yield: Multi-environment trial 2015-2020

The modelling of this dataset commenced with an independent genetic variance model, where no relationship between varieties or environments was considered (DIAG, Table 9). It then progressed incrementally with increasing complexity and resulted in a final model where k_a=3 and k_e = 2. This model had the maximum log-likelihood, the lowest AIC and the largest %VAF. The factor %VAF for the additive effects for the 3 factors are 53%, 21%, and 12%. The FA3-FA1 model would also have been an appropriate choice, especially considering the low percentage of non-additive genetic variance in this dataset.

Table 9: Summary of models fitted to the CAIGE Barley 2015-2020 yield dataset for the additive and non-additive effects; information includes the model, the number of additive and non-additive genetic variance parameters, the log-likelihood (Loglik), log-likelihood ratio test (LRT), Akaike Information Criteria (AIC) and additive, non-additive and total genetic variance accounted for (%). The FA4-FA2 model is the final model.

Model	No. parameters					% Variance Accounted For		
Add-Non	Additive	NonAdditive	Loglik	LRT	AIC	Additive	Non-additive	Total
DIAG			134		169			
DIAG-DIAG	44	44	522	3.33E-16	-520	92	8	
FA1-DIAG	88	44	1128	3.33E-16	-1644	55	6	
FA1-FA1	88	88	1306	3.33E-16	-1912	59	65	60
FA2-FA1	131	88	1410	1.11E-16	-2035	71	71	71
FA3-FA1	173	88	1473	3.66E-15	-2075	87	54	81
FA3-FA2	173	131	1516	7.06E-09	-2076	88	69	83

Selection decisions change depending on the purpose. Figure 4 shows that for the lines evaluated in 2020 the selection purposes will change which lines are selected. To select breeding lines for direct release then the total overall performance (vertical axis, Figure 4) is used, however to select lines for use as parents then their additive overall performance (vertical axis, Figure 4) is of interest. For this dataset, the selection decision will not change as it is essentially a 1:1 line. This result reflects the original observation that the additive to non-additive variance is higher for this dataset than is typically seen in other crops.

