

NIASRA

NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA



National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

08-18

**A Statistical Framework for the Provision of Grower and
Breeder Information on the Frost Susceptibility of Wheat
in Australia.**

Nicole. A. Cocks, Timothy. J. March, Thomas. B. Biddulph,
Alison. B. Smith, and Brian. R. Cullis

*Copyright © 2017 by the National Institute for Applied Statistics Research Australia, UOW.
Work in progress, no part of this paper may be reproduced without permission from the Institute.*

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia Phone +61 2 4221 5076, Fax +61 2 4221 4998.
Email: karink@uow.edu.au

A statistical framework for the provision of grower and breeder information on the frost susceptibility of wheat in Australia.

Nicole. A. Cocks^{a,*}, Timothy. J. March^{b,1}, Thomas. B. Biddulph^c,
Alison. B. Smith^a, Brian. R. Cullis^a

^a*Centre for Bioinformatics and Biometrics, National Institute for Applied Statistics
Research Australia, University of Wollongong, NSW, Australia*

^b*University of Adelaide, School of Agriculture, Food and Wine, SA, Australia*

^c*Department of Agriculture and Food, WA, Australia*



*Corresponding author

Email addresses: ncocks@uow.edu.au (Nicole. A. Cocks), email@timothymarch.com
(Timothy. J. March), Ben.Biddulph@dpird.wa.gov.au (Thomas. B. Biddulph),
alismith@uow.edu.au (Alison. B. Smith), bcullis@uow.edu.au (Brian. R. Cullis)

¹Present Address: Rijk Zwaan Australia Pty Ltd, Daylesford, Vic, Australia.

Abstract

The frost susceptibility of Australian commercial cereal crops, in particular wheat and barley, has become an economically devastating issue for growers. It is crucial that growers are aware of the relative risk to frost damage of the currently available varieties. This information is obtained through the testing of varieties in a series of field experiments at locations susceptible to frost events. The experimental designs, measurement protocols and the resultant data from these frost expression experiments (FEEs) are necessarily complex due to the unpredictability of the timing and severity of frost events, and the maturity of the plants at the time of the events. Design and protocol complexities include the use of multiple sowing dates and the recording of plant maturity. Data difficulties include a high degree of unbalance, and in the instance of multiple frosts in an FEE, there is a longitudinal aspect. In order to accommodate all these elements, we use a linear mixed model analysis. Specification of a suitable model for these data is nontrivial so we utilise the so-called Design Tableau (DT) approach. DT is a general method for determining an appropriate linear mixed model for a comparative experiment. It comprises a series of straight-forward steps aimed at removing the subjectivity in model specification and ensuring that the key principles in the analysis of comparative experiments are not neglected. We describe in detail the use of DT to derive the linear mixed model for a single FEE, then extend to the complete multi-environment trial analysis of 17 experiments. We show that, in our example, the DT approach guards against invalid inference due to false replication and allows appropriate modelling of variety by environment interaction.

1. Introduction

Reproductive frost damage has been estimated to cost Australian grain growers approximately A\$360m in direct and indirect yield losses annually (Baxter, 2015). These losses are most significant for wheat and barley, given their relatively low levels of tolerance to frost damage during flowering (Al-Issawi et al., 2012). Breeding new varieties with reduced frost susceptibility is one possible solution to this problem. Although genetic variation for reproductive frost tolerance has been reported in barley (Reinheimer et al., 2004), the extent of genetic variation across all commercial wheat and barley cultivars remains unknown. This lack of knowledge has been partly a result

of the practical difficulties in accurately measuring reproductive frost damage under field conditions due to the random and spatial variability of frost damage. Furthermore, the unpredictability of frost events aligning with the most sensitive phase of the reproductive stage of development of wheat and barley has also proved challenging.

In contrast to grain yield, statistical design and analysis methods specific to the prediction of a variety's genetic propensity to tolerate frost, hereafter referred to as frost value (FV), have been far less extensively researched. Like many grain quality traits in wheat, determining a variety's FV is a complex process. This is due to issues such as low heritability, difficulty in determining an appropriate phenotype to measure, developmental differences between varieties at the time of a frost event and the unpredictability of when a frost event will occur and its severity. The duration and interaction with other microclimate factors suggests that frost damage is not consistent across each event. The exact definition of what constitutes a frost event is also unclear, as sterility may also be possible under milder climatic conditions termed chilling (Dolferus et. al. in press). Screening for frost risk in Australia is the primary objective of a project funded by the Grains Research and Development Corporation (GRDC), known as the Australian National Frost Project (ANFP). This project operated from 2012-2016 and has since been replaced by a new project with similar aims: Screening for Frost Tolerance in Cereals (SFTC). A key component of these projects is to conduct field experiments to investigate the FV of a range of commercial and advanced varieties of wheat and barley.

A statistical framework for the analysis of such complex data and the subsequent provision of information that is both accurate and invaluable to local growers is a primary output of the ANFP (and SFTC project). In order to achieve these reliable predictions the trait of interest, the measurement protocols and the structure of the data must be identified and understood to formulate an appropriate method for data analysis and subsequent reporting of results. Smith et al. (2015) described an approach to the analysis of crop variety evaluation data, where variety trials across years and locations are collated into a multi-environment trial (MET) dataset. Their approach is based on the factor analytic mixed model of Smith et al. (2001) and they demonstrate the utility of this approach in terms of reporting results at a local or grower level. More recently, Smith and Cullis (2018b) developed a set of summary tools in the context of plant breeding selections for early stage yield evaluation trials. These tools can be generalised to all MET analyses

in the factor analytic mixed model paradigm and are not mutually exclusive to those described in Smith et al. (2015). The results from the analysis can therefore be summarised not only at the grower level but also for breeders to identify and explore the genetic composition of varieties that have lower frost risk than others.

The key trait which is related to reproductive frost damage is the so-called frost induced sterility (FIS) of the plant head (Reinheimer et al., 2004). FIS is simply the ratio of sterile florets to total florets on a per head basis. In order to maximise yield potential, commercial wheat crops are sown in autumn, which typically results in the plants flowering during late winter (barley) or early spring (wheat). Typically, wheat and barley have sufficient cold tolerance in their vegetative stage of development to survive the relatively mild winters in Australia. However, when they transition into the reproductive stages during spring, the temperature overnight can still drop low enough over a prolonged period of time to cause a frost event and thus potential FIS at this vulnerable stage of development.

The level of sensitivity to frost is thought to continue to increase as the spike emerges from the boot and begins to undergo flowering (Frederiks et al., 2015). The experimental phenotyping methodology implemented in the ANFP project was based on that of Reinheimer et al. (2004) and involved multiple sowing dates as separate field blocks at approximately equidistant thermal time. The strategy of multiple sowing dates was implemented to increase the likelihood that varieties would be at the booting or flowering stage of development i.e. susceptible to damage when a frost event occurred. Furthermore, it enables varieties of dissimilar maturity to be compared at the same stage of development for a given frost event. The sample of tillers that have heads at the appropriate stages of development are tagged for a given frost event and FIS is measured 4 to 6 weeks later in the laboratory once the plant has developed and achieved grain set ($\sim Z85$). This is required in order to distinguish between the viable and sterile florets and thus exposure to multiple frost events is unavoidable, where FIS is a combination of both frost damage to anthers and ovules at or around flowering ($Z65$) and developing grains ($Z70.2-79$). Separation of the aborted anthers/ovules from aborted/frosted grains would require tedious dissection, therefore presence or absence is recorded on frozen samples which indicate if a whole grain is present or absent at $\sim Z85$. This process is repeated for all successive frost events, where some tillers may be tagged more than once if they are at two different stages of development for separate events. This protocol results in

a highly unbalanced longitudinal experiment (Brien and Demetrio, 2009), in which there is potentially unbalanced sampling of each field plot in the experiment for each frost event.

The aim of this paper is to develop an approach to the analysis of the MET frost expression experiment (FEE) dataset to provide local growers and breeders with useful and relevant information on the relative frost risk of wheat varieties. The formulation of an appropriate model that accommodates the features of the dataset poses many statistical issues. The MET dataset presents a complex plot structure as the trait measured is not only sparse but also repeated in the instance of multiple frost events which gives rise to longitudinal data. There are also many other variables that have been measured and can be included as potential covariates in the analysis. Finally, we also aim to capture the variety \times environment interaction (VEI) across the MET dataset, where the factor analytic model has been shown to provide a parsimonious and informative model for both VEI and accurate predictions of variety effects (Smith et al., 2015).

The Design Tableau (DT) approach advocated by Smith and Cullis (2018a) for the analysis of comparative experiments was adopted in this setting. DT comprises a series of straight-forward steps aimed at removing the subjectivity in model specification and ensures the key principles in the analysis of these experiments are not neglected. The derivation of a suitable linear mixed model for these data to capture the aims of the analysis was achieved by careful consideration of the experiment design, measurement protocol and data structure.

The paper is arranged as follows. In Section 2 the experiment design for a hypothetical and simple FEE is presented and the measurement protocol for a FEE is outlined. We describe in detail the use of DT to derive the linear mixed model for a single FEE, then extend to the complete multi-environment trial analysis of 17 experiments. The model fitting process and tools for interpreting the final fitted model are also described. The results for the final fitted model to the MET dataset and tools developed to summarise information on frost risk for growers and breeders are presented in Section 3. Section 4 provides a summary of all concepts introduced in the formulation of the model used for the analysis and impact of the tools described in Section 3.

2. Materials and Methods

The FEEs in the MET dataset for this paper were established in three states (NSW, SA and WA) from 2012-2016 inclusive. The locations used included Narrabri (NSW), Loxton (SA), Brookton (WA), Merredin (WA) and Wickepin (WA). Two additional experiments were sown in 2010 and 2011 at Loxton and they were included in the MET dataset as many varieties in these trials were the same as those in 2012-2016 and were grown using the same or similar designs and measurement protocols.

2.1. Experimental Design

There are three elements to be considered in the (model-based) design of comparative experiments; namely the plot structure, treatment structure and the design function (Bailey, 2008). To facilitate the explanation of the experimental design, a hypothetical and simple FEE is considered in which there are three times of sowing (TOS) and 48 varieties. Figure 1 presents the layout of such, where each of the field blocks are not necessarily spatially adjacent. In this example, each block comprises two sub-blocks with 48 field plots in each. The field plots in Figure 1 with an “X” are those with FIS data recorded over the course of the experiment and will be addressed in Section 2.2.

The design function describes the allocation of the treatments to the plots, where in this case a treatment is the combination of a TOS and a variety. The design function therefore involves assignment of each individual TOS to a whole block and allocation of the 48 varieties to field plots within sub-blocks, such that the sub-blocks within a block are resolvable with respect to varieties. The subsequent plot and treatment structures that arise from this design function are presented in Section 2.3 in the context of the real data.



