NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

07-21

Use of Contemporary Groups in Multi-Environment Trial Dataset Construction

Alison Smith, Aanandini Ganesalingam, Christopher Lisle, Gururaj Kadkol, Kristy Hobson, and Brian Cullis

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522, Australia Phone +61 2 4221 5435, Fax +61 2 4221 4998.
Email: karink@uow.edu.au

frontiers

Use of contemporary groups in multi-environment trial dataset construction.

Alison Smith [1,*], **Aanandini Ganesalingam** [2], **Christopher Lisle** [1], **Gururaj Kadkol** [3], **Kristy Hobson** [3] and **Brian Cullis** [1]

[1] *Centre for Bioinformatics and Biometrics, National Institute for Applied Statistics Research Australia, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW, Australia*
[2] *InterGrain Pty Ltd, Bibra Lake, WA, Australia. Former address: Centre for Bioinformatics and Biometrics, National Institute for Applied Statistics Research Australia, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW, Australia*
[3] *Tamworth Agricultural Institute, New South Wales Department of Primary Industries, Calala NSW, Australia*

Correspondence*:
Corresponding Author
alismith@uow.edu.au

ABSTRACT

The plant breeding process can be viewed as a multi-year and multi-cycle collection of data. The inherent structure of the breeding program is underpinned by so called contemporary groups, that is, the grouping of lines that are derived together at the beginning of the breeding cycle. These groupings of lines are then tested together as they progress through the evaluation stages. In this paper we introduce and then demonstrate the use of contemporary groups to construct plant breeding multi-environment trial datasets that optimize the information available for selection decisions. We use an optimality criterion from the model-based design literature to quantify this information. On this basis, datasets constructed using contemporary groups are shown to be superior to other forms, in particular those that relate to a single year alone.

Keywords: Multi-environment trials, Linear Mixed Models, Model-based design, Contemporary Groups, Selection

1 INTRODUCTION

Plant breeding is a process that consists of methods for the creation, selection and fixation of superior plants in terms of productivity or quality (Moose and Mumm, 2008). During this process, the ability to select the best lines and discard others is critical in constantly improving the breeding gene pool (Zamir, 2001). Generally, breeding programs have cycle lengths that span eight to ten years; that is from initial cross to variety commercialisation. The majority of programs follow the modified pedigree breeding method, where traits with high heritability are selected first, and traits of lower heritability selected in later generations when lines become fixed (Collard and Mackill, 2008). The early population stages in such programs are often focused on several traits of commercial importance, including disease resistance, herbicide tolerance, phenology type and functional grain quality. At the advanced evaluation stages selection is focused on grain yield across target production environments (TPE), to appropriately gauge variety by environment

interaction (VEI), as varieties vary in their response to different environments. Note that in this paper the terms "variety" and "line" are used synonymously.

While there are multiple traits of interest under selection, yield is often the trait of foremost interest. In order to achieve efficiency in yield selection, yield data is generated from a series of field trials across years (synonymous with seasons) and geographical locations, known as multi-environment trials (METs). METs are an essential evaluation tool in plant breeding programs, as they enable an effective measure of VEI. This is particularly important in the Australian agricultural environment, which is known to be extremely variable between locations and seasons (Chapman et al., 2003). As a result, advanced evaluation stages in breeding programs use expanded numbers of evaluation environments to appropriately assess across all TPE.

The breeding process can be viewed as a multi-year, multi-cycle collection of data (Arief et al., 2019). In particular, we find that programs are underpinned by a grouping of lines that are derived together from a fixed number of crosses at the crossing block stage. These lines are essentially "born" together and are subsequently tested together. This cohort of lines then progress through the bulk population stages, are derived to fixed lines and tested sequentially in the advanced evaluation stages. We refer to these broad grouping of lines as "contemporary groups" (CGs). As an example we consider the Durum Breeding Australia (DBA) program from one of our motivating examples (see Section 3) in which there are four stages of testing (denoted S1 to S4). Within the time-frame 2015 to 2018, the following four CGs were created: CG15, CG16, CG17 and CG18 corresponding to lines in S1 trials in those years. Then, for example, a subset of the lines from CG15 was progressed to S2 trials in 2016 and so on to S3 in 2017 and finally to S4 in 2018. As lines progress through stages they decrease in number, similar to the narrowing observed in the funnel structure seen in Fig. 1. Following on from this figure, four selection decisions would be made annually on S1, S2, S3 and S4 lines as they progress to the next stage of testing. The final stage selection decision, that is, for S4, is concerned with the submission of elite lines to the Australian national crop variety testing program, the National Variety Trials (NVT).

For plant breeding programs in Australia, the preferred approach for the analysis of MET data is the single stage Factor Analytic Linear Mixed Model (FALMM) of Smith et al. (2001), with modelling of spatial variation for individual trials (Gilmour et al., 1997; Stefanova et al., 2009). This analysis approach is also the preferred method of analysis for the NVT (Smith and Cullis, 2018). Oakey et al. (2007) extended these models with the inclusion of information on genetic relatedness through ancestral information (pedigree), which Beeck et al. (2010) utilised to enable partitioning of additive and non-additive effects for selection in a canola breeding program.

There is a substantial amount of literature demonstrating the improved selection accuracy resulting from the use of a one stage FALMM that incorporates pedigree (ancestry) information (see Beeck et al., 2010; Oakey et al., 2007; Ukrainetz et al., 2018; Smith and Cullis, 2018). The inclusion of pedigree information via the numerator relationship matrix (NRM) provides links between varieties both within and between environments. This enables more reliable estimation of genetic variance parameters and thence more accurate predictions of total genetic effects as required for selection.

It is well known that these developments in methods of analysis contribute to genetic gains, as shown by the wide-spread adoption of the FALMM of Smith et al. (2001) in Australian plant breeding programs (Gogel et al., 2018). In Table 1, we have compiled a concise summary of literature specific to plant breeding METs and the methods of analysis. In the process of compiling this summary, we noted that there are very few papers written on the construction of the underlying MET datasets to realise these genetic gains

from an analysis point of view. For example, the paper by Arief et al. (2019) indicates that a MET dataset across stages of a breeding program in a combined analysis is the most advantageous. However, they used a two-stage linear mixed model without the inclusion of pedigree information. In contrast, Yan and Rajcan (2003) used a similar approach (single stage, without pedigree) and found that a single year MET dataset was sufficient to identify the best and worst selections. Despite the vast amounts of data breeding programs generate, it is clear that there is little consensus on the construction of datasets for selection decisions in the current literature.