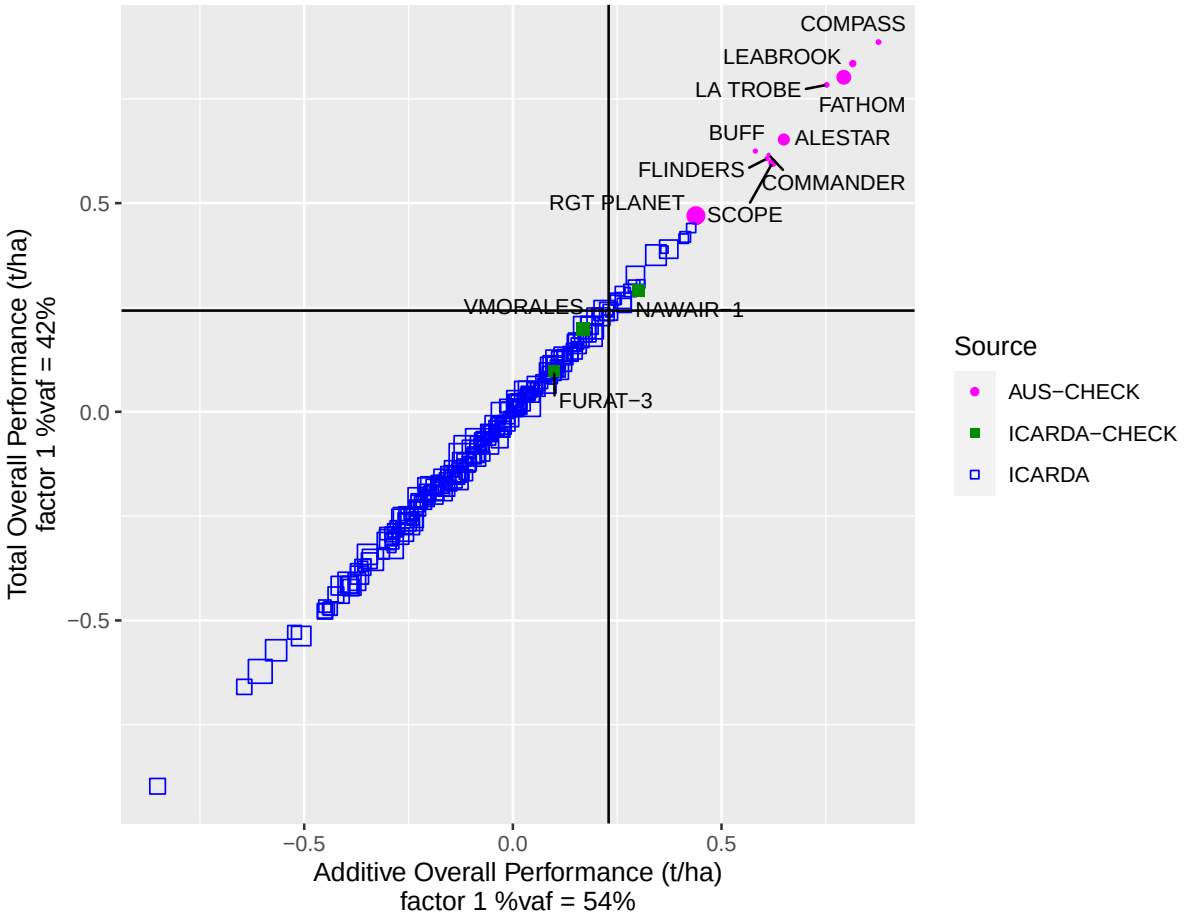

Figure 4: Overall Performance for Total and Additive effects for the 2020 entries in the 2015-20 barley dataset. Points to the top and right of the black horizontal and vertical lines, respectively, represent the top 10% cut-off. Australian and ICARDA checks are labelled. The size of the points indicates the stability (RMSD) of the line with smaller points indicating greater stability.