Figure 1: Field layout of a hypothetical FEE with three TOS and 48 varieties. Each of the three field blocks (Block) contain two sub-blocks and 48 field plots per sub-block. The “X” indicates that there is FIS data recorded for tillers within that plot

2.2. Measurement Protocol

It is known that the susceptibility to frost damage increases with the maturity of the plant (Al-Issawi et al., 2012). The anthers within the florets on the developing wheat spike are responsible for producing pollen that is required for fertilisation of the florets (and thus grain development) and are particularly sensitive to frost damage. In the case of sterile anthers due to frost damage, this process is not initiated and so the number of grains that are formed and viable for harvest is less than the total number of expected grains for that particular head. Pollen development begins while the spike is still contained within the boot and the sensitivity to cold damage increases during pollen meiosis (Barton et al., 2014). As a result, the protocol for these experiments stipulates that the plants be tagged at the booting or flowering stage(s) of development following a frost event in order to capture range in susceptibility over time.

The trait measured (i.e. FIS) is used to derive the genetic response of interest (FV), which depends on the definition of a so-called frost event. Here a frost event is defined by the canopy temperature of a trial falling to 0°C or below. Tillers are tagged if they are at a susceptible stage of development (SOD) i.e. flowering or booting, within (i.e. +/-) 48 hrs. of a frost event. Hence the SOD of each (sampled) tiller is recorded for each frost event (hereafter referred to as tag event).

Ancillary biotic data was also recorded either at the field plot or tiller level. All field plot measurements are taken on templated tillers, where a templated tiller is defined to be a single tiller in that plot chosen to reflect the cohort of remaining tillers tagged for that SOD. The set of field plot level measurements included the percent of tillers within a field plot at the same SOD as the template (hereafter referred to as PercentPlotStage); height of the templated tiller (hereafter referred to as Height); the distance from the flag leaf collar to the base of the spike on the templated tiller (PeduncleLength); the name of the technician who tagged the tillers in a plot at a given tag event (hereafter referred to as the Tagger). The measurements that are collected at the tiller level are the total number of seeds in the head of the plant (TotalNumberOfSeeds), total number of sterile seeds in the head (NumberOfSterileSeeds); the length of the spike on the tagged tiller (hereafter referred to as Length); the name of the technician who counted the total number of seeds and sterile seeds in the head of the tiller (hereafter referred to as Counter).

There are a number of unique growth stages recorded at tagging. The unique identifiers of growth stage at tagging are ear peep, boot, flowering and not recorded. The latter encompasses stages that are synonymous with cases where FIS has been recorded outside the sampling protocol i.e. tillers sampled at random throughout the course of the experiment or tagged at a different SOD to those pre-specified in the protocol. FIS records for tagged tillers that are classified as “not recorded” cannot be included in the statistical analysis and are therefore removed from the final dataset. For the purpose of analysis, the remaining stages of development were then reduced to two stages, namely flowering (Z65) and boot (Z49), where the latter encompasses the stages boot (Z45) and ear peep (Z49). Further data cleaning was undertaken to remove data records where FIS exceeded 1 and/or the total number of floret positions was less than 10 or greater than 60, as these were regarded as either unreliable or unreasonable measurements.

Table 1 presents a summary of key features of the MET dataset. Recall from Figure 1 that the FIS data collected from these experiments is sparse and therefore the number of TOS, field plots and varieties sampled (i.e. that data exists for) is often not the total number for each of these factors in the designed experiment. In other words, the total number of TOS in Figure 1 is three but the number of TOS for which there is FIS data is two.

The number of tag events for each of the FEEs varied from two (**sa11**) to 17 (**nsw12**), while the number of TOS sampled ranged from two to eight. The total number of plots sampled also varied markedly between FEEs and there was a significant degree of sparsity in the data at the TOS, variety and field plot by tag event levels. It is clear that for four of the 17 FEEs (**nsw13**, **nsw16**, **sa11** and **wa12**), all sampled field plots were unique to a single tag event, whereas **nsw12** had 70% of all field plots that were sampled present for at least two of the 17 tag events that occurred. The severity of FIS also varied substantially between FEEs (MeanFIS), while the range in total number of observations for each experiment was 2698 (**sa10**) to 36889 (**sa16**). Finally, with the exception of **nsw16**, **sa10** and **sa11**, there were data records for both SOD for each of the FEEs, however only 14 of the 74 (18.9%) tag events across all 17 experiments had tillers tagged at both SOD (data not shown).

Figure 2 presents the variety connectivity matrix between the 17 FEEs in the MET dataset. The upper triangle presents the number of varieties in common between pairs of experiments, the lower triangle is a heatmap representation of the connectivity and the diagonals are the number of varieties

Expt	TE	TOS	FPlot	TOSxTE	VarietyxTE	FPlotxTE	MeanTExFPlot	MeanFIS	Records
nsw12	17	4	184	0.47	0.38	0.25	1.70	0.32	9340
nsw13	3	3	99	0.33	0.70	0.33	1.00	0.66	4334
nsw14	3	3	148	0.44	0.62	0.33	1.04	0.34	5726
nsw15	4	3	135	0.50	0.51	0.33	1.26	0.12	7816
nsw16	2	2	79	0.50	0.66	0.50	1.00	0.17	3233
sa10	2	2	62	0.75	0.93	0.50	1.53	0.20	2698
sa11	2	2	103	0.50	0.84	0.50	1.00	0.02	2782
sa12	4	6	384	0.67	0.82	0.17	1.57	0.17	13432
sa13	4	3	225	0.58	0.77	0.33	1.69	0.08	8838
sa14	3	3	291	0.44	0.53	0.33	1.04	0.74	10011
sa15	2	6	566	0.67	0.96	0.17	1.07	0.23	26985
sa16	4	5	688	0.55	0.95	0.20	1.32	0.30	36889
wa12	3	3	107	0.78	0.42	0.33	1.00	0.08	3151
wa13	5	6	250	0.57	0.61	0.17	1.19	0.05	8904
wa14	7	8	534	0.48	0.64	0.12	1.10	0.15	19828
wa15	6	8	812	0.58	0.88	0.12	1.36	0.15	35126
wa16	3	7	500	0.76	0.84	0.14	1.09	0.65	16908

Table 1: Summary of the total number of tag events (TE), TOS and field plots (FPlot) sampled for each of the 17 FEEs in the MET dataset. A measure of sparsity is also provided for $TOS \times TE$, $Variety \times TE$ and $FPlot \times TE$ combinations that exist, presented as a proportion of all possible combinations for each experiment. The mean number of TE for each FPlot, mean FIS and total number of data records are also summarised for each experiment in the MET dataset

that were sampled in each FEE. With the exception of **sa10** and **sa11**, the connectivity between experiments in general is quite reasonable (more than 20 varieties in common between pairs of FEEs) considering the stochastic nature of frost events and subsequent alignment with the required phenological development stages of the crop.

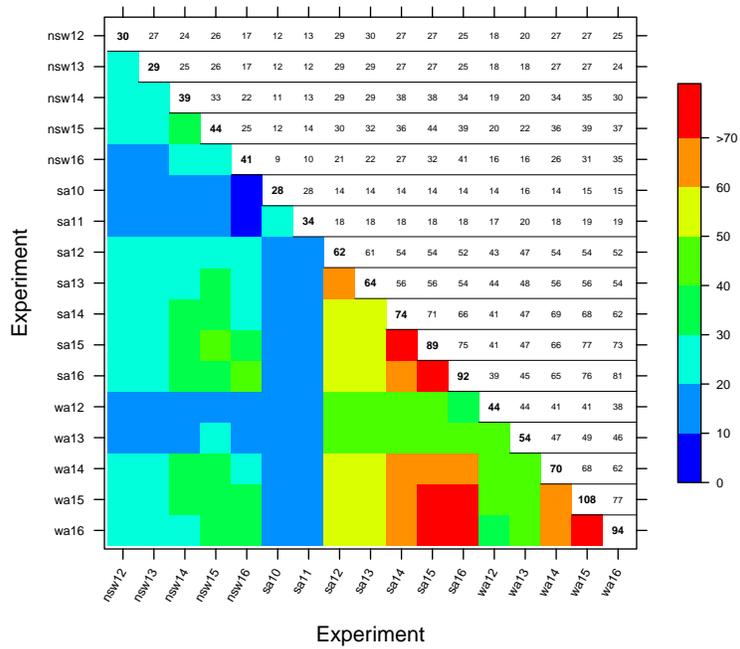


Figure 2: Connectivity of varieties for each of the 17 FEEs, where the number of varieties sown at each trial are on the main diagonal and those in common between pairs of trials are presented on the off-diagonal

2.3. Statistical Methods

The analysis of the MET dataset needs to consider the aims, the design function and nature of the experimental design, the longitudinal nature of the data and the high degree of unbalance in the data within and between FEEs. The aim of the experiment is to provide (the most) accurate FV predictions for varieties. This requires formulation of an efficient and valid analysis which incorporates all sources of variation and accounts for dependence through the longitudinal structure of the data. Our analysis is embedded within the framework of linear mixed models (LMMs) and this requires enumeration of the terms in the fixed and random model formulae, which will ensure the analysis is linked to classical or randomisation based analyses of comparative experiments (see for example Bailey (2008)).

The so-called design DT approach of Smith and Cullis (2018a) is adopted here in order to achieve this. We begin by consideration of a single FEE with one tag event and extend this to accommodate multiple tag events for a single FEE. Finally, this is extended to the multiple environment case which is the construct of the FEE MET dataset. In the following development, the inclusion of ancillary data (described in Section 2.2) is only addressed in Section 2.3.3.

2.3.1. LMM for FEE with one tag event

Implementation of the DT approach of Smith and Cullis (2018a) (for a single phase experiment) has seven steps. The first four require definitions of the treatments, the factors associated with the treatments, the plots (observational units, OU) and the set of factors that index the OU. Lastly, the treatment and plot structures are constructed using the design function and ensuring that the set of OU can be uniquely indexed from the set of plot factors. Furthermore, some plot factors will be involved in the design function while some may not and are merely required to uniquely identify the OU. Note that use of the `font` in the following indicates that the factors will be used as terms in the subsequent LMM as opposed to descriptions in the text.

Recall the hypothetical example of a simplified FEE in Figure 1. To extend this example and accommodate the sub-sampling of tillers within each plot, an additional factor is introduced that indexes the OU that were tagged (and phenotyped at maturity). The plot and treatment factors for a single FEE with one tag event are therefore defined as

Plot Factors: `Block`, `SBlock`, `FPlot`, `Tiller`

Treatment Factors: TOS, Variety

where the sub-blocks are denoted by **SBlock** and individual field plots by **FPlot** in order to distinguish the latter from the set of “plot” factors.

The treatments are combinations of the levels of **TOS** and **Variety**. The treatments are applied to **OU** such that (a) all tillers within an **FPlot** receive the same **Variety**, (b) each **FPlot** within an **SBlock** is allocated a different **Variety** but the same **TOS**, (c) each **SBlock** within a **Block** contains one replicate of each **Variety** and a single **TOS** and (d) each **Block** receives a different **TOS**. Each of the tillers are only tagged if they are at one of the two pre-specified **SOD**. Therefore, **SOD** is a factor that has not been applied/randomised but has effects, including interactions with the treatment factors that are of interest. In this setting, **SOD** reflects the inherent characteristics of each **Tiller** and is therefore referred to as an *anatomical variable* (Smith and Cullis, 2018a). This design function gives rise to the following plot and treatment structures

Plot Structure: Block/SBlock/FPlot/Tiller

Treatment Structure: TOS*Variety*SOD.