The inherent structure of the breeding program (stages within a year) and thus selection on a stage basis often results in the analysis of plant breeding datasets comprising a series of single year based analyses (Arief et al., 2019). Bernal-Vasquez et al. (2017) finds similarly that breeding program datasets are often analysed on a year basis and not over years, due to the reasons: that it is simpler/faster and that it is difficult to estimate variation across years due to the lack of common lines between breeding stages. However due to the commercial nature of plant breeding programs, there is limited literature on the dataset construction of plant breeding METs, within the context of a commercial breeding operation.

One of the central aims of plant breeding programs is to select high yielding and stable varieties across environments. This is why plant breeding programs evaluate lines over a large number of locations and years as it enables estimates of random occurring cycles of normal and extreme conditions in the TPE (Rosielle and Hamblin, 1981). In the Australian context large and complex VEI has been reported specifically for wheat breeding (Bänziger and Cooper, 2001). For barley, wheat, oats, lupins, peas, lentils and canola Cullis et al. (2000) found that the main sources of VEI are what they define as "non-static", that is linked to seasonal influences. This alone contributed to 41% of the total variance. Frensham et al. (1998) similarly found with a Southern Australian oat breeding program that the variety by year by location (VYL) variance component accounted for 41.1% of the total phenotypic variance. Arief et al. (2015) summarised multiple plant breeding studies, finding that the VYL variance component as a proportion of total phenotypic variance was 29% for wheat and 25% for navy beans breeding programs in Australia. Therefore, there is a need for multiple years of data to accurately quantify variety performance in the presence of substantial VEI in order to make accurate selection decisions. This is clearly not addressed by single year/single stage based data analysis in a plant breeding program.

The aim of this paper is to demonstrate the utilisation of the CG concept for dataset construction. The basic premise is to include sufficient trials to optimize the amount of data on the lines under consideration for selection. By tracing CGs across stages and years it is possible to form a MET dataset with the desired properties. In order to quantify the impact of this approach we use the A-optimality criterion from model-based design theory. The paper is arranged as follows. Section 2 outlines the methodology for MET dataset construction using CGs. The use of A-optimality for comparing datasets is described. In Section 3 the methods are applied to two motivating examples from Australian plant breeding programs. Some concluding remarks are given in Section 4.

2 METHODS FOR MET DATASET CONSTRUCTION

The CG concept for MET data construction is first illustrated using a hypothetical breeding program with four stages of testing (S1 to S4) and in which lines progress through stages without fast-tracking (skipping stages) or retention (remaining in a stage for more years of testing). The aim is to construct a dataset to enable selection decisions for S1 to S4 in 2018. The simplest process would be to trace the lines in 2018 as far back as possible. This would suggest a separate analysis for each of the selection decisions, based on combining trials in the following stages:

- Selection decision S1: S1 2018
- Selection decision S2: S2 2018 + S1 2017
- Selection decision S3: S3 2018 + S2 2017 + S1 2016
- Selection decision S4: S4 2018 + S3 2017 + S2 2016 + S1 2015

It is instructive to illustrate this compilation of sequences of stages across years using tables such as Table 2. In this table the diagonal bands of stages across years are labelled as A to I, with the labels A to D being assigned in such a way that they align with the S1 to S4 trials in the year of selection (here 2018). In the absence of retention or fast-tracking all the lines in S1, S2, S3 and S4 in 2018 are members of CG18, CG17, CG16 and CG15 respectively. Thus, all the lines within a stage in 2018 belong to a single CG only and the entire selection history for any of these lines is captured in the associated band. The generalisation to more complex scenario will be discussed in the context of the motivating examples (see Section 3).

In terms of information available for each of the four selection decisions it is instructive to differentiate between "direct" and "indirect" information. The former relates to observed data so is maximised by including all trials in which the lines of interest have been grown. In the hypothetical example this corresponds to the bands so suggests the conduct of four analyses each based on a separate band (A, B, C and D). However, the use of a FALMM for analysis creates the possibility of also using indirect information derived from genetically related lines in other trials. We would therefore recommend undertaking a single analysis using data combined across these bands. This recommendation can be justified by applying the method described in Section 2.1 to quantify information for selection. Finally, we note that in the MET analysis, VEI is modelled with reference to environments which are defined to be combinations of trial locations and years. Combining across bands may lead to the presence of multiple trials at a single location within a year. For example, in any given year, locations with S1 trials also typically include S2, S3 and S4 trials. We refer to such trials as "co-located". This will be discussed further in Section 3.

2.1 Quantifying information for selection in MET datasets

In order to discriminate among possible MET datasets in terms of the amount of information available for selection decisions, we note that the problem has strong links with optimal (model-based) design. As Butler et al. (2014) state, "The goal of optimal design is to discriminate among competing designs in an effort to maximise the treatment information from a fixed number of experimental units." This requires the use of an optimality criteria, and, in the context of plant breeding trials in which the treatments are varieties and the aim is selection, the A-optimality criteria is appropriate since this aligns with minimising the probability of an incorrect selection decision (Bueno Filho and Gilmour, 2003). A-optimality is based on the so-called A-value which is the average pairwise variance of elementary treatment contrasts. We therefore propose to use A-values to quantify the treatment (variety) information available in any given MET dataset.

In model-based design, A-values are computed under a pre-specified Linear Mixed Model (LMM) which we will term the design model. Specification of the design model requires specification of the fixed and random effects, the variance models for the random effects and residuals and the values of the associated variance parameters. The design model is usually chosen to be as close as possible to that expected for the analysis. Additionally, the variance parameter values are chosen as being "typical" so may be based on historic analyses. The model proposed in this paper for the analysis of MET data is the FALMM. Variety selections using this model are typically focussed on the measure of overall variety performance (across environments) as presented in Smith and Cullis (2018). However the factor analytic variance parameters are specific to the individual environments in the dataset so that typical values do not exist. Therefore a more generic, but still realistic design model is required for assessing MET dataset information. We have

chosen a variance component model that involves random variety main effects and random VEI effects, both of which are partitioned into additive and non-additive effects. This is, in fact, a sub-model of the FALMM. The A-values are then computed for the total (additive plus non-additive) variety main effects since these provide a measure of average performance of varieties across environments.