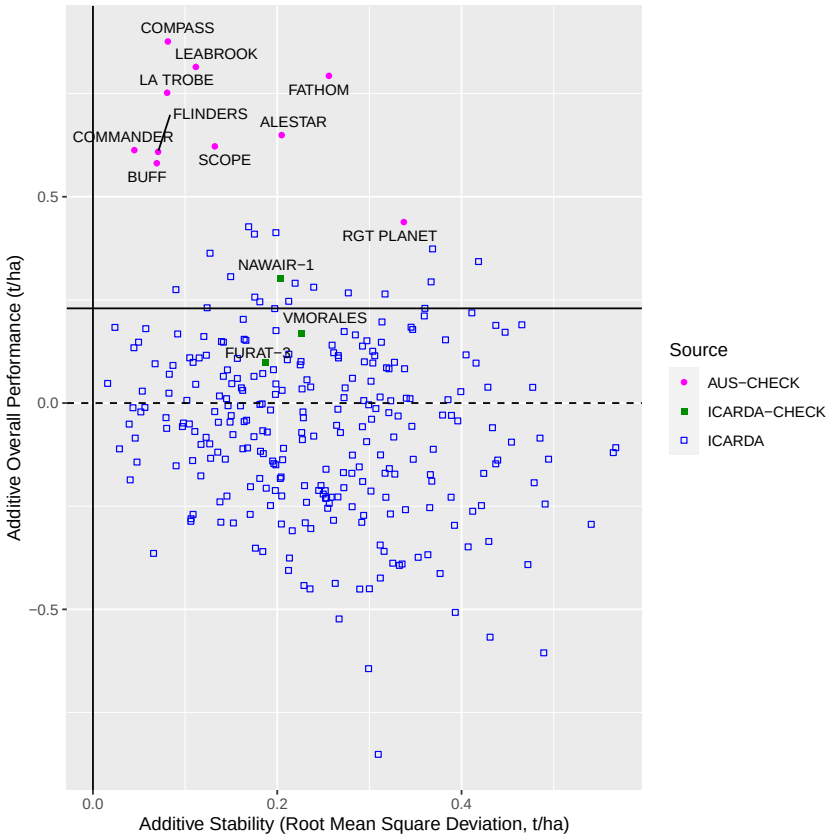

Figure 5: Factor Analytic Selection Tools: Overall Performance versus Stability (RMSD) for the additive genetic effects of the lines evaluated in 2020. The legend indicates the germplasm source. Australian and ICARDA checks are labelled. The solid horizontal line indicates the top 10% of additive overall performance for all entries from 2015-2020.

Overall performance (OP) and root mean square deviation (RMSD) for the additive effects summarise the line performance (Figure 5). The first factor for the additive effects, on which OP and RMSD is based, accounts for 54% of the additive genetic variance. Points in the top left hand corner represent lines with good additive performance and stability across the sampled environments. Figure 5 shows the subset of 2020 entries only from the full 2015-2020 dataset. Figures 4 and 5 demonstrate that none of the imported lines outperform the Australian varieties although there are a number in the top 10% for both additive and total OP (Figure 4).

The responsiveness statistics from FAST describe the remaining 34% of the 88% variance accounted for in the additive effects. Figure 6 presents the additive overall performance (OP, t/ha) compared to the responsiveness for factors 2 and 3, each in a separate panel. An inspection of the contrasting performance of the check varieties can inform as to the underlying environmental covariates driving these factors. For example, for factor 2 the responsiveness of RGT PLANET and VMORALES contrast whereas in factor 3 these two varieties are similarly responsive.