Following Bailey (2008) and Smith and Cullis (2018a), the treatment structure is initially fitted as fixed and the plot structure as random in the LMM that gives rise to the initial DT presented in Table 2.

We modify this initial DT in the following way. First recall the aim of the analysis is to determine the FV of each variety and thus rank their relative risk to frost. Therefore BLUPs are preferable over BLUEs as the probability of selection errors are minimised under the former approach (Bueno Filho and Gilmour, 2007). Thus the main effect of **Variety** and all interactions with **Variety** will be fitted as random effects. Furthermore, each **Block** is representative of a different **TOS** and hence **TOS** is aliased with **Block**. The convention in DT Smith and Cullis (2018a) is to replace occurrences of **TOS** with **TOS[Block]** and occurrences of **Block** with **Block[TOS]**. The implication for the fixed effects of **TOS[Block]** is that there can be no hypothesis testing for **TOS** since there is no residual term against which it can be tested. The implication for random effects involving **TOS[Block]** or **Block[TOS]** is that associated variance parameter estimates may be attributable to either **Block** or **TOS** and we cannot make the distinction. This simple re-labelling of terms therefore guards against invalid inference due to false replication. Therefore, **TOS[Block]** is retained in the fixed model formula as demonstrated in Smith and Cullis (2018a) with respect to the overall mean. Therefore the final model is given by Table 3, where “-” indicates that the term

Source	Model Term	Fixed/Random	Variance Model
Mean	1	F	
TOS	TOS	F	
Variety	Variety	F	
SOD	SOD	F	
TOS:Variety	TOS:Variety	F	
TOS:SOD	TOS:SOD	F	
Variety:SOD	Variety:SOD	F	
TOS:Variety:SOD	TOS:Variety:SOD	F	
<hr style="border-top: 1px dashed black;"/>			
Block	Block	R	$\sigma_B^2 \mathbf{I}$
Block:SBlock	Block:SBlock	R	$\sigma_{BSb}^2 \mathbf{I}$
Block:SBlock:FPlot	Block:SBlock:FPlot	R	$\sigma_{BSbFp}^2 \mathbf{I}$
Block:SBlock:FPlot:Tiller	units	R	$\sigma^2 \mathbf{I}$

Table 2: Initial DT for a single FEE with a single tag event. The **Source** terms are those that result from the expansion of the plot and treatment structures, the **Model Term** is what will be fitted in the linear mixed model (LMM) for the corresponding **Source**, the **Fixed/Random** column indicates whether that term will be in the fixed or random model formula in the LMM and if the term is random (“R”) the corresponding **Variance Model** is provided in the final column. The set of terms presented above the dashed line correspond to the treatment structure and the terms below the dashed line correspond to the plot structure

has already been fitted in the LMM. Note that the dimension of the identity matrix (\mathbf{I}) for each variance model specified in Table 3 is not presented for ease of presentation and is equivalent to the levels associated with the model term fitted.

2.3.2. LMM for the analysis of a FEE with multiple tag events

The baseline LMM in Table 3 is now extended to accommodate multiple tag events for a single FEE. The allocation of treatments to plot factors are as described in Section 2.3.1 and the multiple frost events that can lead to (possible) repeated measurements on the same FPlot are now addressed. Thus the factor **TagEvent** is introduced as a second anatomical variable which allows us to account for the longitudinal nature of the data.

Plot Factors: Block, SBlock, Plot, Tiller

Treatment Factors: TOS, Variety

Anatomical Variables: SOD, TagEvent

Following Brien and Demetrio (2009), **TagEvent** is considered to be an

Source	Model Term	Fixed/Random	Variance Model
Mean	1	F	
TOS	TOS[Block]	F	
Variety	Variety	R	$\sigma_V^2 \mathbf{I}$
SOD	SOD	F	
TOS:Variety	TOS[Block]:Variety	R	$\sigma_{BV}^2 \mathbf{I}$
TOS:SOD	TOS[Block]:SOD	F	
Variety:SOD	Variety:SOD	R	$\sigma_{VS}^2 \mathbf{I}$
TOS:Variety:SOD	TOS[Block]:Variety:SOD	R	$\sigma_{BVS}^2 \mathbf{I}$
Block	-	-	-
Block:SBlock	Block[TOS]:SBlock	R	$\sigma_{BSb}^2 \mathbf{I}$
Block:SBlock:FPlot	Block[TOS]:SBlock:FPlot	R	$\sigma_{BSbFp}^2 \mathbf{I}$
Block:SBlock:FPlot:Tiller	units	R	$\sigma^2 \mathbf{I}$

Table 3: Final DT for a single FEE with a single tag event. Note that “-” for any Model Term indicates that the term is already fitted in the LMM and TOS[Block]/Block[TOS] indicates the aliasing of TOS with Block. The dimension of the identity matrix (\mathbf{I}) for each variance model is not specified for ease of presentation but is equivalent to the levels associated with the model term fitted

integral component of the plot structure and hence it is crossed with the plot factors that are potentially involved in the repeated measurement process i.e. Block, SBlock and FPlot but not Tiller. We also allow for the interaction of TagEvent with some all of the treatment factors as they are again of interest. For the case of multiple tag events for a single FEE, the following set of plot and treatment structures arise

Plot Structure: ((Block/SBlock/FPlot)*TagEvent)/Tiller

Treatment Structure: TOS*Variety*SOD*TagEvent

Since the trait measured (FIS) requires the plants to be close to physiological maturity, i.e. grains fully set in order to distinguish between the viable and sterile florets, they must remain in the field until such time and thus exposure to multiple frost events is unavoidable. As a result, the individual frost events (TagEvent) in an experiment are not regarded as treatments but rather all tagged tillers for an event are associated in a cohort of similarity, as TagEvent is not representative of the effect of an individual frost event. Therefore, the term TagEvent and all interactions of remaining factors with TagEvent are fitted as random. All remaining terms are as previously described and fitted as in Table 3. The terms fitted in the LMM for a single

Source	Model Term	Fixed/Random	Variance Model
Mean	1	F	
TOS	TOS[Block]	F	
Variety	Variety	R	$\sigma_V^2 \mathbf{I}$
SOD	SOD	F	
TagEvent	TagEvent	R	$\sigma_T^2 \mathbf{I}$
TOS:Variety	TOS[Block]:Variety	R	$\sigma_{BV}^2 \mathbf{I}$
TOS:SOD	TOS[Block]:SOD	F	
TOS:TagEvent	TOS[Block]:TagEvent	R	$\sigma_{BT}^2 \mathbf{I}$
Variety:SOD	Variety:SOD	R	$\sigma_{VS}^2 \mathbf{I}$
Variety:TagEvent	Variety:TagEvent	R	$\sigma_{VT}^2 \mathbf{I}$
SOD:TagEvent	SOD:TagEvent	R	$\sigma_{ST}^2 \mathbf{I}$
TOS:Variety:SOD	TOS[Block]:Variety:SOD	R	$\sigma_{BVS}^2 \mathbf{I}$
TOS:Variety:TagEvent	TOS[Block]:Variety:TagEvent	R	$\sigma_{BVT}^2 \mathbf{I}$
TOS:SOD:TagEvent	TOS[Block]:SOD:TagEvent	R	$\sigma_{BST}^2 \mathbf{I}$
Variety:SOD:TagEvent	Variety:SOD:TagEvent	R	$\sigma_{VST}^2 \mathbf{I}$
TOS:Variety:SOD:TagEvent	TOS[Block]:Variety:SOD:TagEvent	R	$\sigma_{BVST}^2 \mathbf{I}$
Block	-	-	-
Block:SBlock	Block[TOS]:SBlock	R	$\sigma_{BSb}^2 \mathbf{I}$
Block:SBlock:FPlot	Block[TOS]:SBlock:FPlot	R	$\sigma_{BSbFp}^2 \mathbf{I}$
TagEvent	-	-	-
Block:TagEvent	-	-	-
Block:SBlock:TagEvent	Block[TOS]:SBlock:TagEvent	R	$\sigma_{BSbT}^2 \mathbf{I}$
Block:SBlock:FPlot:TagEvent	Block[TOS]:SBlock:FPlot:TagEvent	R	$\sigma_{BSbFpT}^2 \mathbf{I}$
Block:SBlock:FPlot:TagEvent:Tiller	units	R	$\sigma^2 \mathbf{I}$

Table 4: Final DT for a single FEE with multiple tag events. Note that “-” for any Model Term indicates that the term is already fitted in the LMM and TOS[Block]/Block[TOS] indicates the aliasing of TOS with Block. The dimension of the identity matrix (\mathbf{I}) for each variance model is not specified for ease of presentation but is equivalent to the levels associated with the model term fitted

FEE with more than one TagEvent are presented in Table 4, where those that have already been fitted are represented by a “-”.

2.3.3. LMM for the MET FEE dataset

The model for a single site FEE is now extended to accommodate the analysis of the MET dataset. In the following derivation the data is ordered as OU within each of the 17 FEEs. Other terms that may be potential sources of variation (and not genetically determined) but are not necessarily of interest (Diggle and Chetwynd, 2011) are also considered here. In this setting, these would include the people who have tagged the individual tillers (Tagger) and those who counted the sterile and total numbers of grains

in each head (**Counter**). These terms are referred to by Smith and Cullis (2018a) as extraneous variables. The set of plot and treatment factors and anatomical and extraneous variables for the MET dataset are

Plot Factors: **Expt**, **Block**, **SBlock**, **FPlot**, **Tiller**

Treatment Factors: **TOS**, **Variety**

Anatomical Variables: **Expt**, **SOD**, **TagEvent**

Extraneous Variables: **Tagger**, **Counter**

Expt is considered to be an anatomical variable as it is not only an inherent component of the plot structure but there is also particular interest in modelling and accounting for variety \times environment interaction (VEI). Formulation of the plot structure for the MET dataset is a simple extension of the plot structure derived in Section 2.3.2, where all model terms are now nested within **Expt** as each of their respective levels and interactions are unique to each experiment.

The treatment structure for the MET dataset however is quite complex. Firstly, there are varieties that are in common (at least 9 or more) between each of the 17 FEEs (see Figure 2) and tillers were only tagged at two pre-specified **SOD**. Therefore, both **Variety** and **SOD** are crossed with **Expt** as both the main effects and interactions of these terms are meaningful and consistent across the MET dataset. Each **TagEvent** however, is unique to (nested within) each experiment and therefore is not consistent across all experiments. Similarly, the levels of **TOS** are also nested within **Expt** as each **TOS** is not only aliased with **Block** but is also inconsistent across the 17 FEEs.

Finally, both **Tagger** and **Counter** in some instances are consistent across FEEs managed in the same state if a casual is re-hired the following year. However, this is only the case for very few people and there is not enough information to know with certainty if it is the same person nor that the protocol remained unchanged across years. Therefore, **Tagger** and **Counter** are assumed to be unique to each of the 17 FEEs and therefore the levels associated with each are nested within **Expt**. The resulting plot and treatment structures for the MET dataset are given by

Plot Structure: **Expt**/((**Block**/**SBlock**/**FPlot**)***TagEvent**)/**Tiller**

Treatment Structure: (**Expt**/(**TOS*****TagEvent**))***Variety*****SOD**

where **Expt:Tagger** and **Expt:Counter** are also fitted as random model terms to capture non-genetic variation in the dataset. The initial DT for the MET FEE dataset by expansion of the above structures is given by Table 5.

