



CBB
CENTRE FOR BIOINFORMATICS
AND BIOMETRICS

NIASRA
NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA



**UNIVERSITY OF
WOLLONGONG**



GRDC
GRAINS RESEARCH
& DEVELOPMENT
CORPORATION

Bioinformatics and Biometrics for the Australian Grains Industry

CAIGE Bread Wheat Analysis 2019

GRDC Project No.: US00073

Nicholas Lambert and Ky Mathews on behalf of the BBAGI team
email:nlambert@uow.edu.au

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

Julie Nicol, Sandra Micallef and Richard Trethowan
email:julie.nicol@sydney.edu.au

The Plant Breeding Institute
The University of Sydney

September 15, 2020

CONTENTS

Contents

- 1 Executive Summary** **3**

- 2 Introduction** **5**

- 3 Materials and Methods** **5**
 - 3.1 Experimental Designs 5
 - 3.2 Data cleaning 7
 - 3.3 Statistical Methods 7

- 4 Results** **8**
 - 4.1 Yield 8
 - 4.1.1 Modelling and percent variance accounted for (%vaf) 8
 - 4.1.2 Correlations and predictions 9
 - 4.2 Phenology, height and frost traits 11

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. As such, this is a confidential, internal document and it should not be distributed outside of the project stakeholders without prior written permission from the authors. The project stakeholders are GRDC staff with a need to know, BBAGI staff and the researcher named on this document. 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.

1 Executive Summary

This report describes the yield analysis of the CIMMYT Australia ICARDA Germplasm Evaluation (CAIGE) Bread Wheat MET dataset for 2019. Additional analyses were conducted on phenology traits, however these were performed on the individual-environment level. The report is structured as follows: an Executive Summary is provided in Section 1 presenting the key results, an Introduction in Section 2 provides background information on the CAIGE project; the materials and methods are described in Section 3 and the results presented in Section 4.

1 Executive Summary

- The 2019 CAIGE bread wheat dataset contained 12 environments and 312 unique varieties. The trial locations, environment by CAIGE nursery distribution and experimental designs for all environments are provided in Tables 1-3. The traits of interest are yield, days to heading, days to maturity, height and % frost damage.
- There were 10 environments with good quality yield data used in the final analysis. Balaklava and Roseworthy were both severely affected by frost and not included in the across environment yield analysis. The remaining traits were only measured at Narrabri. A *Septoria tritici* blotch score was measured at Balaklava but only for Block 1 and hence cannot be analysed.
- The variety connectivity between environments is excellent (≥ 170). The dataset is not balanced with not all varieties being represented in all environments, and many equally replicated within an environment
- An FA4 model was fitted to the yield dataset and the total variance accounted for was 67.68%. There was 1 environment where the %vaf was $< 50\%$. See Section 4.1 for further details.
- There was substantial variety by environment interaction (VEI) as presented in the between environment genetic correlation heatmap for yield Figure 2.
- Days to heading, maturity, height and %frost damage were analysed for the single environment Narrabri in the linear mixed model framework and the empirical best linear unbiased predictions provided.
- The EBLUPS of the CVE effects, environment loadings, variety scores, between environment genetic correlation matrix and variety connectivity matrix are provided in the accompanying Excel file, *CAIGE Bread Wheat 2019-results.xlsx*.
- The yield results are available in the Production Value Plus (<https://nvt.sagi.shinyapps.io/CAIGEPVPLUS/>) system (Smith et al., 2015) include factor analytic selection tools (FAST) (Smith & Cullis, 2018) to collaborators

1 Executive Summary

of the CAIGE project. It is password protected. Please contact Dr. Julie Nicol for access to the APP.

3 Materials and Methods

2 Introduction

CAIGE is a GRDC funded project to evaluate Bread Wheat, Durum Wheat and Barley germplasm developed by the International Maize and Wheat Improvement Centre (CIMMYT) and the International Centre for Agricultural Research in the Dry Areas (ICARDA).

The key objective of CAIGE is to evaluate germplasm developed by CIMMYT and ICARDA for potential introgression into Australian Wheat Breeding Programs. The germplasm is evaluated in different environments across the Australian wheat belt and selected by breeding companies to be included in their breeding programs and ultimately released to Australian wheat growers.

Trials were designed as p -rep trials by BBAGI staff for eight locations across four states of Australia; NSW (Edgeroi, Narrabri and Junee), VIC (Horsham), SA (Roseworthy and Balaklava) and WA (Toodyay and Dandaragan).