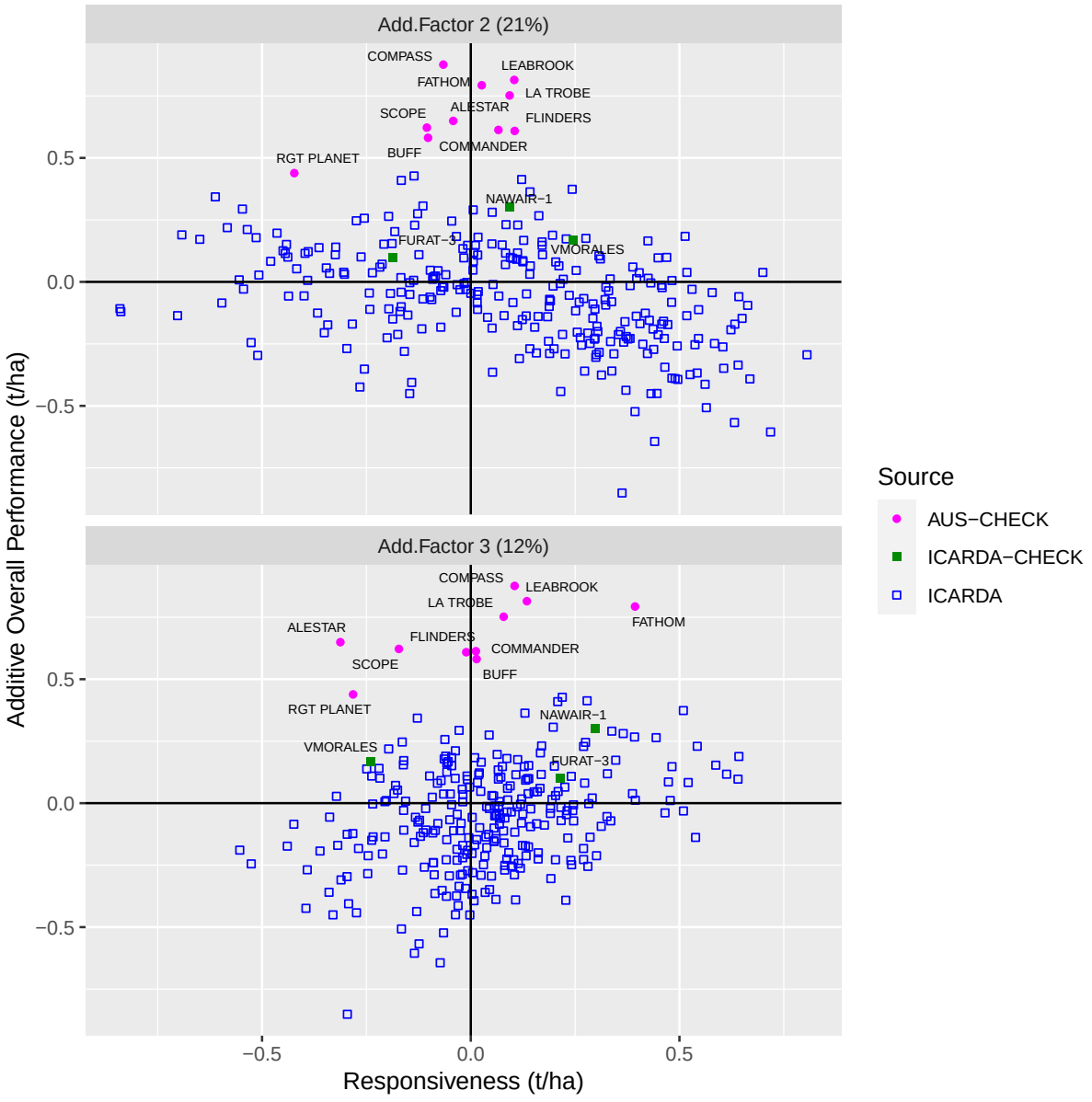

Figure 6: Factor Analytic Selection Tools: Overall Performance versus Responsiveness for the 2020 entries for the 2nd and 3rd factors fitted to the additive genetic variance-covariance matrix. The legend indicates the germplasm source. Australian and ICARDA checks are labelled.

5 Results

The number of lines by germplasm source which occur in the top 10% additive and total overall performers in each year cohort is of interest (Table 10). The numbers in this table have been summarised from the additive and total overall performance calculated for each line from the full dataset. The table shows that in each year that are lines from both CIMMYT and ICARDA which are in the top 10% additive and/or total effects which, based on other traits, may be of interest to the collaborating breeder.

Table 10: Number of lines by year and germplasm source that were evaluated, percentage in the top 10% additive and total overall performers of the CAIGE barley 2015-2020 yield dataset

Year	Australian	ICARDA
Number of lines		
2015	16	120
2016	17	133
2017	12	111
2018	14	139
2019	11	89
2020	10	276
Percentage of lines in top 10% additive		
2015	62	2
2015	62	2
2016	65	2
2016	71	2
2017	92	0
2017	92	0
Percentage of lines in top 10% total		
2018	86	1
2018	86	1
2019	82	0
2019	82	0
2020	100	7
2020	100	7

It is important to understand that the FAST measures OP and RMSD summarise the results from a FA-LMM concisely but they are based on only the first factor estimated from the data. In this dataset the first factor for the additive component explains 54% of the 88% additive genetic variance accounted for by the model. The remaining 34% is the variety by environment interaction effect, which is important to consider when selecting parents for Australian environments. The re-ranking of the lines within environments is clear when the the additive overall performance is plotted against the additive common variety by environment best linear unbiased predictions (additive CVE-eBLUPS) for each environment (Figure 7). This plot shows the results for the trials evaluated in 2020. The percentage of additive to total genetic variance for these four environments was 39, 46, 49 and 46%, respectively (Table 6). An example of specific additive adaptation are the Australian varieties RGT PLANET and FATHOM. Both varieties are high overall performers (above the horizontal 10% line), however FATHOM is in the top 10% additive CVE-eBLUPs in both B20GATT4 and B20YORK but not B20KALK3 or B20NARR2. RGT PLANET has an opposite performance, performing well in B20KALK3 and B20NARR2 but less so in B20GATT4 and B20YORK.