In order to determine reasonable values for the variance parameters in this design model we consider Cullis et al. (2000) who conducted variance component analyses of grain yield in 22 MET datasets from Australian crop variety evaluation programs. The environments in those datasets were classified according to the year, the geographic region and possibly location within region so Cullis et al. (2000) partitioned VEI accordingly. In our motivating examples we do not have regional information nor are trials typically located in identical positions from year to year. However, we recognise the importance of variety by year interaction so maintain this as a separate source in the design model. Thus we have used the variety main effects (V), variety by year interaction (VY) and Error sources of variation from Cullis et al. (2000), and have added together the remaining sources to form residual variety by environment (VE) interaction. The mean percentage contributions for each of these sources across all 22 datasets was 13.77% (V), 8.59% (VY), 37.91% (residual VE) and 39.73% (Error) (see Table 3). In the model-based design literature, and without loss of generality, a value of one is typically assumed for the error variance. We adopt the same approach here (see second row in Table 3).

The analyses in Cullis et al. (2000) do not involve information on genetic relatedness. We therefore make the further assumption that additive variance comprises 80% of total variance. This represents an average from the analyses of numerous Australian plant breeding datasets. The final values for the variance parameters in the design model are given in the third and fourth rows of Table 3. All A-values in this paper were computed using ASReml-R (Butler et al., 2009). The code is provided in Appendix A.

3 RESULTS

In this section we show the application of the methods presented in section 2 to two motivating examples.

3.1 Durum breeding program

Durum wheat (*Triticum durum desf.*) breeding in Australia is currently resourced on agronomic zones of production and funded by New South Wales Department of Primary Industry (NSW DPI), The University of Adelaide and the Grains Research and Development Corporation (GRDC) under the Durum Breeding Australia (DBA) project. The motivating example of our paper is the DBA North program that operates out of Tamworth Agricultural Institute, capturing TPEs in New South Wales (NSW) up to and including central Queensland (QLD). The aim of this paper (and this breeding program) is to evaluate the performance of the 2018 S1, S2, S3 and S4 lines for selection and progression to the next stage of breeding.

The structure of the program is illustrated in Fig. 1. The S1 in any year contains on average 1120 lines evaluated across one to two sites. As the program progresses the line numbers decrease to on average, 60 lines in the S4, evaluated at six sites. The numbers of test lines for each stage and for the years 2013 to 2018 in the durum breeding program are given in Table 4. Note that test lines refer only to the lines under consideration for selection, so excludes check varieties, for example. At any stage of selection, a line may be selected to progress to the next stage of testing, retained in the same stage or rejected. In contrast to the hypothetical example, lines are often retained within later stages for additional year/s of testing. Retentions may occur due to limitations in seed production, or even a holding pattern while awaiting disease and/or quality data. This has resulted in the lines in later stages comprising a mixture of CGs. The distribution across CGs for 2018 lines are given as the final columns in Table 4. For example, the majority (66) of lines in S3 in 2018 correspond to CG16 so have followed the simple progression along band C (that is,

they progressed from S1 trials in 2016 to S2 trials in 2017 to S3 trials in 2018). But a fair number (22) correspond to CG15 and progressed from S1 trials in 2015 to S2 trials in 2016 to S3 trials in 2017 and were then retained in S3 in 2018. Finally, five lines correspond to CG14 and progressed from S1 trials in 2014 to S2 trials in 2015 to S3 trials in 2016 and were then retained in S3 in 2017 and 2018. This has implications for construction of the MET dataset.

The starting point for MET dataset construction for selection decisions on the 2018 lines (S1 to S4) involves bands A-D as described in the hypothetical example. With the retention of lines it is clear that this would fail to capture much of the data on the 30 S3 and S4 lines in 2018 that belonged to CG14 and CG13 (see Table 4). For example, Table 5 shows there are nine lines in S4 for which five years of data would be missing if the dataset comprised only bands A-D; another nine lines for which four years would be missing and a further seven lines missing two or three years of data. We believe this is unacceptable. We therefore investigate the addition of bands E and F to the data. Table 5 shows the improvement in capturing more of the data on the lines of interest. Complete data could be obtained by adding band G but we caution against this because we do not have the full selection history for many of the lines in band G since it commenced in 2012 (see Table 2). The inclusion of band G, or indeed bands H and I (so that the full rectangle of data is included) may result in "selection bias", that is, bias in the estimates of the genetic variance parameters (Thompson, 1973) so is not recommended. The final dataset is therefore chosen to comprise bands A-F. With this dataset only five lines under scrutiny for selection (in S4) are missing data and the amount missing is small (one or two years out of a total of six years).

The final MET dataset (bands A-F) for analysis comprised yield data on 6,951 lines from 21,660 plots corresponding to 97 trials across 30 environments. Each field trial was sown as a rectangular array indexed by field rows and columns. Trials were sown in a serpentine sequence and harvested in the row direction with all other management regimes applied via pathways in the column dimension. Trials were designed as grid-plot, partially replicated (*p*-rep) (Cullis et al., 2006) or randomised complete block designs, with two to three replicates. Summary information for the 30 environments is given in Table 6. There were 15 environments with co-located trials, ranging in number from two to 13. The co-located trials related either to different stages or to multiple trials for stage S1 or S2 (due to physical restrictions on trial sizes). We note that co-located trials are only deemed to comprise a single environment when they are all managed in the same way, that is, they are sown and harvested within a similar time frame and subjected to the same agronomy practices including fertiliser, herbicide and pathway regimes.

The pedigree information associated with the above trial data contained 7,628 records. All lines in the MET data set had pedigree information. This was the first time a pedigree file had been created for this breeding program and included in the analysis. This was a significant undertaking as the durum breeding program was established in the 1960's and a comprehensive pedigree file (outside annual crossing block information) had not existed in the program since this time. The NRM was formed using the *pedicure* package (Butler, 2019) in R (R Development Core Team, 2015). The inbreeding coefficients of lines with phenotypic data ranged from 0.750 to 0.998 with mean of 0.905.