There are many characteristics of Table 5 that require explanation and defining. Firstly, for any terms where the prefix **Expt|a.i**, $i=1\dots 23$ is

Source	Model Term	Fixed/Random	Variance Model
Mean	1	F	
Expt	Expt	F	
Variety	Variety	F	
SOD	SOD	F	
Expt:Variety	Expt:Variety	F	
Expt:TOS	(Expt a_1):TOS	F	
Expt:SOD	(Expt a_2):SOD	F	
Expt:TagEvent	(Expt a_3):TagEvent	F	
Expt:TOS:Variety	(Expt a_4):TOS:Variety	F	
Expt:TOS:SOD	(Expt a_5):TOS:SOD	F	
Expt:TOS:TagEvent	(Expt a_6):TOS:TagEvent	F	
Expt:Variety:SOD	(Expt a_7):Variety:SOD	F	
Expt:Variety:TagEvent	(Expt a_8):Variety:TagEvent	F	
Expt:SOD:TagEvent	(Expt a_9):SOD:TagEvent	F	
Expt:TOS:Variety:SOD	(Expt a_10):TOS:Variety:SOD	F	
Expt:TOS:Variety:TagEvent	(Expt a_11):TOS:Variety:TagEvent	F	
Expt:TOS:SOD:TagEvent	(Expt a_12):TOS:SOD:TagEvent	F	
Expt:Variety:SOD:TagEvent	(Expt a_13):Variety:SOD:TagEvent	F	
Expt:TOS:Variety:SOD:TagEvent	(Expt a_14):TOS:Variety:SOD:TagEvent	F	
Expt:Tagger	(Expt a_15):Tagger	R	$\oplus(\mathbf{G}_{1_i})$
Expt:Counter	(Expt a_16):Counter	R	$\oplus(\mathbf{G}_{2_i})$
Expt	Expt	R	$\oplus(\mathbf{G}_{3_i})$
Expt:Block	(Expt a_17)Block	R	$\oplus(\mathbf{G}_{4_i})$
Expt:Block:SBlock	(Expt a_18):Block:SBlock	R	$\oplus(\mathbf{G}_{5_i})$
Expt:Block:SBlock:FPlot	(Expt a_19):Block:SBlock:FPlot	R	$\oplus(\mathbf{G}_{6_i})$
Expt:TagEvent	(Expt a_20):TagEvent	R	$\oplus(\mathbf{G}_{7_i})$
Expt:Block:TagEvent	(Expt a_21):Block:TagEvent	R	$\oplus(\mathbf{G}_{8_i})$
Expt:Block:SBlock:TagEvent	(Expt a_22):Block:SBlock:TagEvent	R	$\oplus(\mathbf{G}_{9_i})$
Expt:Block:SBlock:FPlot:TagEvent	(Expt a_23):Block:SBlock:FPlot:TagEvent	R	$\oplus(\mathbf{G}_{10_i})$
Expt:Block:SBlock:FPlot:TagEvent:Tiller	units	R	$\oplus(\mathbf{R}_i)$

Table 5: Initial DT for the FEE MET dataset. The expansion of the treatment structure is presented first, then the extraneous variables after the first dashed line then expansion of the plot structure after the final dashed line. Note that the \oplus operator indicates the overall variance matrix is block diagonal with each block corresponding to a separate experiment

fitted, the levels contained in \mathbf{a}_i are the names of those experiments for which that term must be fitted and is hereafter a conditioned factor. For example, $(\text{Expt}|\mathbf{a}_2):\text{SOD}$ will only be fitted for those experiments that are listed in \mathbf{a}_2 , which are those experiments with data measured at both levels of SOD and that are not aliased with any other term.

The \oplus operator indicates that the overall variance matrix for that fitted term is block diagonal, with each block corresponding to a separate experiment, i . For each \mathbf{G}_{k_i} , $k = 1, 2, \dots, 10$ the variance matrix is given by $\mathbf{G}_{k_i} = \sigma_{k_i}^2 \mathbf{I}_{b_{k_i}}$. In other words, the value of σ_k^2 is calculated for each experiment i contained in \mathbf{a}_k , $k = 15, \dots, 23$ and for ease of presentation was not explicit in Table 5. Hence, for $(\text{Expt}|\mathbf{a}_{15}):\text{Tagger}$ for example, the corresponding variance model is $\oplus(\sigma_{15_i}^2 \mathbf{I}_{b_{15_i}})$, where i indexes each of the ex-

periments contained in \mathbf{a}_{15} and b_{15_i} is the dimension of \mathbf{I} that is equivalent to the total number of levels associated with **Tagger** for experiment i . Furthermore, the variance for the OU, $\oplus(\mathbf{R}_i)$ is a block diagonal matrix, where $\mathbf{R}_i = \sigma_i^2 \mathbf{I}_{b_i}$ for $i = 1, 2, \dots, 17$ and the dimension b_i is equal to the number of OU for experiment i .

Recall from Section 2.3.1 that the aliasing of **TOS** with **Block** is denoted by **TOS[Block]** (and as **Block[TOS]** in the expansion of the plot structure). Furthermore, the term **Variety** and all associated interactions of **Variety** with other factors are fitted in the random model, given the aim of the analysis is selection of less frost susceptible varieties. From Section 2.3.2, it was deduced the term **TagEvent** and all interactions of remaining factors with **TagEvent** are fitted as random, as **TagEvent** is not representative of each individual frost event within an **Expt**. Finally, the main effect of **Expt** is present in both the fixed and random model formula and therefore we must decide which is more appropriate. Since we wish to centre the predicted FV of each variety within environment on the mean FV for that environment, **Expt** is retained in the fixed model. Furthermore, the variance of each experiment and the correlations between experiments can be estimated within this framework. The intermediate LMM for the MET FEE dataset is given by Table 6.

The current model in Table 6 includes a main effect for varieties and an interaction term between varieties and experiments and assumes a simple variance component structure. This is commensurate with a so-called compound symmetric variance structure which assumes equal variance at each experiment and equal covariance between pairs of experiments. A more general variance structure can be obtained by excluding the **Variety** main effect from the model so that the term **Expt:Variety** no longer represents the interaction effects but rather the variety effects for each level of experiment, that is, the nested effects of varieties within experiment. The variance model for **Expt:Variety** is chosen to be a factor analytic (FA) model after Smith et al. (2015) and has been shown to provide a parsimonious model for VEI and also provide accurate predictions for the variety effects (Smith et al., 2015).

Using the notation of Smith et al. (2015), the FA model of order k for the

Source	Model Term	Fixed/Random	Variance Model
Mean	1	F	
Expt	Expt	F	
Variety	Variety	R	$\sigma_V^2 \mathbf{I}$
SOD	SOD	F	
Expt:Variety	Expt:Variety	R	$\sigma_{EV}^2 \mathbf{I}$
Expt:TOS	(Expt a_1):TOS[Block]	F	
Expt:SOD	(Expt a_2):SOD	F	
Expt:TagEvent	(Expt a_3):TagEvent	R	$\oplus(\mathbf{G}_{1_i})$
Expt:TOS:Variety	(Expt a_4):TOS[Block]:Variety	R	$\oplus(\mathbf{G}_{2_i})$
Expt:TOS:SOD	(Expt a_5):TOS[Block]:SOD	F	
Expt:TOS:TagEvent	(Expt a_6):TOS[Block]:TagEvent	R	$\oplus(\mathbf{G}_{3_i})$
Expt:Variety:SOD	(Expt a_7):Variety:SOD	R	$\oplus(\mathbf{G}_{4_i})$
Expt:Variety:TagEvent	(Expt a_8):Variety:TagEvent	R	$\oplus(\mathbf{G}_{5_i})$
Expt:SOD:TagEvent	(Expt a_9):SOD:TagEvent	R	$\oplus(\mathbf{G}_{6_i})$
Expt:TOS:Variety:SOD	(Expt a_10):TOS[Block]:Variety:SOD	R	$\oplus(\mathbf{G}_{7_i})$
Expt:TOS:Variety:TagEvent	(Expt a_11):TOS[Block]:Variety:TagEvent	R	$\oplus(\mathbf{G}_{8_i})$
Expt:TOS:SOD:TagEvent	(Expt a_12):TOS[Block]:SOD:TagEvent	R	$\oplus(\mathbf{G}_{9_i})$
Expt:Variety:SOD:TagEvent	(Expt a_13):Variety:SOD:TagEvent	R	$\oplus(\mathbf{G}_{10_i})$
Expt:TOS:Variety:SOD:TagEvent	(Expt a_14):TOS[Block]:Variety:SOD:TagEvent	R	$\oplus(\mathbf{G}_{11_i})$
Expt:Tagger	(Expt a_15):Tagger	R	$\oplus(\mathbf{G}_{12_i})$
Expt:Counter	(Expt a_16):Counter	R	$\oplus(\mathbf{G}_{13_i})$
Expt	-	-	-
Expt:Block	-	-	-
Expt:Block:SBlock	(Expt a_17):Block[TOS]:SBlock	R	$\oplus(\mathbf{G}_{14_i})$
Expt:Block:SBlock:FPlot	(Expt a_18):Block[TOS]:SBlock:FPlot	R	$\oplus(\mathbf{G}_{15_i})$
Expt:TagEvent	-	-	-
Expt:Block:TagEvent	-	-	-
Expt:Block:SBlock:TagEvent	(Expt a_19):Block[TOS]:SBlock:TagEvent	R	$\oplus(\mathbf{G}_{16_i})$
Expt:Block:SBlock:FPlot:TagEvent	(Expt a_20):Block[TOS]:SBlock:FPlot:TagEvent	R	$\oplus(\mathbf{G}_{17_i})$
Expt:Block:SBlock:FPlot:TagEvent:Tiller	units	R	$\oplus(\mathbf{R}_i)$

Table 6: Intermediate DT for the FEE MET dataset. Note that “-” indicates that the term has already been fitted prior and that TOS[Block]/Block[TOS] represents the aliasing of TOS with Block

vector \mathbf{u} of variety effects within experiment can be written as

$$\mathbf{u} = (\lambda_1 \otimes \mathbf{I}_m) \mathbf{f}_1 + (\lambda_2 \otimes \mathbf{I}_m) \mathbf{f}_2 + \dots + (\lambda_k \otimes \mathbf{I}_m) \mathbf{f}_k + \delta \quad (1)$$

$$\mathbf{u} = (\mathbf{\Lambda} \otimes \mathbf{I}_m) \mathbf{f} + \delta$$

where $\mathbf{\Lambda}$ is the $t \times k$ matrix of loadings, \mathbf{f} is the mk vector of scores, δ is the mt vector of genetic regression residuals, $t = 17$ is the number of experiments and $m = 167$ is the number of varieties. The variance matrices for \mathbf{f} and δ are $\text{Var}(\mathbf{f}) = \mathbf{I}_{mk}$ and $\text{Var}(\delta) = \mathbf{\Psi} \otimes \mathbf{I}_m$ respectively, where $\mathbf{\Psi}$ is a $t \times t$ diagonal matrix with a variance (called a specific variance) for each experiment. Hence,

$$\text{Var}(\mathbf{u}) = (\mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{\Psi}) \otimes \mathbf{I}_m$$

where $\mathbf{G}_e = (\mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Psi})$ is defined to be the between experiment (aka environment) genetic variance matrix.