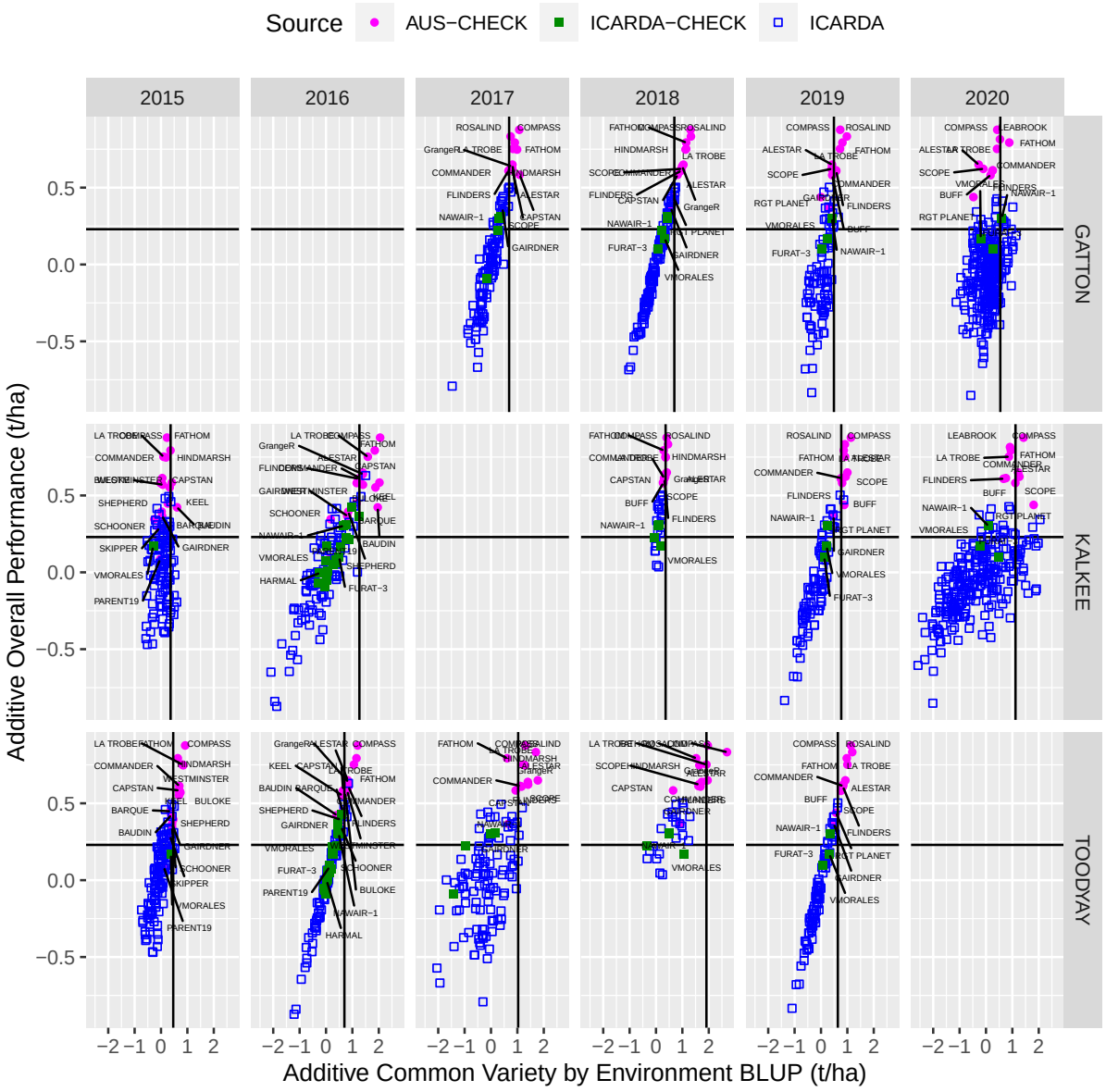

Figure 7: Additive Overall Performance compared to Additive Best Linear Unbiased Predictions (BLUPs) for lines evaluated at Gatton, Kalkee and Toodyay in 2015-2020. Points above and to the right of the black horizontal and vertical lines, respectively, represent the top 10% lines based on overall performance (above) and individual environment BLUPs (right). The legend indicates the germplasm source. Australian and ICARDA checks are labelled.

5 Results

A feature of these results is that the Australian checks outperform the ICARDA lines by a reasonable amount on an overall basis and in some individual environment. There is a change to this in the 2020 entries which reflects the newest phase of the CAIGE project, where the percentage of ICARDA entries in the top 10% increased from 1-2% in previous years to 7% for the 2020 entries. It is relevant to highlight that summary tables such as Table 10 can be misleading to those who do not take care with interpretation and relevance. It is not always the number of entries that help to improve a breeding program but the entries themselves which may have other traits, such as disease resistance, which contribute significantly to the improvement of germplasm. Figure 7 is more informative as it demonstrates that some 2020 entries are similarly performing to Australian checks in these environments. It also shows which environments benefit from the importation of ICARDA environments.

Table 11 presents the correlations between variety by environment BLUPS from the DIAG, DIAG-DIAG (with pedigree) and the final FA3-FA2 model. The correlation between the VE-eBLUPS of the most basic model (no pedigree and no VEI) and the Total VE-eBLUPS from the final model (pedigree and VEI modelled) have a correlation of 0.936. This correlation