This report describes the analysis of the CAIGE Bread Wheat multi-environment trial (MET) analysis of yield for 2019, and the single environment analyses for phenology, height and %frost. The data are available via the CAIGE website (<http://www.caigeproject.org.au/>).

The fully efficient one-stage Factor Analytic (FA) approach of [Smith et al. \(2001\)](#) and [Gogel et al. \(2018\)](#) has been adopted for the analysis of CAIGE yield MET datasets. The results of this analysis are reported according to [Smith et al. \(2015\)](#) and [Smith & Cullis \(2018\)](#).

3 Materials and Methods

The CAIGE bread wheat 2019 trials were grown in 12 locations across the Australian grain growing region. The list of trial locations with institution responsible for managing the trial is presented in Table 1. In 2019, the CAIGE project was instructed by GRDC to only release results for trials that were funded by GRDC, these are indicated as Type ‘Public’ in Table 1.

3.1 Experimental Designs

The CAIGE 2019 bread wheat trials were designed in collaboration with Dr Julie Nicol and Professor Richard Trethowan. Julie provided the list of entries and numbers of seed packets available for each entry together with the design specifications for each environment. BBAGI performed the allocation of entries to environments and then the randomisation of entries to plots within each environment. A report was not produced for those designs by BBAGI so we will briefly summarise that process here.

3 Materials and Methods

Table 1: Trial locations for the CAIGE bread wheat 2019 trials.

Type	State	Location	GPS.Coordinates	Institution	SowingDate	HarvestDate
Commercial	NSW	Breeza	-31.2, 150.41	SWS Seeds	17-08-19	16-12-19
Commercial	NSW	Junea	-34.74, 147.6	LPB	01-06-19	09-12-19
Public	NSW	Narrabri	-30.28, 149.8	USYD	12-05-19	02-11-19
Commercial	NSW	Northstar	-28.93, 150.39	AGT	15-05-19	15-10-19
Commercial	SA	Balaklava	-34.33, 138.48	LPB	17-05-19	05-12-19
Commercial	SA	Roseworthy	-34.51, 138.7	AGT	17-04-19	26-11-19
Commercial	VIC	Kalkee	-36.51, 142.15	Intergrain	10-06-19	12-12-19
Public	VIC	Swan Hill	-35.31, 143.22	USYD (KAYLX)	12-05-19	12-11-19
Public	WA	Corrigin	-32.26, 117.94	USYD (KAYLX)	02-06-19	29-11-19
Commercial	WA	Dandaragan	-30.69, 115.68	Intergrain	16-06-19	15-12-19
Commercial	WA	Goomalling	-31.55, 116.99	LPB	29-05-19	10-12-19
Commercial	WA	Toodyay	-31.32, 116.45	EdStar Genetics	28-06-19	28-11-19

The entry list contains 14 Australian check varieties and 298 breeding lines from CIMMYT nurseries (SBP = Soil Borne Pathogen, RAV = Ravi Singh selections, ZSA = Stress Adaptive Trait Yield, ZWB = Spring Wheat Breeder selections and ZWY = Wheat Yield Consortium Yield Trials), ICARDA nurseries (ZIF and ZIZ = Spring Bread Wheat, ZRT = Spring Bread Wheat selected by Richard Trethowan) and selections from Dr Rajaram (ZIG). There were 218 entries that were not present at all sites. A summary of the number of entries by site and nursery is provided in Table 2.

Table 2: Number of entries evaluated at each site by CAIGE nursery in the 2019 Bread Wheat MET dataset.

Environment	CHECK	RAV15	SBP17	SBP18	ZIF16	ZIG18	ZIZ17	ZIZ18	ZRT15	ZSA18	ZWB17	ZWB18	ZWY18	Total
Balaklava	14	16	5	7	0	7	1	43	4	5	1	89	11	203
Breeza	14	16	5	12	1	10	1	58	4	6	1	93	13	234
Corrigin	14	16	5	13	1	8	0	61	4	7	1	101	14	245
Dandaragan	14	16	4	10	1	7	1	54	4	5	1	93	11	221
Goomalling	14	16	5	12	0	9	1	64	4	7	1	96	16	245
Junea	14	16	4	14	1	7	2	66	4	8	1	94	15	246
Kalkee	14	16	5	10	1	8	1	49	4	6	1	90	14	219
Narrabri	14	16	5	20	1	17	2	97	4	10	1	107	18	312
North Star	14	16	4	3	1	5	1	35	4	5	1	83	8	180
Roseworthy	14	10	5	20	1	16	2	83	0	8	1	85	14	259
Swan Hill	14	16	5	10	1	9	1	58	4	9	1	95	10	233
Toodyay	14	16	5	8	1	8	1	50	4	7	1	93	13	221
Total	14	16	5	20	1	17	2	97	4	10	1	107	18	312