Finally, given the sparsity of the data collected (see Table 1) it is not possible to fit the full saturated LMM outlined in Table 6. The list of terms present in Table 4 represents the so-called “maximal single site” mixed model for each of the FEEs. Table 7 presents a summary of the number of levels associated with each term in the maximal single site mixed model for each experiment. The numbers that are **bold** indicate that the corresponding term can and was fitted in the LMM for that experiment. Of particular interest is the lack of information on **Variety:SOD** for most experiments and we return to this point in the next section. All remaining terms in the table that are not **bold** were excluded for an experiment due to either structural aliasing or implicit aliasing with another term(s). This was determined by using an algorithm designed to detect singularities in the average information matrix to identify aliasing of fixed and random effects that were not already explicit but due to the sparsity in the data.

	nsw12	nsw13	nsw14	nsw15	nsw16	sa10	sa11	sa12	sa13	sa14	sa15	sa16	wa12	wa13	wa14	wa15	wa16
Mean	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TOS[Block]	4	3	3	3	2	2	2	6	3	3	6	5	3	6	8	8	7
Variety	30	29	39	44	41	28	34	62	64	74	89	92	44	54	70	108	94
SOD	2	2	2	2	1	1	1	2	2	2	2	2	2	2	2	2	2
TagEvent	17	3	3	4	2	2	2	4	4	3	2	4	3	5	7	6	3
TOS[Block]:Variety	98	61	92	77	54	38	57	245	126	169	339	390	56	140	280	431	257
TOS[Block]:SOD	8	3	4	4	2	2	2	11	4	5	6	8	4	10	10	13	12
TOS[Block]:TagEvent	32	3	4	6	2	3	2	16	7	4	8	11	7	17	27	28	16
Variety:SOD	60	40	75	73	41	28	34	121	103	124	151	183	55	106	104	178	157
Variety:TagEvent	194	61	72	89	54	52	57	204	198	118	171	350	55	164	312	573	237
SOD:TagEvent	19	3	4	5	2	2	2	5	4	4	3	6	3	6	7	8	5
TOS[Block]:Variety:SOD	177	61	93	105	54	38	57	395	165	218	339	536	56	173	291	499	261
TOS[Block]:Variety:TagEvent	239	61	97	114	54	62	57	435	239	177	382	570	56	176	315	593	280
TOS[Block]:SOD:TagEvent	33	3	5	7	2	3	2	17	7	6	8	13	7	17	27	31	20
Variety:SOD:TagEvent	198	61	97	89	54	52	57	260	198	167	231	437	55	170	312	574	279
TOS[Block]:Variety:SOD:TagEvent	239	61	98	114	54	62	57	435	239	226	382	613	56	176	315	594	281
Block[TOS]:SBlock	8	6	6	6	4	4	4	12	6	6	12	10	6	12	16	15	14
Block[TOS]:SBlock:FPlot	184	99	148	135	79	62	103	384	225	291	566	688	107	250	534	812	500
Block[TOS]:SBlock:TagEvent	51	6	8	12	4	6	4	31	14	8	16	21	14	32	52	55	32
Block[TOS]:SBlock:FPlot:TagEvent	313	99	154	170	79	95	103	602	380	302	605	908	107	298	588	1105	544
Tagger	0	0	2	5	2	0	0	0	0	0	0	9	0	6	4	9	9
Counter	0	1	6	1	3	0	0	0	0	0	4	6	0	5	9	9	13

Table 7: Number of levels associated with each term in the LMM for each of the 17 FEE. Also note that those in bold were fitted for each of the respective experiments in the MET analysis

Following Table 7, the final LMM for the MET dataset is presented in Table 8. Note that a “*” indicates that the term was not fitted for any of the 17 FEEs due to implicit aliasing as a result of the sparsity of the data. Furthermore, “**” was used to indicate the main effect of **Variety** was not fitted due to the FA model fitted to the **Expt:Variety** effects.

2.4. Model fitting and tools for interpretation

All models in this paper were fitted using the **ASReml-R** package (Butler et al., 2015) in the R statistical computing environment (R Core Team, 2015). The variance parameters are estimated using residual maximum likelihood (REML). In the FA model, the variance parameters are the loadings and specific variances, where the REML estimates of these are denoted by $\hat{\lambda}_{rj}$ and $\hat{\Psi}_j$ ($r = 1, \dots, k; j = 1, \dots, 17$). Note when $k > 1$, $\mathbf{\Lambda}$ is not unique so estimation requires imposition of constraints. The approach used in **ASReml-R** fixes the $k(k-1)/2$ elements in the upper triangle of $\hat{\mathbf{\Lambda}}$ to zero. Once the REML estimate of $\mathbf{\Lambda}$ is available, the matrix $\hat{\mathbf{\Lambda}}$ is rotated to the principal components solution (see Smith et al. (2001) for details) and henceforth assume $\hat{\mathbf{\Lambda}}$ relates to the rotated solution.

Given the set of REML estimates of the variance parameters, the empirical best linear unbiased predictors (EBLUPs) are obtained of the random effects for \mathbf{f} and δ . These are denoted by $\tilde{\mathbf{f}}_{r_i}$ and $\tilde{\delta}_{ij}$ ($r = 1, \dots, k; i = 1, \dots, 167; j = 1, \dots, 17$). The model fitting process involved fitting a sequence of FA models with increasing order to the **Expt:Variety** term. Note that all other terms remain in the LMM during the FA modelling.

Subject to the parameter estimate, the model in equation (1) resembles a multiple regression in which there are k environmental covariates ($\hat{\lambda}_1, \dots, \hat{\lambda}_k$) and for each covariate, each variety has a different slope that is given by the corresponding predicted score ($\tilde{\mathbf{f}}$). The final term in equation (1) represents the residual or lack of fit of the regression model. The choice for k is determined by the data and may be based on likelihood ratio tests or information criteria. A pragmatic approach is used here as advocated by Smith et al. (2015) and k is chosen such that at least 80% of genetic variance is accounted for by the regression part of the model. The percentage variance accounted for (%VAF) for experiment j is defined as

$$v_j = 100 \sum_{r=1}^k \hat{\lambda}_{rj}^2 / (\sum_{r=1}^k \hat{\lambda}_{rj}^2 + \hat{\Psi}_j)$$

Source	Model Term	Fixed/Random	Variance Model
Mean	1	F	
Expt	Expt	F	
Variety	**		
SOD	SOD	F	
Expt:Variety	Expt:Variety	R	$(\Lambda\Lambda^\top + \Psi) \otimes \mathbf{I}$
Expt:TOS	(Expt a_1):TOS[Block]	F	
Expt:SOD	(Expt a_2):SOD	F	
Expt:TagEvent	(Expt a_3):TagEvent	R	$\oplus(\mathbf{G}_{1_i})$
Expt:TOS:Variety	(Expt a_4):TOS[Block]:Variety	R	$\oplus(\mathbf{G}_{2_i})$
Expt:TOS:SOD	(Expt a_5):TOS[Block]:SOD	F	
Expt:TOS:TagEvent	(Expt a_6):TOS[Block]:TagEvent	R	$\oplus(\mathbf{G}_{3_i})$
Expt:Variety:SOD	(Expt a_7):Variety:SOD	R	$\oplus(\mathbf{G}_{4_i})$
Expt:Variety:TagEvent	(Expt a_8):Variety:TagEvent	R	$\oplus(\mathbf{G}_{5_i})$
Expt:SOD:TagEvent	*		
Expt:TOS:Variety:SOD	(Expt a_9):TOS[Block]:Variety:SOD	R	$\oplus(\mathbf{G}_{6_i})$
Expt:TOS:Variety:TagEvent	(Expt a_10):TOS[Block]:Variety:TagEvent	R	$\oplus(\mathbf{G}_{7_i})$
Expt:TOS:SOD:TagEvent	*		
Expt:Variety:SOD:TagEvent	*		
Expt:TOS:Variety:SOD:TagEvent	*		
Expt:Tagger	(Expt a_11):Tagger	R	$\oplus(\mathbf{G}_{8_i})$
Expt:Counter	(Expt a_12):Counter	R	$\oplus(\mathbf{G}_{9_i})$
Expt	-	-	-
Expt:Block	-	-	-
Expt:Block:SBlock	(Expt a_13):Block[TOS]:SBlock	R	$\oplus(\mathbf{G}_{10_i})$
Expt:Block:SBlock:FPlot	(Expt a_14):Block[TOS]:SBlock:FPlot	R	$\oplus(\mathbf{G}_{11_i})$
Expt:TagEvent	-	-	-
Expt:Block:TagEvent	-	-	-
Expt:Block:SBlock:TagEvent	(Expt a_15):Block[TOS]:SBlock:TagEvent	R	$\oplus(\mathbf{G}_{12_i})$
Expt:Block:SBlock:FPlot:TagEvent	(Expt a_16):Block[TOS]:SBlock:FPlot:TagEvent	R	$\oplus(\mathbf{G}_{13_i})$
Expt:Block:SBlock:FPlot:TagEvent:Tiller	units	R	$\oplus(\mathbf{R}_i)$

Table 8: Final DT for the FEE MET dataset. Note that “-” indicates that the term has already been fitted prior and “*” means that the term could not be fitted for any of the 17 FEEs due to the sparsity in the data. The main effect of **Variety** (“**”) is also not fitted because an FA model is fitted to the **Expt:Variety** term. The treatment factors, extraneous variables and plot factors are separated by the dashed lines

and across all experiments as

$$\bar{v} = 100tr(\hat{\Lambda}\hat{\Lambda}^\top)/tr(\hat{\Lambda}\hat{\Lambda}^\top + \Psi). \quad (2)$$

3. Results

The FIS data required transformation in order to more appropriately satisfy the Gaussian residual assumptions of the LMM. The transformation for the phenotypic trait of interest was

$$y = \sqrt[3]{\frac{\text{No.sterileseeds} + 1.5}{\text{TotalNo.Seeds}}}$$

Initially, a baseline diagonal variance model (**Diag**) was fitted for the **Expt:Variety** effects, which is equivalent to analysing each of the 17 FEEs independently of one another. It was determined from this preliminary analysis that there was no genetic variance at the **sa14** site (due to severe frost damage of all varieties) and so this trial was dropped from the subsequent MET analyses. The MET analysis of the remaining 16 FEEs consisted of fitting a series of **FA** models of increasing order k to the **Expt:Variety** effects. The REML log-likelihood test statistic ($-2\Delta\text{REML.loglik}$), difference in number of $V \times E$ parameters ($\Delta V \times E$) for each model compared to its predecessor, p -value associated with the REML log-likelihood ratio test (REML-LRT) for each model compared to its predecessor and %VAF (i.e. \bar{v} from equation (2)) by each model are presented in Table 9. It is clear that the **FA1** model is superior to the **Diag** model ($-2\Delta\text{REML.loglik} = 427.51$), where the REML-LRT returned a p -value of $\ll 0.001$. The increase in REML.loglik between the **FA2** and **FA1** models was also shown to be statistically significant (p -value < 0.001 , $-2\Delta\text{REML.loglik} = 48.72$). The higher order **FA3** model was not found to be statistically significant compared to the **FA2** model (p -value = 0.22), however the %VAF by the **FA2** model was not sufficiently high enough (%VAF = 77%). Therefore, the **FA3** model was the final model fitted and accounted for 82% of genetic variance for the MET dataset. We will therefore proceed with the results from the **FA3** model for the FV predictions of the unique breeding lines grown across the 16 FEEs.