Finally, the approach described in section 2.1 for comparing MET datasets in terms of the information for selection was applied for each stage of selection and for three types of MET dataset, namely the 2018 data for each stage, the diagonal band of data for each stage and the final dataset (bands A-F). Thus, for S4 selections, the three datasets comprised data from S4 trials in 2018 alone; data from trials in band D (S4 trials in 2018 + S3 trials in 2017 + S2 trials in 2016 + S1 trials in 2015) and the final dataset. For S3 selections, the three datasets comprised data from S3 trials in 2018 alone; data from trials in band C (S3 trials in 2018 + S2 trials in 2017 + S1 trials in 2016) and the final dataset. For S2 selections, the three

datasets comprised data from S2 trials in 2018 alone; data from trials in band B (S2 trials in 2018 + S1 trials in 2017) and the final dataset. Note that for S1 selections, the single year dataset (S1 trials in 2018) is equivalent to the band dataset (band A). The resultant A-values are shown in Figure 2. This clearly shows the superiority of the final dataset in each case. The reduction in A-values is largely driven by an increase in the amount of direct information (as reflected in the mean numbers of environments per line) but there is also a strong contribution from indirect information. For example, the S1 lines under consideration for selection in 2018 were only grown in a single environment so there is no difference in direct information between using the 2018 data alone for this stage compared with the final dataset. However the A-value for the final dataset is much lower, indicating the impact of indirect information from relatives of the S1 lines.

3.2 Chickpea breeding program

In Australia, the chickpea (*Cicer arietinum* L.) breeding program is managed under the umbrella of Pulse Breeding Australia (PBA). PBA is an Australian government funded project through the Grains Research and Development Corporation (GRDC). PBA coordinates and funds the breeding activities of the four pulse crops of economic importance - chickpeas, field peas, faba beans and lentils across the Australian growing environment. The chickpeas program under PBA comprises three sub-programs, which are based on the following germplasm streams - desi (north and south) and kabuli. We will focus on the desi south sub program managed by NSW DPI.

The chickpea breeding program evaluates lines annually, across southern NSW, Victoria and South Australia in three stages, namely S1, S2 and S3 across multiple locations. Here we consider the MET dataset construction to evaluate the performance of test lines in all three stages in 2019.

The structure of the program is similar to that of the program illustrated in Fig. 1. The S1 in any year contains on average 650 lines evaluated across five sites. As the program progresses the line numbers decrease to on average, 70 lines in the S3, evaluated at eight sites. The numbers of test lines for each stage and for the years 2016 to 2019 in the chickpea breeding program are given in Table 7. The distribution across CGs for 2019 lines are given as the final columns in Table 7. Thus, for example, the majority (46) of lines in S3 correspond to CG17 (so have followed the simple progression along band C) but 11 correspond to CG16 and one to CG14.

The starting point for MET dataset construction for selection decisions on the 2019 lines (S1 to S3) involves bands A-C. However this would fail to capture all the data on the 11 S3 lines in 2019 that belonged to CG16 and the single line that belonged to CG14 (also see Table 8). Addition of band D accommodates the former and provides an additional year for the latter. Complete data could be obtained by using bands A-F but once again we caution against this because it introduces many lines with incomplete selection histories. The final dataset is therefore chosen to comprise bands A-D. With this dataset only a single line (in S3) is missing data. We note that this line has been retained in S3 for a number of years and is of less interest in terms of promotion to the next stage of testing (in the NVT).

The final MET dataset (bands A-D) for analysis comprised yield data on 2,448 lines from 18,936 plots corresponding to 56 trials across 28 environments. Summary information for the 28 environments is given in Table 9. There were 18 environments with co-located trials, with between two to three trials.

The pedigree information associated with the above trial data contained 2,983 records. All lines in the MET data set had pedigree information. The inbreeding coefficients of lines with phenotypic data ranged from 0.500 to 0.999 with mean of 0.7404.

The information available for selection at each stage was assessed using a similar approach to the durum example, namely computing A-values for test lines for three types of MET dataset. For S3 selections, the

three datasets comprised data from S3 trials in 2019 alone; data from trials in band C (S3 trials in 2019 + S2 trials in 2018 + S1 trials in 2017) and the final dataset (trials in bands A-D). For S2 selections, the three datasets comprised data from S2 trials in 2019 alone; data from trials in band B (S2 trials in 2019 + S1 trials in 2018) and the final dataset. Note that for S1 selections, the single year dataset (S1 trials in 2019) is equivalent to the band dataset (band A). The resultant A-values are shown in Figure 3. Once again the superiority of the final dataset in each case is clearly shown, with greatly reduced A-values compared to the single year and single band datasets.

4 DISCUSSION

In this paper we have addressed a void in the literature, and provided a rigorous framework for the construction of MET datasets for selection in plant breeding programs. We have described a method that aims to optimize the amount of information available on the "lines of interest", that is, the lines under consideration for selection. The method is intuitive and involves several simple steps. A key aspect is to identify contemporary groups (CGs), that is, groups of lines that entered their first stage of testing (S1) in the same year. This allows a complete enumeration of the trials in which the lines of interest have been grown. In addition to defining CGs, which relate to lines, we have also introduced the concept of data bands, which relate to trials. Data bands align with the testing system, namely the progression through stages (for example S1 through to S4) from year to year. By sequentially building the MET dataset using data bands we can form a dataset that captures as much information as possible for the lines of interest.

We have also developed a method for quantifying this information for any given MET dataset. The method uses fundamental concepts from model-based design theory. Thus information is quantified using an A-value (average pairwise variance) for the lines of interest from a pre-specified linear mixed model. The application of this approach to the two motivating examples in this paper clearly showed the superiority of the MET datasets constructed using the CG approach, in particular when compared with the more common approach of simply using trials from a single stage and year. The information gains for all selection stages were associated both with direct information, that is, from the trials in which the lines of interest were grown and also indirect information derived from trials in which genetically related lines were grown.

MET datasets constructed using the CG approach encompass multiple bands of data and this suggests a further and crucial gain for S1 selection decisions since these lines have only been grown in a single year and often at only one or two locations. The key driver here is that locations with S1 trials also typically include later stage trials. The environment is then defined as the amalgamation of all of these (co-located) trials and thence includes a far wider set of lines than the S1 lines alone. This provides links between environments (both within and between years) for the S1 lines. Then application of the FALMM will allow examination of VEI for the S1 lines of interest across a wide range of environments. This is particularly important in an Australian context as variety by location by year interaction is often the largest component of VEI (Cullis et al., 2000; Frensham et al., 1998). Hence the additional seasons of data will lessen the impact of selection on a single year of data, which could be a seasonally extreme year and therefore outside what is expected for the range of TPE. This issue has also been pointed out by Arief et al. (2015).