decrease as less information is included in the prediction from the final model. For example, the Total CVE-eBLUPS do not include the specific environmental effects and their correlation with the base model VE-eBLUPS is 0.868, Table 11. These correlation support the across year MET pedigree analysis results indicating that the results from this informative, complex model, are in line with the simpler models many collaborators are familiar with.

Table 11: Correlation between additive and total Best Linear Unbiased Predictions (BLUPS) from the DIAG (no pedigree, no VEI), DIAG-DIAG (pedigree, no VEI) and final FA3FA2 (pedigree and VEI) models for the CAIGE barley 2015-2020 yield dataset

Model	BLUP	DIAG	DIAG-DIAG		FA3-FA2			
		Total	Additive	Total	Additive CVE	Additive VE	Total CVE	Total VE
DIAG	Total	1.000	0.956	0.973	0.794	0.843	0.868	0.936
DIAG-DIAG	Additive	0.956	1.000	0.988	0.830	0.886	0.885	0.944
DIAG-DIAG	Total	0.973	0.988	1.000	0.830	0.881	0.891	0.956
FA3-FA2	Additive CVE	0.794	0.830	0.830	1.000	0.974	0.945	0.906
FA3-FA2	Additive VE	0.843	0.886	0.881	0.974	1.000	0.929	0.936
FA3-FA2	Total CVE	0.868	0.885	0.891	0.945	0.929	1.000	0.965
FA3-FA2	Total VE	0.936	0.944	0.956	0.906	0.936	0.965	1.000

The following results are provided in Excel worksheets:

- **VEeffects** - variety by environment effects for all variety by environment combinations. Columns indicating the number of replicates, presence of entry in the current year together with all passport information are provided alongside the additive and total CVE- and VE-eBLUPs, standard errors and ranks.

- **Trials** - environment level statistics including indications of TrialType (public or private), total and additive genetic variance and additive and total loadings for all factors. (Note, this tab name should probably be changed to **Environment**).