Model	$-2\Delta\text{REML.loglik}$	$\Delta V \times E$ param.	pval	%VAF
Diag	-	-		
FA1	-427.51	16	< 0.001	67
FA2	-48.72	15	< 0.001	77
FA3	-17.68	14	0.222	82

Table 9: Summary of models fitted to **Expt:Variety** effects (**Diag**, **FA1**, **FA2** and **FA3**) including the REML log-likelihood test statistic ($-2\Delta\text{REML.loglik}$), increase in number of $V \times E$ parameters fitted, p -value associated with the REML log-likelihood ratio test of the current model to the one above and %VAF

Table 10 presents the REML estimates of variance parameters on an in-

dividual experiment basis, expressed as a percentage of the total variance for that experiment. The genetic variance of each experiment (diagonal elements of $(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top + \hat{\mathbf{\Psi}})$) accounted for between 1.73% (**wa16**) to 40.45% (**sa10**) of the total variance of each experiment, while the largest sources of non-genetic variation was due to the residual (24.10% to 80.28%) and **TagEvent** (0% to 60.27%). The variance associated with **TOS[Block]:Variety** ranged from 0% to 17.01% (**wa12**).

There was also variance associated with the interaction of **Variety** and **TagEvent** detected for some of the experiments. In other words, at any given **TagEvent**, the FIS score for a **Variety** will not always be higher than another **Variety** or vice versa for the same **SOD**. Finally, the data was too sparse in order for the term **SOD:TagEvent** to be fitted and therefore it is not possible to determine if there is an interaction between the susceptible **SOD** (i.e. flowering and booting) and frost events.

	nsw12	nsw13	nsw14	nsw15	nsw16	sa10	sa11	sa12	sa13	sa15	sa16	wa12	wa13	wa14	wa15	wa16
Variety	5.56	4.44	12.96	8.61	22.32	40.45	2.90	13.67	17.09	8.86	15.09	28.14	15.73	2.30	5.78	1.73
TagEvent	0.32		0.00	60.27		0.76		12.19	7.39	0.00	3.85	0.00	0.00	46.36	16.13	20.00
TOS[Block]:Variety	2.70	4.32	2.94	0.00	8.38	9.67	0.00	2.44	0.00	0.00	0.20	17.01	6.22	1.99	4.11	5.06
TOS[Block]:TagEvent	1.16							0.30	0.00	0.00	1.47			0.00	2.22	5.54
Variety:SOD	0.00			0.00				0.00	0.00		0.16		0.00	0.21	0.00	3.46
Variety:TagEvent	0.00		0.00	3.03		4.01		0.31	2.16	0.00	0.41		0.00	3.20	2.77	0.00
TOS[Block]:Variety:SOD	0.00							0.00			4.84					
TOS[Block]:Variety:TagEvent								0.36	0.00	4.06	0.37					
Block[TOS]:SBlock	2.75	0.00	0.00	0.00	0.00	0.06	1.01	2.94	0.15	0.00	3.70	0.00	0.34	0.00	0.76	0.00
Block[TOS]:SBlock:FPlot	6.53	10.96	11.05	0.85	4.58	0.00	18.17	9.72	2.28	6.08	7.04	13.36	2.11	2.48	4.70	0.00
Block[TOS]:SBlock:TagEvent	0.00			0.00		0.00		0.16	0.00	25.11	0.00		0.00	0.00	0.00	0.17
Block[TOS]:SBlock:FPlot:TagEvent	4.79			0.00		12.92		6.37	6.10		3.66		3.79	2.02	1.57	18.70
Tagger			0.00	0.03	7.33						0.68		0.00	0.19	0.65	0.29
Counter			1.43		0.00					31.78	24.26		6.14	8.12	4.96	9.94
units	76.20	80.28	71.62	27.21	57.39	32.13	77.92	51.53	64.83	24.10	34.27	41.49	65.67	33.12	56.35	35.10

Table 10: Percent variance associated with random effects in the LMM for each of the 16 FEEs

Figure 3 presents the REML estimate of the between environment correlations and the genetic variance of each experiment for the fitted FA3 model. The between environment correlations are presented on the upper triangle of the figure and as a heatmap on the lower triangle. The genetic variance of each experiment is on the main diagonal and scaled by $\times 10^3$. It is worth noting the negative correlation of **sa10** and **wa16** with **nsw12** and **nsw13**, indicative of real and substantive VEI. The correlations between experiments, with the exception of **sa10** and **nsw13** with all remaining trials, is at least 0.4.

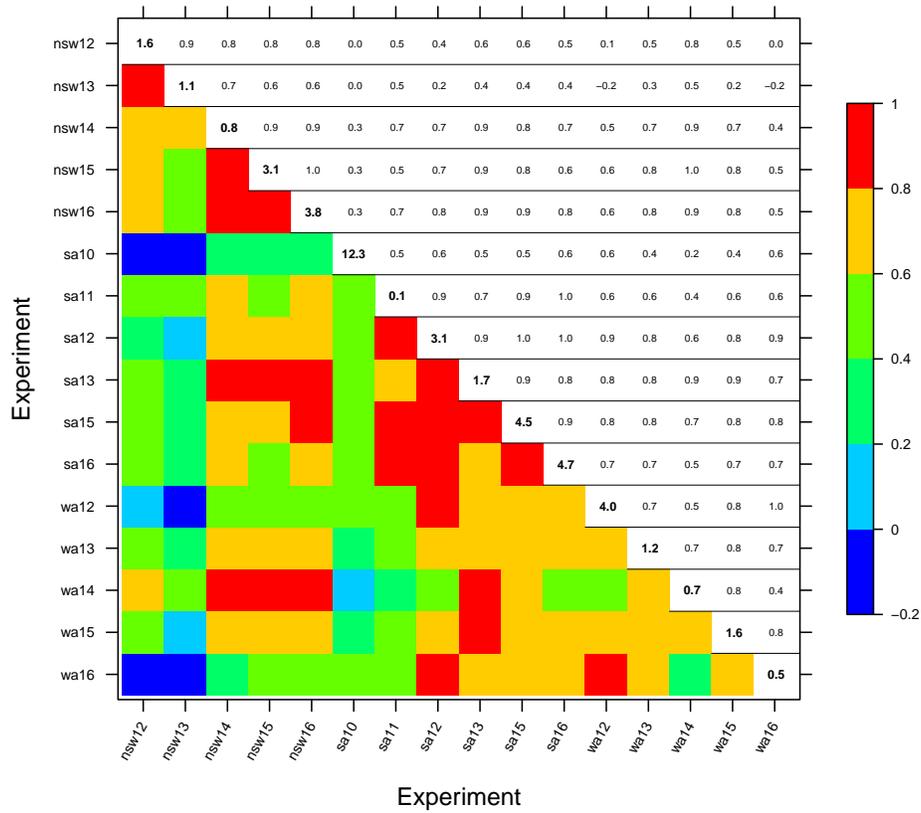


Figure 3: REML estimates of the between experiment genetic correlations presented on the off-diagonal and the genetic variance of each experiment presented on the diagonal scaled by $\times 10^3$

3.1. Selection of Low Frost Risk Varieties

The primary aim of the FEE MET analysis is to inform local growers of the relative frost susceptibility of commercial and advanced wheat varieties in a range of environments. The PV-PLUS graphs (hereafter referred to as FV-PLUS) introduced by Smith et al. (2015) are an appropriate tool for summarising variety performance at a level where growers can exploit the environmental conditions represented by individual experiments and apply that information to their own farming systems. Given the nature of the analysis, it will also be possible to provide information that is of use to the breeders and researchers. Varieties that consistently demonstrate lower frost risk (high adaptation) or low risk in specific environments only (specific adaptation) are also of interest. The genetic predisposition of these highly and/or specifically adapted varieties can then be explored with the aim of introducing reduced susceptibility to frost in future breeding.

With this in mind, it would be sensible to initially use summary tools such as those presented in Smith and Cullis (2018b), which are based on the regression interpretation of the FA model as described in Section 2.4. It is noted that the rotated environment loadings for the first factor of the fitted FA3 model are all positive. This indicates that the VEI accounted for by the first factor is driven by scale. In other words, the rankings of each variety in reference to frost risk do not change across each environment (for the first factor) however the magnitude of the differences between each variety may change thus leading to scale VEI. It is therefore meaningful to then determine the so-called overall performance (OP) for each of the varieties using the fitted values in the regression for the first rotated factor, which are given by $(\hat{\mathbf{A}}_1 \otimes \mathbf{I}_m)\tilde{\mathbf{f}}_1$ (see equation (1)). Thus the OP for variety i is given by $\bar{\mathbf{A}}_1\tilde{\mathbf{f}}_1$, where $\bar{\mathbf{A}}_1 = \frac{1}{16} \sum_{j=1}^{16} \hat{\lambda}_{1,j}$ (Smith and Cullis, 2018b).

The growers, breeders and researchers are also interested in measures of stability of FV across experiments. Smith and Cullis (2018b) provide a measure for FA models that is based on factors 2, \dots , k (i.e. excluding the first). It is based on the sum of squares of fitted values for the second and higher order regression terms and is referred to as Root Mean Square Deviation (RMSD). Both OP and RMSD are expressed in the units of measurement of the trait being analysed, in this case the transformed FIS value. Figure 4 illustrates the OP vs RMSD for the 167 varieties that were sown in the FEE MET dataset on the scale of the (transformed) data. Those varieties with a low OP and RMSD are considered to be superior in the sense that they have low frost risk (concurrent with a low FV) and are relatively stable across all

16 environments. Therefore, those varieties in the bottom left corner should be those of interest to particularly the breeders for low susceptibility to frost damage.

The points in Figure 4 have been shaded lighter if that variety was only grown in four or fewer environments of the possible 16 in the MET dataset. Although this is not an exact measure of the accuracy of OP, it cautions the user on interpretation of the OP for this set of varieties that were only grown in a minimal subset of the environments in the MET dataset. Figure 4 shows that the most frost susceptible varieties are the IMI herbicide resistant wheat varieties, namely *Impose CL Plus* and *Impress CL Plus* (note that the latter is present in only four environments, first grown in WA and SA in 2015). The most stable and low risk varieties were *AUS30323* and *Suntop* and the most unstable variety sampled in all 16 environments was *Ventura*.