We note that combining trials utilising the CG approach may result in an unbalanced dataset, with poor connectivity (low numbers of lines in common) between some environments. Whilst one of the advantages of using an FALMM is the ability to handle unbalanced data, there has been concern about the reliability of estimation of FA parameters in extreme cases when connectivity is very poor. It is well known that poorly estimated genetic variance parameters will result in a reduction in genetic gain (Sales and Hill, 1976a,b). The A-value approach does not take this into account since the variance parameters are assumed known.

Historically, variety connectivity was thought to be the key determinant of the reliability of estimation of FA parameters. More recently Lisle et al. (in preparation) have developed a formal information based diagnostic for this purpose. It is superior to connectivity in the sense of better forecasting the uncertainty of variance parameter estimates and being applicable for both additive and non-additive genetic variance parameters. It may therefore be applied jointly with the A-value approach in order to balance variety information and reliability of variance parameter estimation in the search for an optimum MET dataset.

The applications in this paper are based on inbred annual crops. However we note that our methods can be easily modified for situations that include hybrid crops, which require evaluation of inbred parental lines in addition to the hybrids themselves. Furthermore, the methods are applicable for parental evaluation in perennial crops such as radiata pine, in which parental trees are evaluated by the performance of their progeny in (field) trials across numerous years and seasons (Smith and Cullis, 2018).

Our methodology for constructing a MET dataset has been programmed within the R computing environment (R Development Core Team, 2015) and the code is available upon request. The method for quantifying information involves the use of ASReml-R (Butler et al., 2009) and example code is given in Appendix A. We have implemented these methods in a number of plant breeding programs with the result of superior MET datasets and also improved breeder confidence and understanding.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

BC and AS conceived the ideas and developed the methodology. AS wrote all sections of the manuscript apart from the Introduction which was written by AG. AG and CL prepared and curated the datasets. GK and KH provided the data and input on plant breeding perspectives. AS and CL performed the statistical calculations. CL and BC contributed to manuscript revision. All authors read and approved the submitted version.

FUNDING

AS was supported by the Grains Research and Development Corporation (GRDC) through the EssCargoT project (UW00010). The PBA Chickpea program received investment from GRDC, New South Wales Department of Primary Industries (NSW DPI), Queensland Department of Agriculture and Fisheries (DAF), Agriculture Victoria Research (AVR) and South Australian Research and Development Institute (SARDI).

ACKNOWLEDGMENTS

The authors thank the reviewers whose comments have led to an improved manuscript.

REFERENCES

Arief, V. N., DeLacy, I. H., Crossa, J., Payne, T., Singh, R., Braun, H.-J., et al. (2015). Evaluating testing strategies for plant breeding field trials: redesigning a CIMMYT international wheat nursery. *Crop Science* 55, 164–177

Arief, V. N., Desmae, H., Hardner, C., DeLacy, I. H., Gilmour, A., Bull, J. K., et al. (2019). Utilization of multiyear plant breeding data to better predict genotype performance. *Crop Science* 59, 480–490

Bänziger, M. and Cooper, M. (2001). Breeding for low input conditions and consequences for participatory plant breeding examples from tropical maize and wheat. *Euphytica* 122, 503–519

Beeck, C., Cowling, W., Smith, A., and Cullis, B. R. (2010). Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. *Genome* 53, 992–1001

Bernal-Vasquez, A.-M., Gordillo, A., Schmidt, M., and Piepho, H.-P. (2017). Genomic prediction in early selection stages using multi-year data in a hybrid rye breeding program. *BMC genetics* 18, 51

Bueno Filho, J. S. and Gilmour, S. G. (2003). Planning incomplete block experiments when treatments are genetically related. *Biometrics* 59, 375–381

Butler, D. (2019). *Pedicure: pedigree tools*. R package version 2.0.0

Butler, D., Cullis, B., Gilmour, A. R., and Gogel, B. (2009). *Mixed models for S language environments, ASReml-R reference manual.* Training and development series, No QE02001, QLD Department of Primary Industries and Fisheries, Brisbane, QLD.

Butler, D. G., Smith, A. B., and Cullis, B. R. (2014). On the design of field experiments with correlated treatment effects. *Journal of Agricultural, Biological and Environmental Statistics* 19, 539–555

Chapman, S., Cooper, M., Podlich, D., and Hammer, G. (2003). Evaluating plant breeding strategies by simulating gene action and dryland environment effects. *Agronomy Journal* 95, 99–113

Collard, B. C. Y. and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 363, 557–572

Cullis, B., Gogel, B., Verbyla, A., and Thompson, R. (1998). Spatial analysis of multi-environment early generation variety trials. *Biometrics* 54, 1–18

Cullis, B., Jefferson, P., Thompson, R., and Smith, A. (2014). Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a Pinus radiata breeding programme. *Theoretical and Applied Genetics* 127, 2193–2210. doi:10.1007/s00122-014-2373-0

Cullis, B., Smith, A., and Coombes, N. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics* 11, 381–393. doi:10.1198/108571106X154443

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

Cullis, B. R., Smith, A. B., Hunt, C., and Gilmour, A. R. (2000). An examination of the efficiency of Australian crop variety evaluation programmes. *Journal of Agricultural Science, Cambridge* 135, 213–222

Frensham, A., Barr, A., Cullis, B., and Pelham, S. (1998). A mixed model analysis of 10 years of oat evaluation data: use of agronomic information to explain genotype by environment interaction. *Euphytica* 99, 43–56

Gilmour, A. R., Cullis, B. R., and Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics* 2, 269–293

Gogel, B., Smith, A., and Cullis, B. (2018). Comparison of a one-and two-stage mixed model analysis of Australia's National Variety Trial Southern Region wheat data. *Euphytica* 44

Guilly, S., Dumont, T., Thong-Chane, A., Barau, L., and Hoarau, J.-Y. (2017). Analysis of multienvironment trials (MET) in the sugarcane breeding program of Réunion Island. *Euphytica* 213

Lisle, C., Smith, A., and Cullis, B. (in preparation). Multi-environment trial dataset construction: a diagnostic to assess information available for variance parameter estimation.