The experimental design specifications, such as number of ranges and rows for each trial were provided by Julie Nicol. For 10 trials the blocks were in the range direction, whilst at Breeza and Toodyay the blocks were in the row direction. BBAGI used the design specifications from all sites together to allocate the entries to plots such that the average pairwise variance within and across sites is minimised. The resulting MET design specifications are detailed in Table 3. This design was performed in the R statistical software (R Core Team, 2018) using the optimal design library (od) (Butler & Cullis, 2019).

3 Materials and Methods

Table 3: Summary of the experimental design parameters for the 2019 CAIGE bread wheat trials, including the number of plots, ranges, runs, blocks and varieties. The number of varieties with 1, 2 and 3 plots, respectively, is presented with the % p -replicate.

Site	Plots	Range	Run	Block	Varieties	$p=1$	$p=2$	$p=3$	% p -rep
Balaklava	288	12	24	2	203	142	37	24	30
Breeza	288	12	24	2	234	191	32	11	18.4
Corrigin	288	12	24	2	245	202	43	0	17.6
Dandaragan	288	12	24	2	221	177	21	23	19.9
Goomalling	288	12	24	2	245	202	43	0	17.6
June	288	12	24	2	246	204	42	0	17.1
Kalkee	288	12	24	2	219	173	23	23	21
Narrabri	624	24	26	2	312	0	312	0	100
North Star	288	24	12	2	180	89	74	17	50.6
Roseworthy	288	24	12	2	259	230	29	0	11.2
Swan Hill	288	12	24	2	233	189	33	11	18.9
Toodyay	288	12	24	2	221	177	21	23	19.9
Total	3792				312				

3.2 Data cleaning

There were 6 traits, including yield, provided in the bread wheat dataset. Yield (t/ha) is the key trait of interest but other traits may be used to explain yield performance. Some of these traits could potentially be used as covariates (e.g % frost damage at Narrabri) and some, such as establishment scores, were not intended for analysis but rather to assist with trial management. Other traits of interest, all measured at Narrabri only, are days to heading, days to maturity and height (cm). *Septoria tritici* blotch (STB) was also scored at Balaklava.

The following issues were found in the data file:

1. the data worksheet and the list of entries in the provided xlsx file *caigeyieldtrials2019_breadwheatyielddata.xlsx* has different headings due to lower and upper case and punctuation. The information in the data worksheet was used.
2. in the data file, at Breeza Line = EGA GREGORY in Range 12, Run19 has QCode = ZWB18 and QNo = 263. Fixed in data file.
3. in the data file, there was inconsistency in the labelling of NA values for GID, pedigree and selection history with values of “-”, NA and 0 used interchangeably for the same Line across sites.

3.3 Statistical Methods

For full documentation on the method and theory of the FA approach used in the one-stage MET analysis, readers are referred to [Gogel et al. \(2018\)](#), although a brief summary is provided here. The MET analysis involved two steps: (i) individual trial analyses were performed in order to determine the most adequate spatial models that accounted for the

4 Results

presence of global and extraneous trends exhibited in the field. In this step variety effects were fitted as fixed and outliers identified and removed from all subsequent analyses. All models fitted here were saved and stored for inclusion during the following step. An independent structure was then fitted to the between-environment genetic variance matrix (i.e. variety effects now fitted as random and environments treated as independent) to examine whether any environments had zero genetic variance. Next, (ii) FA structures of order 1 through k were fitted to the between environment genetic variance-covariance matrix, which models heterogeneous environment variances and heterogeneous covariance between environments. The order of the FA model, k , is increased incrementally in order to achieve a satisfactory level of the total VE variance accounted for by the k common factors. The final model is determined using formal statistical criterion (AIC and Log-likelihood ratio tests) and once a satisfactory amount of VE variance is explained ($\%vaf > 80\%$, where possible).