- **Varieties** - variety level statistics with passport information including the number of plots, environments and years an entry was evaluated in along side additive and total overall performance (OP), root mean square deviation (RMSD), responsiveness and scores.

- **Correlation matrices** - additive and total between environment correlation matrices

- **Connectivity matrices** - entry and parent level between environment connectivity matrices

The following yield result files were provided to the CAIGE team on 20 February 2021:

- *CAIGE Barley 2015-2020-METresults-ALL-2021-02-20.xlsx* - all results for all trials

- *CAIGE Barley 2015-2020-METresults-PUBLIC-2021-02-20.xlsx* - results only for publicly funded trials. FAST not provided because these are based on the whole dataset

- *CAIGE Barley 2015-2020top 10% and top 10 rank entries-2021-02-20.xlsx* - top 10% and top 10 ranked entries

- *CAIGE-Barley_DataCompilationFormat-2021-02-20.xlsx* - all results for the 2020 entries only provided in wide format for CAIGE team to add disease and other data too.

5.3 Agronomic traits: 2020

The phenology and height data were not consistently measured at all CAIGE barley trials in 2020 (Table 7). Consequently, all agronomic traits were modelled on an individual environment basis with pedigree in order to estimate the additive effects. The additive and non-additive genetic variances for each trait by environment combination is presented in Table 12.

Table 12: Additive and non-additive genetic variances for the phenology and height traits in the CAIGE Barley 2020 trials.

Trait	Env	Additive	NonAdditive
height	B20GATT4	15.78	0.00
height	B20YORK6	35.22	0.00
dth	B20NARR2	14.47	0.00
dth	B20YORK6	5.80	3.45
dtm	B20NARR2	3.67	0.00
relmaturity	B20GATT4	0.43	0.00
zadoks	B20KALK3	11.34	2.14

For the agronomic traits the individual environment variety additive and total BLUPs

were estimated and provided in wide format in the file *CAIGE-Barley-AgronomicTraitResults-2021-03-08.xlsx*.

6 Resources

This section is primarily as information for the author and team members of BBAGI.

The folder for this work is found on the BBAGI Dropbox account under `UOW-EIS-NIASRA-BBAGI/Projects/CAIGE/CAIGE20/Barley/Analysis/2-MET`.

The scripts for the analysis are located within the `1Scripts` folder and described below:

- 0-MET-analysis.R - initial diag model to determine best spatial models and test potential covariates
- 0a.CovariateCheck.R - checking covariates for use in the MET model
- 1-METped-analysis.R - pedigree MET analysis for the 2015-2020 dataset
- 2-METped-Results.R - results from the pedigree MET analysis of the 2015-2020 dataset.
- 4-AgronomicTraits-analysis-2020.R - head file for the analysis of the agronomic traits. From this file the 4a-HT-analysis.R, 4b-DTH-analysis.R, 4c-DTM-analysis.R, 4d-ZADOKS-analysis.R and 4e-RelMAT-analysis.R files are stepped into and then summarised at the end of the script.
- 4a-HT-analysis.R - height analysis
- 4b-DTH-analysis.R - days to heading analysis
- 4c-DTM-analysis.R - days to maturity analysis
- 4d-ZADOKS-analysis.R - maturity zadoks analysis
- 4e-RelMAT-analysis.R - relative maturity analysis

The purpose of the remaining folders is:

- `1Plots` – plots generated throughout the code
- `1RData` – .RData files saved throughout code
- `1Results` – output files sent to CAIGE staff.

The scripts for this report are found in
`UOW-EIS-NIASRA-BBAGI/Projects/CAIGE/CAIGE20/Barley/Analysis/Reports/1METped`.

References

Burgueno, J., Crossa, J., Cornelius, P. L., McLaren, G., Trethowan, R. & Krishnamachari, A. (2007). Modeling additve x environment and additive x additve x environment using genetic covariances of relatives of wheat genotypes. *Crop Science* **47**, 311–320.