Moose, S. P. and Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant physiology* 147, 969–977

Oakey, H., Verbyla, A., Cullis, B., Wei, X., and Pitchford, W. (2007). Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics* 114, 1319–1332. doi:10.1007/s00122-007-0515-3

Perez-Elizalde, S., Jarquin, D., and Crossa, J. (2012). A general Bayesian estimation method of linear–bilinear models applied to plant breeding trials with genotype×environment interaction. *Journal of Agricultural, Biological, and Environmental Statistics* 17, 15–37

R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

Redden, R., DeLacy, I., Butler, D., and Usher, T. (2000). Analysis of line x environment interactions for yield in navy beans. 2. Pattern analysis of lines and environment within years. *Australian Journal of Agricultural Research* 51, 607–617

Rosielle, A. and Hamblin, J. (1981). Theoretical aspects of selection for yield in stress and non-stress environment. *Crop science* 21, 943–946

Sabaghnia, N., Sabaghpour, S., and Dehghani, H. (2008). The use of an AMMI model and its parameters to analyse yield stability in multi-environment trials. *The Journal of Agricultural Science* 146, 571

Sales, J. and Hill, W. (1976a). The effect of sampling errors on the efficiency of selection indices. 1. Use of information on relatives for single trait improvement. *Animal Production* 22, 1–17

Sales, J. and Hill, W. (1976b). The effect of sampling errors on the efficiency of selection indices. 2. Use of information on associated traits for improvement of a single trait. *Animal Production* 23, 1–14

Smith, A. and Cullis, B. (2018). Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica* 214, 143. doi:10.1007/s10681-018-2220-5

Smith, A. B., Cullis, B. R., and Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57, 1138–1147. doi:10.1111/j.0006-341X.2001.01138.x

Stefanova, K. T., Smith, A. B., and Cullis, B. R. (2009). Enhanced diagnostics for the spatial analysis of field trials. *Journal of Agricultural, Biological, and Environmental Statistics* 14, 392–410. doi:10.1198/jabes.2009.07098

Thomason, W. and Phillips, S. (2006). Methods to evaluate wheat cultivar testing environments and improve cultivar selection protocols. *Field Crops Research* 99, 87–95

Thompson, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics* 29, 527–550

Ukrainetz, N. K., Yanchuk, A. D., and Mansfield, S. D. (2018). Climatic drivers of genotype–environment interactions in lodgepole pine based on multi-environment trial data and a factor analytic model of additive covariance. *Canadian Journal of Forest Research* 48, 835–854

Yan, W. and Rajcan, I. (2003). Prediction of cultivar performance based on single-versus multiple-year tests in soybean. *Crop Science* 43, 549–555

Yan, W. and Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian journal of plant science* 86, 623–645

Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nature reviews genetics* 2, 983–989

FIGURE CAPTIONS

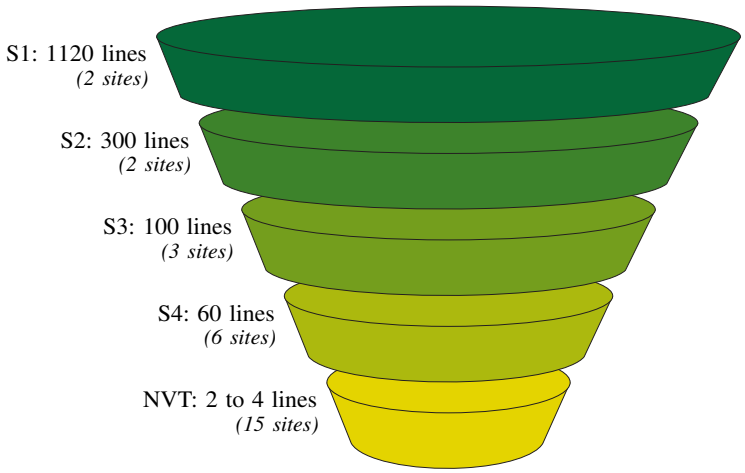

Figure 1. Summary of the typical number of lines and trial sites (in parentheses) at each testing stage for one of our motivating examples, the durum breeding program. There are four breeding stages of testing: stage 1 (S1) to stage 4 (S4) and a final stage corresponding to the independent Australian testing system, the National Variety Trials (NVT).

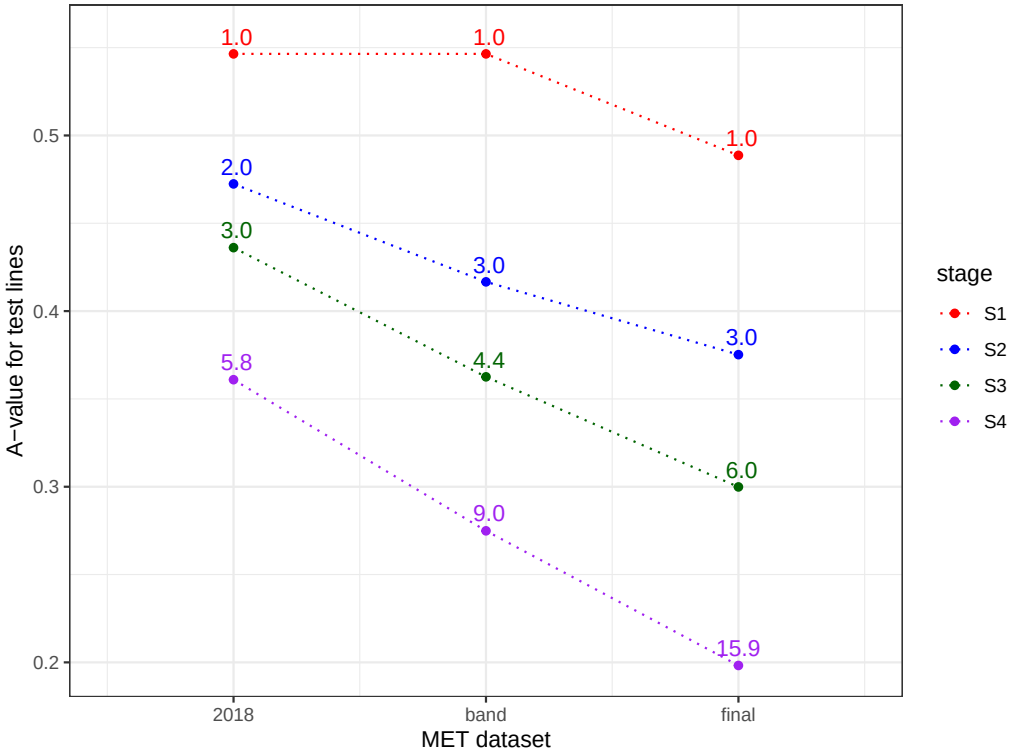

Figure 2. A-values for lines under consideration for selection in durum breeding program in 2018 (stage S4: 56 lines, S3: 93 lines, S2: 315 lines and S1: 1148 lines). A-values are given for three types of MET dataset, namely 2018 data for each stage; diagonal band of data for each stage and the final dataset constructed as in section 3.1. The points are labelled with the associated mean numbers of environments in which these lines were grown. Note that for S1 the 2018 and band datasets are the same.