All analyses were conducted in the R software environment (R Core Team, 2018) with the ASReml-R (v4.0) package (Butler et al., in prep.) used for fitting the linear mixed models.

In 2019 GRDC requested that there is now a delineation between publicly (GRDC) and breeder funded trials in the CAIGE project. Three of the twelve bread wheat CAIGE trials in 2019 are publicly funded (see Table 1). In order to provide the best possible predictions for this dataset all trials have been included in one analysis. However, only results from the publicly funded trials are available to all parties. Results from the commercially funded trials are only available to collaborators as defined by CAIGE staff.

4 Results

In this section, we provide an overview summary of all traits and then specifically comment on the models fitted, the between environment genetic correlations and provide a PV-PLUS plot for the publicly funded trials and corresponding genetic correlation heatmap.

4.1 Yield

4.1.1 Modelling and percent variance accounted for ($\%vaf$)

Yield data were measured on all 12 trials, however, Balaklava and Roseworthy trials were severely affected by frost and were not included in the across environment yield analysis. The final dataset contained 10 environments and 312 varieties.

Table 4 presents a summary of the models fitted to the CAIGE Bread wheat 2019 MET yield data set. The final model selected was an FA model of order 4, where the total $\%vaf$ was 67.68% and the factor $\%vaf$ for the 4 factors were 34.5, 15.5, 8.9 and 8.7, respectively. The $FA4$ model was determined based on a log-likelihood ratio test (LRT) to be the best model. The $FA5$ model accounted for more of the genetic variance $\%vaf$ (72%) but it is not significantly different from the $FA4$ model using a formal LRT. The environment mean

4 Results

yield and the percentage variance accounted for in each public environment is presented in Table 5. The results for the full dataset is provided in *CAIGE Bread Wheat 2019-results.xlsx*.

Table 4: Summary of the models fitted to the CAIGE Bread wheat MET yield dataset including the residual model fitted, the number of genetic variance parameters estimated, the residual log-likelihood (Residual Loglik), Akaike information criterion (AIC), the percentage of total variety by environment variance explained by the common factors (i.e. %vaf) in terms of the number of environments below 50%, above 80% and overall mean.

Model	Residual	No. parameters	Residual Loglik	AIC	%vaf		Mean
					< 50	> 80	
DIAG	AR1×AR1	10	1960.7	-3805.5	1	1	
FA1	AR1×AR1	20	2068	-3999.9	7	0	37.5
FA2	AR1×AR1	29	2111.9	-4069.8	4	0	53.9
FA3	AR1×AR1	37	2121.2	-4072.5	1	0	61.8
FA4	AR1×AR1	44	2125.4	-4066.9	1	2	67.7
FA5	AR1×AR1	50	2128.2	-4060.4	0	3	72.1

Table 5: Environment mean yield and site percentage variance accounted for in the public trials in the CAIGE Bread Wheat 2019 MET yield dataset

Environment	Mean Yield (t/ha)	site %vaf
Corrigin	0.86	50.17
Narrabri	3.76	56.14
Swan Hill	0.62	46.25

4.1.2 Correlations and predictions

The between environment genetic correlations for all publicly funded trials is presented as a heatmap in Figures 2. This heatmap shows that there is substantial cross-over VEI present in bread wheat for yield.

The empirical best linear unbiased predictions (EBLUPs) of the common variety by environment (CVE) effects, environment loadings, variety scores, between environment genetic correlation matrix and environment variety connectivity matrix are provided to CAIGE staff in the accompanying excel spreadsheet, *CAIGE Bread Wheat 2019-results.xlsx*.

In addition, these results are available in the Production Value Plus (PV-PLUS) system (Smith et al., 2015), which is one recommended way to interrogate the MET analysis results from a factor analytic linear mixed model. For people with access to the publicly funded trials only PV-Plus and correlation heatmaps are available. For people with access to the full set of results there are PV-PLUS plots, correlation heatmaps, latent regression plots and factor analytic selection tools (FAST) as per Smith & Cullis (2018). Please contact Dr. Julie Nicol if you would like access to the commercially funded trial results.