Butler, D. (2019). *pedicure: pedigree tools*. www.mmade.org. R package version 2.0.0.

Butler, D. & Cullis, B. (2019). *od: Generate optimal experimental designs*. R package version 2.0.0, www.mmade.org.

Butler, D., Cullis, B. & Thompson, R. (in prep.). Asreml: An R package to fit the linear mixed model. *Journal of Statistical Software* .

Cullis, B., Cocks, N., Smith, A. & Butler, D. (2018). Sparse multi-environment trial designs for early stage selection experiments in plant breeding programmes. *National Institute for Applied Statistics Research of Australia Working Paper Series, University of Wollongong, (https://niasra.uow.edu.au/workingpapers/index.html)* .

Cullis, B., Smith, A., Beeck, C. & Cowling, W. (2010). Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. *Genome* **53**, 1002–1016.

Cullis, B. R., Smith, A. B., Cocks, N. A. & Butler, D. G. (2020). The Design of Early-Stage Plant Breeding Trials Using Genetic Relatedness. *Journal of Agricultural, Biological, and Environmental Statistics* .

Cullis, B. R., Smith, A. B. & Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 381–393.

Ganesalingam, A., Smith, A. B., Butler, D. G. & Cullis, B. (2019). Incomplete MET designs for early stage selection. Wheat Breeding Assembly, Adelaide.

Gilmour, A., Cullis, B., Welham, S., Gogel, B. & Thompson, R. (2004). An efficient computing strategy for prediction in mixed linear models. *Computational Statistics and Data Analysis* **44**, 571–586.

Gilmour, A. R., Cullis, B. R. & Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric., Biol. and Environ. Statist.* **2**, 269–293.

Gogel, B., Smith, A. & Cullis, B. (2018). Comparison of a one- and two-stage mixed model analysis of Australia' s National Variety Trial Southern Region wheat data. *Euphytica* **214**, 1–21.

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* , 69–83.

REFERENCES

Jordan, M. & Cullis, B. (2020). The Numerator Relationship Matrix. Tech. Rep. 1, Center of Bioinformatics and Biometrics, University of Wollongong.

Mathews, K., Wilson, R., Baranwal, D., Nicol, J. & Trethowan, R. (2021). CAIGE Bread Wheat 2017-2020 Across Year analysis. Tech. rep., University of Wollongong.

Mathews, K. L. (2021). Curation of CAIGE pedigree datasets. Tech. rep., BBAGI, University of Wollongong.

Meuwissen, T. H. E. & Luo, Z. (1992). Computing inbreeding coefficients in large populations. *Genetics, Selection and Evolution* **24**, 305–315.

Oakey, H., Verbyla, A., Pitchford, W., Cullis, B. R. & Kuchel, H. (2006). Joint Modelling of Additive and Non-Additive Genetic Line Effects in Single Field Trials. *Theoretical and Applied Genetics* **113**, 809–819.

Oakey, H., Verbyla, A. P., Cullis, B. R., Wei, X. & Pitchford, W. S. (2007). Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics* **114**, 1319–1332.

Pirathiban, R., Nicol, J., Trethowan, R. & Mathews, K. (2021). CAIGE Barley Designs 2021. Tech. rep., BBAGI, University of Wollongong.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Smith, A., Cullis, B. & Thompson, R. (2001). Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics* **57**, 1138–1147.

Smith, A. & Cullis, B. R. (2018). Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica* **214**, 1–19.

Smith, A. B., Ganesalingam, A., Kuchel, H. & Cullis, B. R. (2015). Factor analytic mixed models for the provision of grower information from national crop variety testing programs. *Theoretical and applied genetics.* **128**, 55–72.

Tolhurst, D. J., Mathews, K. L., Smith, A. B. & Cullis, B. R. (2019). Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. *Journal of Animal Breeding and Genetics* **136**, 279–300.

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist* , 330–8.

REFERENCES