Varietal stability as illustrated in Figure 4 is estimated across the full set of 16 environments in the MET dataset. It is also pertinent that growers are provided with accurate and efficient estimation of performance that is also relevant to their local area. The complexity of the VEI necessitates systems such as these to guide optimal selection decisions. For example, a variety that appears unstable may perform favourably and be specifically adapted to a particular growing region. The FV-PLUS graphs for the set of varieties that are commonly grown in NSW, SA and WA are presented, where *Suntop* was the most stable.

In the following FV-PLUS graphs, the predicted FV is shown as a shape specific to each variety and the corresponding standard errors of each prediction as error bars. If a variety was not sown in a particular year for that state, its predicted FV is not presented. Figure 5, Figure 6 and Figure 7 present the predicted FV of the regression part (equation (1)) of the FA3 model for each experiment in NSW, SA and WA respectively. The number of tag events and mean FIS for each experiment is also presented to provide the user with more detailed information about the environment that is represented, i.e. frequency and severity (mean FIS) of the frost events. This tool is available online (<https://www.nvtonline.com.au/frost/>) to all growers in an interactive format, where the user can manage the varieties, region, year and crop type (wheat or barley) presented.

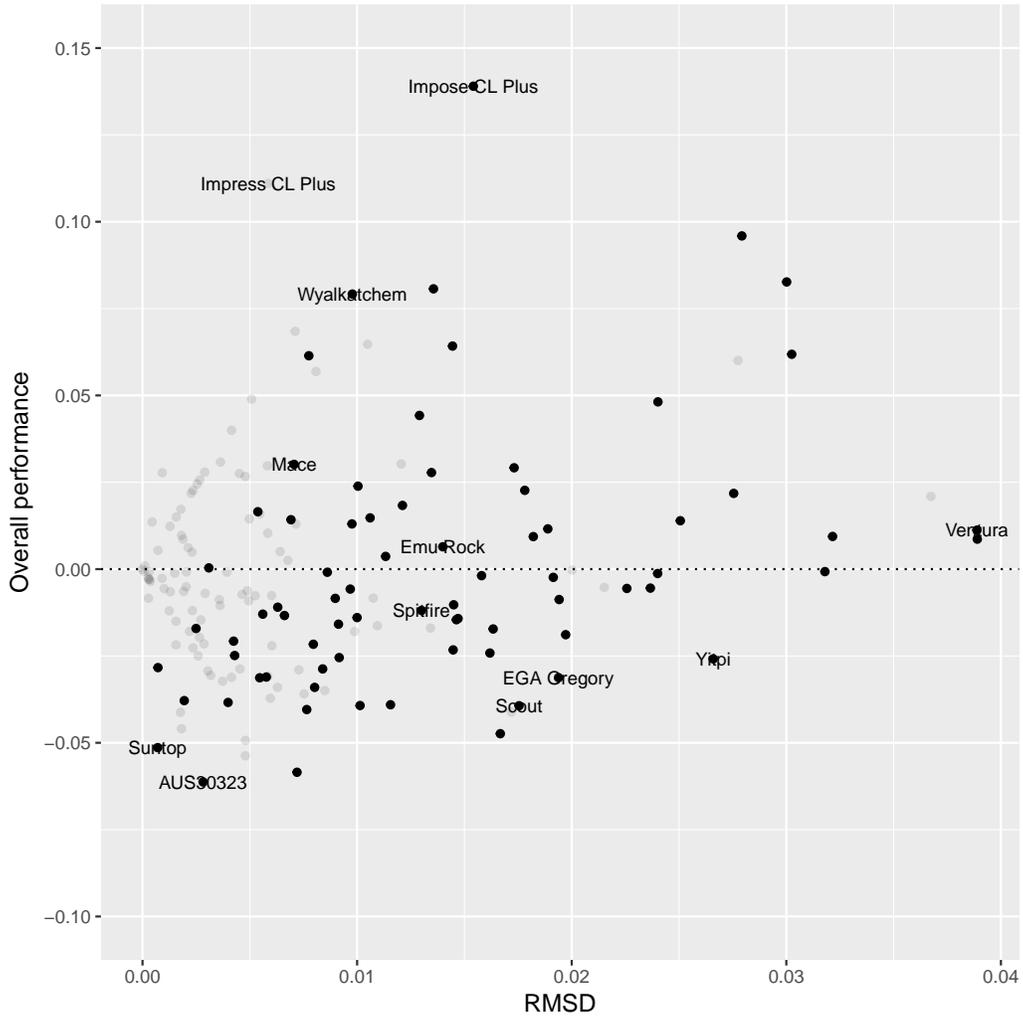


Figure 4: OP vs RMSD for FV of all varieties sown in the FEE MET dataset. Varieties in the bottom left corner are considered to be the least susceptible to frost and stable performing across all environments. Those varieties that were present in less than five experiments appear more transparent than the remaining varieties present in five experiments or more

NSW Region

(the dashed line represents an average variety for that experiment)

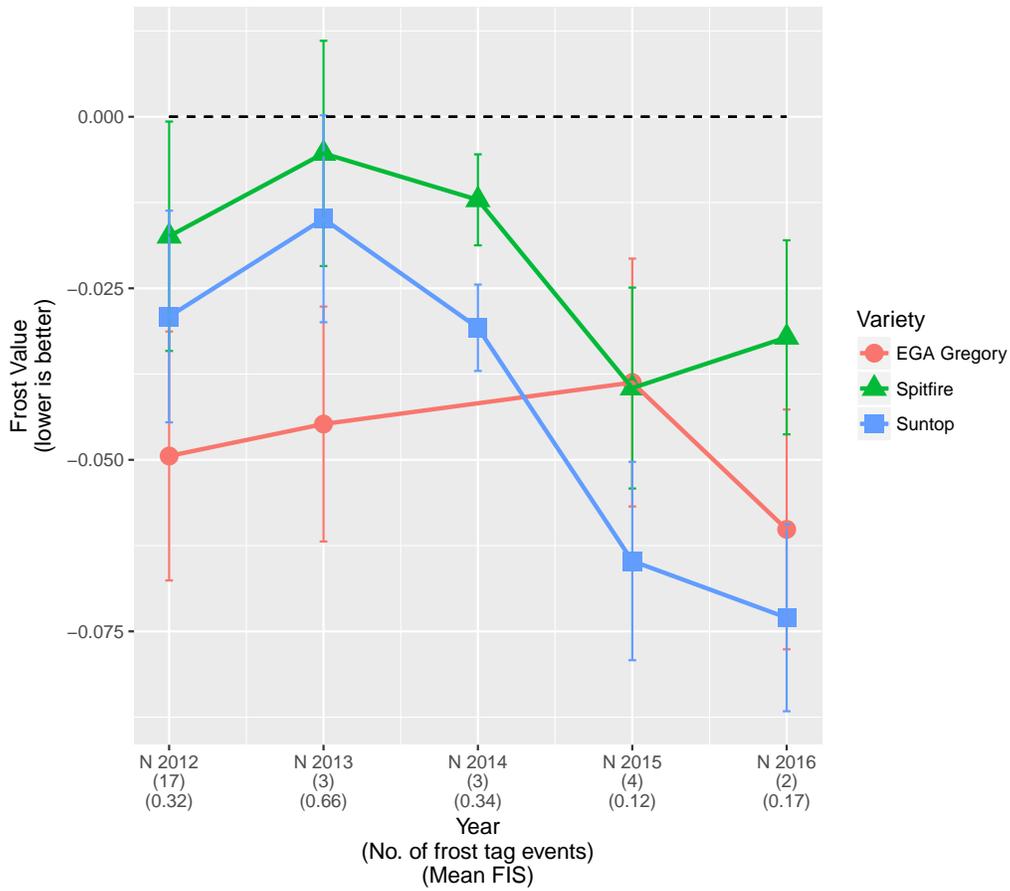


Figure 5: FV-PLUS plot for NSW. Each of the NSW experiments from 2012-2016 are listed on the x -axis and the corresponding predicted FV for each of the three varieties on the y -axis together with standard error bars. The number of tag events and mean FIS for each experiment is also presented underneath the experiment name. Predicted FV are not presented for varieties that are not sown in an experiment

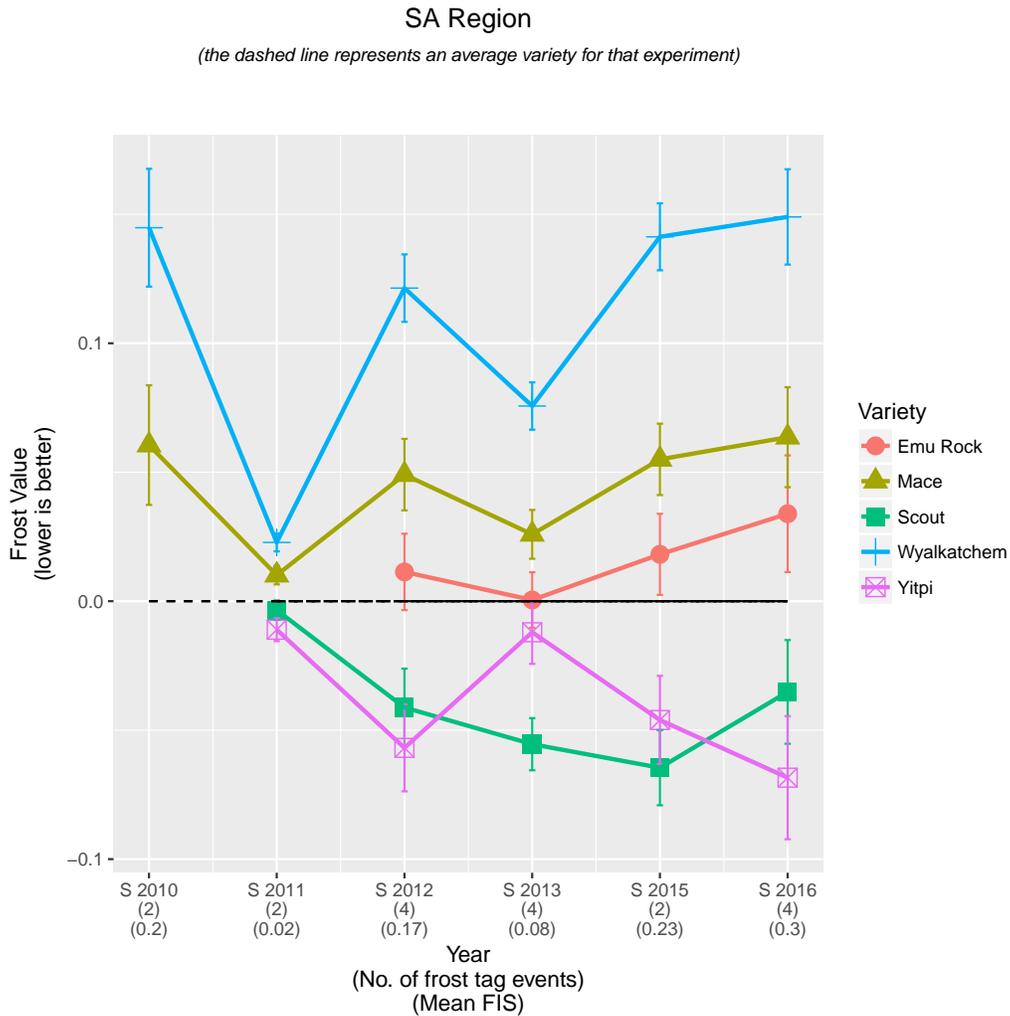


Figure 6: FV-PLUS plot for SA. Each of the SA experiments from 2010-2016 are listed on the x -axis and the corresponding predicted FV for each of the five varieties on the y -axis together with standard error bars. The number of tag events and mean FIS for each experiment is also presented underneath the experiment name. Predicted FV are not presented for varieties that are not sown in an experiment