4 Results

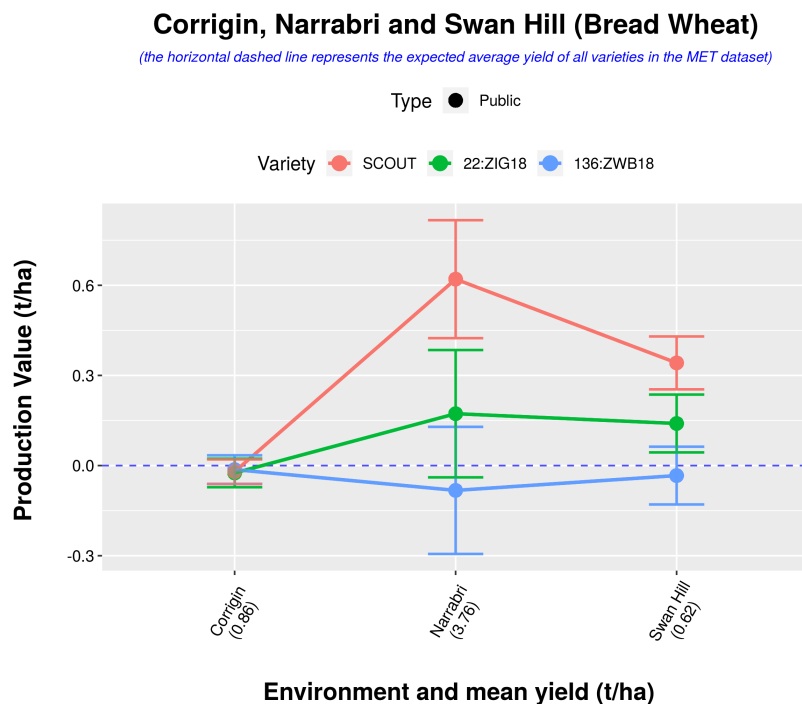


Figure 1: PV-PLUS plot for public CAIGE bread wheat trials 2019.

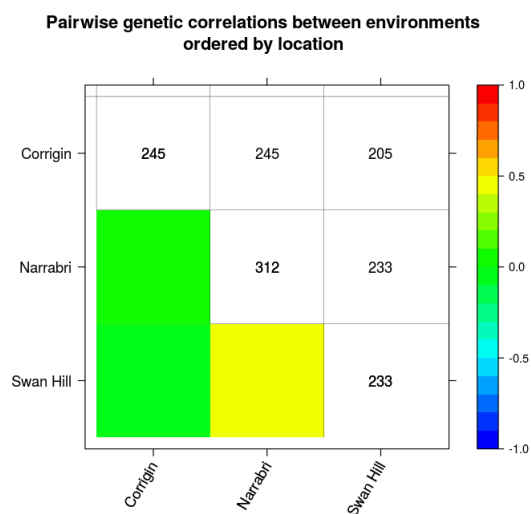


Figure 2: Heatmap of the residual maximum likelihood (REML) estimate of the total genetic correlation matrix (with connectivity on upper triangle) for public CAIGE bread wheat trials, 2019.

4 Results

4.2 Phenology, height and frost traits

Days to heading, days to maturity, height and percentage frost damage were recorded at the home site, Narrabri. A best spatial model was fitted for each trait where the **Variety** term is fitted as a random effect in line with the aim to quantify the amount of genetic variation and variety rank for these traits. The residual estimated maximum likelihood (REML) variance parameter estimates for all terms in each model, together with the reliability for each trait is provided in Table 6.

Table 6: REML variance parameter estimates and reliability for days to heading, maturity, height and %frost damage measured at the CAIGE bread wheat trial at Narrabri, 2019.

Term	Heading	Maturity	Height	Frost
Line	7.73	13.15	40.82	195.88
Block	0.11	0.00	0.05	16.38
Range	0.26	0.49	3.12	75.42
Run	0.10			
Range:Run!R	0.55	2.33	9.98	530.64
Range:Run!Range!cor	0.16	0.14	0.12	0.29
Range:Run!Run!cor	0.26	0.21	0.06	0.10
Reliability	0.96	0.92	0.88	0.14

The best linear unbiased predictions (BLUPs) for both these traits are presented in the worksheet *Extra Traits* in *CAIGE Bread Wheat 2019-results.xlsx*.

REFERENCES

References

- BUTLER, D. & CULLIS, B. (2019). *od: Generate optimal experimental designs*. R package version 2.0.0.
- BUTLER, D., CULLIS, B. & THOMPSON, R. (in prep.). Asreml: An R package to fit the linear mixed model. *Journal of Statistical Software* .
- 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.
- R CORE TEAM (2018). *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. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **128**, 55–72.

REFERENCES