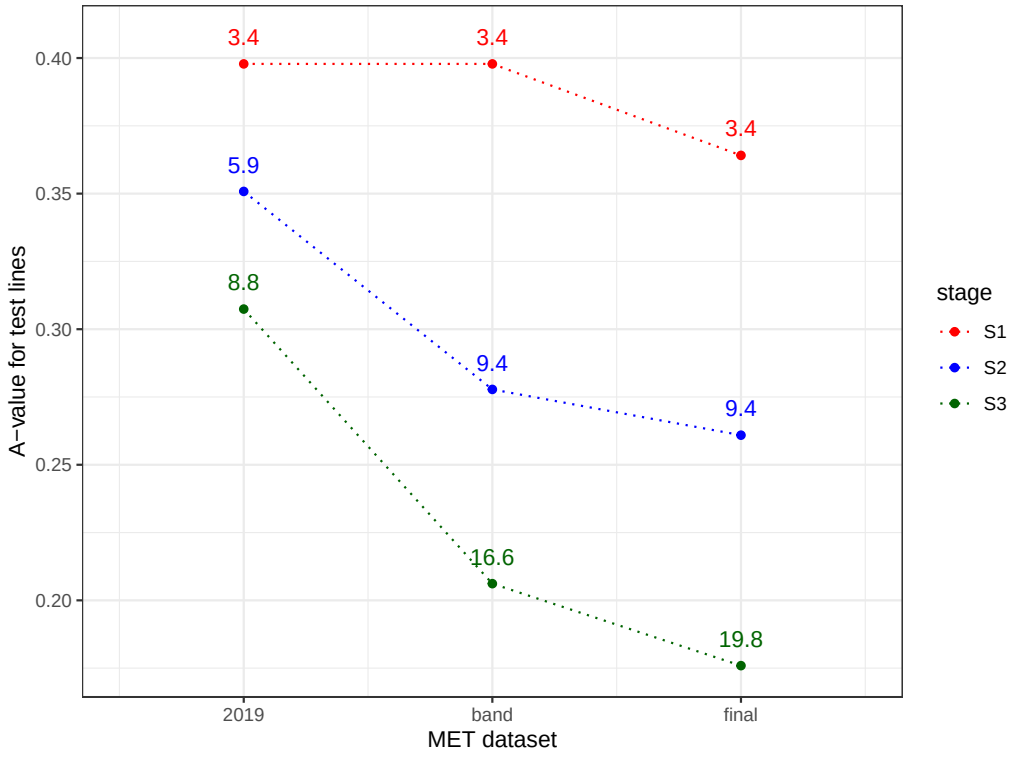

Figure 3. A-values for lines under consideration for selection in chickpea breeding program in 2019 (stage S3: 58 lines, S2: 176 lines and S1: 638 lines). A-values are given for three types of MET dataset, namely 2019 data for each stage; diagonal band of data for each stage and the final dataset constructed as in section 3.2. The points are labelled with the associated mean numbers of environments in which these lines were grown. Note that for S1 the 2019 and band datasets are the same.

Table 1. Summary of studies based on plant breeding METs, their dataset composition and methods of analysis. The Selection column indicates if the aim of the paper was selection on the given dataset. Abbreviations are: GY = Grain Yield; SD = Stem Diameter; TH = Tree Height; AMMI = Additive Main effect and Multiplicative Interaction; ARAM = Approximate Reduced Animal Model; FALMM = Factor Analytic Linear Mixed Model; LMM = Linear Mixed Model; VCLMM = Variance Component Linear Miexed Model

Crop - Trait	Selection?	Aim	Dataset	Analysis	Reference
Barley & Wheat, GY	N	Present an approach for the analysis of early stage breeding METs	Barley S3 comprising 125 lines at 3 sites in 1992 & Wheat early generation yield trials sown at 3 sites in 1991	1 & 2 stage LMM	Cullis et al. (1998)
Oat, GY	N	Identify long-term sources of variation	S4 trials of 10-20 lines across 22 locations across 2 years, (174 trials spanning 1985 to 1994)	2 stage LMM	Frensham et al. (1998)
Navy beans, GY	N	Report the results of Pattern analysis on on TPE	MET comprising 15 locations across the years 1983 to 1989	VCLMM	Redden et al. (2000)
Soy bean, GY	N	Evaluate multi-year data vs single year data of variety performance	MET dataset comprising variety trials at 4 locations per year, from 1991 to 2000	LMM	Yan and Rajcan (2003)
Wheat, GY	N	Review the principles of biplot analysis for MET data	MET dataset comprising 18 winter wheat varieties tested at 9 Ontario locations in 1993	LMM	Yan and Tinker (2006)
Wheat, GY	Y	Identify relevant testing environments and improve predictive value of data	MET comprising: 22 varieties evaluated in 32 site years	LMM	Thomason and Phillips (2006)
Sugarcane	Y	Demonstrate a statistical approach for METs to enable selection of parents and best performing lines	Stage 2 & 3 clones at 6 sites across 2 years (2002 to 2003)	FALMM & pedigree	Oakey et al. (2007)
Lentils, GY	Y	Explore the AMMI model for selection	MET dataset comprising 11 varieties evaluated at 7 locations over 2-3 years	AMMI	Sabaghnia et al. (2008)
Canola, GY & Oil	Y	Develop tools to explore VEI	MET comprising 19 trials in advanced stage of breeding across southern Australian during 2007 to 2008	FALMM & pedigree	Beeck et al. (2010); Cullis et al. (2010)
Maize, GY	Y	Investigate VEI on selection decisions	MET comprising, 12 varieties and 25 environments evaluated in two consecutive years	AMMI	Perez-Elizalde et al. (2012)
Pine Breeding, SD	Y	Present an approach for investigating additive VEI in an outcrossing plant species	77 trials, located in sites across Australia & New Zealand with planting dates spanning the period 1968 to 2005	ARAM & FALMM & pedigree	Cullis et al. (2014)
Wheat, 21 traits	Y	Evaluate methods to obtain reliable estimates of variance components	MET comprising single cycle of breeding nurseries: 466 unique locations across 30 years	VCLMM	Arief et al. (2015)
Sugarcane, 3 traits	Y	Characterise the varieties in the final selection stage of breeding	MET dataset comprising four consecutive variety series (S00, S03, S04, and S05) planted in the years 2011 to 2014 at 7 sites	VCLMM	Guilly et al. (2017)
Pine Breeding, SD	Y	Obtain predicted estimated breeding values for parental selection	107 trials planted across the years 1968 to 2005	FALMM & pedigree	Smith and Cullis (2018)
Lodgepole pine, TH	Y	To model patterns of VEI for a tree breeding program	MET dataset comprising 28 second generation progeny test sites established during 2002 to 2006 across 5 breeding zones	ARAM & FALMM & pedigree	Ukrainetz et al. (2018)
Maize, GY	Y	Evaluate the benefit of multi-year and multi-stage data for seection	MET comprising stages 1 to 6 across 100s of locations (some containing multiple trials) across 6 years	2 stage FALMM	Arief et al. (2019)