WA Region

(the dashed line represents an average variety for that experiment)

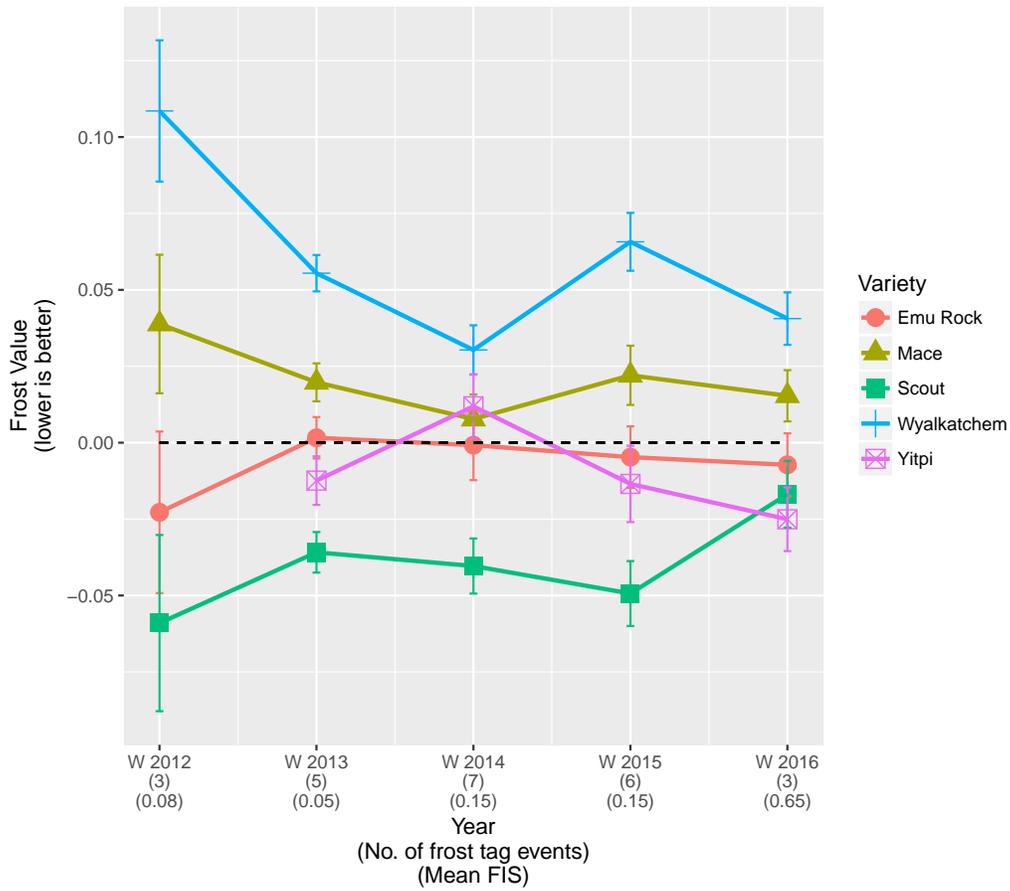


Figure 7: FV-PLUS plot for WA. Each of the WA experiments from 2012-2016 are listed on the x -axis and the corresponding predicted FV for each of the five varieties on the y -axis together with standard error bars. The number of tag events and mean FIS for each experiment is also presented underneath the experiment name. Predicted FV are not presented for varieties that are not sown in an experiment

4. Discussion

Screening cereals for frost risk in Australia has become increasingly important over the years. The enormity of the economic and mental impact of frost on local growers has brought the issue to the forefront as one of the biggest challenges facing the industry. There has been little research undertaken in this space until now as the challenges of not only designing a suitable experiment but also formulation of an appropriate model for the analysis is not to be underestimated. Furthermore, the provision of this information to local growers, breeders and researchers in an informative, valuable and user-friendly format is crucial. We discuss this in detail in the following.

In order to achieve the key outputs of the ANFP (and SFTC project), the trait of interest (FIS), measurement protocols and derivation of the plot and treatment structures for a single FEE and the MET dataset were fundamental to the statistical approach for analysis. Since the FEEs are sown across the Australian wheatbelt, it was expected that real and substantial VEI exists in the dataset. This is of importance to the growers, breeders and researchers. The growers would wish to examine the performance of varieties in environments that experience similar conditions to their own farms. The breeders and researchers on the other hand would be interested in exploiting this information across all environments in identifying varieties that are highly adaptive and therefore consistently exhibit low frost risk in all environments to introgress this trait into their own breeding program.

The DT methodology introduced by Smith and Cullis (2018a) provided an effective and statistically sound approach to the derivation of the plot and treatment structures for the final LMM used for the analysis. The resulting structures reflected the inherent nature of the randomisation of treatments to plots, the longitudinal nature of the data arising from potential repeated measurements over the course of the experiment and features of interest reflected in the measurement protocol such as SOD at tagging. The elegance of this process allowed for immediate identification of statistical aliasing of plot and treatment factors such as with TOS and field block, voiding the potential for invalid inferences to be made for either term. The approach advocated by Brien and Demetrio (2009) for longitudinal factors was also extended in this context to a generic framework that lead to the introduction of a set of terms, namely *anatomical variables*. These variables can be but are not limited to, longitudinal factors such as tag event, environment factors for the analysis of MET data and other factors that are integral to the data

structure such as SOD. Finally, inclusion of factors for which it is anticipated variation exists but are not necessarily of interest are included in the LMM as *extraneous variables*.

The longitudinal MET data was also modelled using a novel yet simplistic approach, by way of only fitting an FA variance model to the `Expt:Variety` term rather than the `Expt:Variety:TagEvent` term as in Smith et al. (2007). The three-way interaction was not plausible in the context of FEEs due to not only the sparsity of the data but also that `TagEvent` is nested within each experiment. In other words, each tag event does not correspond to the same frost event in each experiment and therefore we simply model the terms that are crossed, which is `Expt:Variety`. The VEI in the MET dataset was captured by fitting an FA model to `Expt:Variety` as recently adopted by Smith et al. (2015), in order to provide FV predictions for each variety for every environment in the MET analysis.

The results from the fitted model have allowed us to disseminate information at both the regional level for growers and across the growing regions as a whole. The FV-PLUS plots (Smith et al., 2015) provide an excellent interactive summary to the growers of variety performance across years within state or across all states and years if they wish. They can then filter the environments based on those that are representative of the environmental influences that they also commonly expect for their own farms.

The selection tools more recently developed by Smith and Cullis (2018b) in the plant breeding context are also useful in this setting. Not only would these summary tools provide information to the growers but would also prove useful to both breeders and researchers. Growers would use these tools as an initial overview to then identify a set of varieties to inspect using FV-PLUS graphs to further understand their behaviour. A breeder would be able to isolate a set of high performing lines (low OP) and determine characteristics of each with the aim of introducing these traits into their breeding program. They can also investigate those that appear to be at higher risk and make use of the information that distinguishes these from the remaining main season wheat varieties to understand frost damage within the crop.

There is still further research required in the frost risk of not only cereals but all crops in Australia. It is important that our understanding of the physiology of these crops continues to evolve in order to modify the protocols used in the current FEEs if appropriate. These modifications are then catered for in the statistical approach taken to analyse the resulting data set and therefore the methods outlined in this paper are subject to change due to

this and in order to reduce the level of unexplained variation in the current modeling approach.

Acknowledgements

We gratefully acknowledge the financial support of the Grains Research and Development Corporation (GRDC) of Australia and thank them for use of the data. We acknowledge the work of Simon Diffey in the initial development of the shiny app used to develop the FV-PLUS graphs. We would also like to thank Paul McGowan for his efforts in data collation and auditing. We thank the referees for comments that will improve the manuscript.

Author Contribution Statement

BC and AS developed the statistical approach, with further refinements made by NC. NC and TM were responsible for data auditing and integrity. TM coordinated and managed trials in SA and NSW and BB in WA. Manuscript was prepared by NC, BC and AS. All authors have read and approved the final manuscript.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Al-Issawi, M., Rihan, H., El-Sarkassy, M., Fuller, M., 2012. Frost Hardiness Expression and Characterisation in Wheat at Ear Emergence. *Journal of Agronomy and Crop Science* 199, 66–74.
- Bailey, R., 2008. *Design of Comparative Experiments*. Cambridge University Press, New York.
- Barton, D., Cantrill, L., Law, A., Phillips, C., Sutton, B., Overall, R., 2014. Chilling to zero degrees disrupts pollen formation but not meiotic microtubule arrays in *Triticum aestivum* L. *Plant, Cell and Environment* 37, 2781–2794.
- Baxter, N., March 2015. Trials deliver frost-susceptibility data. *Ground-Cover*.
- Brien, C., Demetrio, C., 2009. Formulating Mixed Models for Experiments, Including Longitudinal Experiments. *Journal of Agricultural, Biological, and Environmental Statistics* 14 (3), 253–280.
- Bueno Filho, J., Gilmour, S., 2007. Block designs for random treatment effects. *Journal of Statistical Planning and Inference* 137 (4), 1446–1451.
- Butler, D., Smith, A., Cullis, B., Thompson, R., 2015. *Asreml: An R package to fit the linear mixed model*.
- Diggle, P., Chetwynd, A., 2011. *Statistics and Scientific Method: An Introduction for Students and Researchers*. Oxford University Press, Oxford.
- Frederiks, T., Christopher, J., Sutherland, M., Borrell, A., 2015. Post-head-emergence frost in wheat and barley : defining the problem , assessing the damage , and identifying resistance. *Journal of Experimental Botany* 66 (12), 3487–3498.
- R Core Team, 2015. *R: A Language and Environment for Statistical Computing*.
- Reinheimer, J. L., Barr, A. R., Eglinton, J. K., 2004. QTL mapping of chromosomal regions conferring reproductive frost tolerance in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 109 (6), 1267–1274.

- Smith, A., Cullis, B., 2018a. Design Tableau : An Aid to Specifying the Linear Mixed Model for a Comparative Experiment. Tech. rep., NIASRA Working Paper Series, Univeristy of Wollongong, Paper 05-18.
- Smith, A., Cullis, B., 2018b. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data.
- Smith, A., Cullis, B. R., Thompson, R., 2001. Analysing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57, 1138–1147.
- Smith, A., Ganesalingam, A., Kuchel, H., Cullis, B., 2015. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. *Theoretical and Applied Genetics* 128 (1), 55–72.
- Smith, A., Stringer, J., Wei, X., Cullis, B., 2007. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. *Euphytica* 157 (1-2), 253–266.