Table 2. Sequences of stages and years for potential inclusion in a MET dataset for selection decisions in 2018 from a breeding program with four stages of selection (S1 to S4). Sequences correspond to the diagonal bands labelled as A to I.

	Year					
Stage	2013	2014	2015	2016	2017	2018
S1	F	E	D	C	B	A
S2	G	F	E	D	C	B
S3	H	G	F	E	D	C
S4	I	H	G	F	E	D

Table 3. Variance parameter values for design model. Rows in the table are: means of percentages in Cullis et al. (2000); associated (total) variance parameter values assuming error variance of one; additive variance parameter values (numerator is 80% of total values and denominator, $\bar{a}$, is the mean of the diagonal elements of NRM); non-additive variance parameter values (20% of total values).

	V	VY	VE	Error
Mean % from Cullis et al. (2000)	13.77	8.59	37.91	39.73
Total variance parameter	0.35	0.22	0.95	1.00
Additive variance parameter	$0.28/\bar{a}$	$0.18/\bar{a}$	$0.76/\bar{a}$	
Non-additive variance parameter	0.07	0.04	0.19	

Table 4. Number of test lines in each stage and year in the durum breeding program with data bands indicated as superscripts. The final columns give the number of lines in each contemporary group (CG18-CG13) for lines under consideration for selection in 2018.

	Number of test lines						Number of 2018 test lines					
Stage	2013	2014	2015	2016	2017	2018	CG18	CG17	CG16	CG15	CG14	CG13
S1	582^{F}	1485^{E}	1000^{D}	1163^{C}	1303^{B}	1148^{A}	1148	0	0	0	0	0
S2	105^{G}	361^{F}	413^{E}	388^{D}	379^{C}	315^{B}	0	315	0	0	0	0
S3	30^{H}	92^{G}	92^{F}	92^{E}	90^{D}	93^{C}	0	0	66	22	5	0
S4	25^{I}	41^{H}	57^{G}	55^{F}	53^{E}	56^{D}	0	0	0	31	12	13

Table 5. MET dataset construction for 2018 selection decisions in durum breeding program. Number of lines missing years of data in datasets comprising bands A-D, A-E, A-F and A-G. Total number of lines for selection: 1148, 315, 93 and 56 (for stages S1 to S4).

	# years	Bands in dataset			
Stage	missing	A-D	A-E	A-F	A-G
S1	0	1148	1148	1148	1148
S2	0	315	315	315	315
S3	0	88	93	93	93
	2	5	0	0	0
S4	0	31	42	51	56
	1	0	1	3	0
	2	6	0	2	0
	3	1	0	0	0
	4	9	13	0	0
	5	9	0	0	0

Table 6. Summary of environments in the durum MET dataset: number of trials for each stage of testing (S1, S2, S3, S4) and total number of trials. Number of plots (nplot) and lines (nline); mean yield (t/ha).

	Number of trials							Mean
Environment	S1	S2	S3	S4	Total	nplot	nline	yield
2013-Breeza	2	0	0	0	2	720	585	2.88
2014-Breeza	3	0	0	0	3	1296	937	4.28
2014-Edgeroi	0	4	0	0	4	768	364	2.78
2014-Tworth	2	4	0	0	6	1468	915	3.93
2015-Breeza	0	0	2	0	2	384	96	5.15
2015-Edgeroi	0	4	1	0	5	1056	498	1.83
2015-Nstar	0	0	1	0	1	192	96	5.05
2015-Tworth	6	4	1	0	11	2244	1499	4.00
2016-Breeza	0	0	1	1	2	372	152	4.35
2016-Edgeroi	0	0	0	1	1	180	60	4.79
2016-Gurley	0	0	0	1	1	180	60	5.62
2016-Nstar	0	0	1	2	3	552	152	5.49
2016-Tworth	6	3	1	1	11	2628	1704	4.81
2017-Blbgra	0	0	0	1	1	180	60	1.12
2017-Breeza	0	0	1	1	2	384	158	5.31
2017-Bribbaree	0	0	0	1	1	180	60	1.20
2017-Coonamble	0	0	0	1	1	180	60	1.61
2017-Edgeroi	0	0	0	1	1	180	60	3.93
2017-Garah	0	0	0	1	1	180	60	1.84
2017-Gurley	0	0	0	1	1	180	60	2.12
2017-Nstar	0	0	1	1	2	384	158	3.41
2017-Tworth	7	3	1	2	13	3014	1836	4.26
2017-Westmar	0	0	0	1	1	180	60	2.24
2018-Blbgra	0	0	0	1	1	198	66	1.24
2018-Breeza	6	3	1	1	11	2502	1629	5.53
2018-Coonamble	0	0	0	1	1	198	66	1.55
2018-Gurley	0	0	1	0	1	210	105	2.23
2018-Moree	0	0	0	1	1	198	66	1.51
2018-Trangie	0	0	0	1	1	198	66	1.02
2018-Tworth	0	3	1	1	5	1074	481	2.24

Table 7. Number of test lines in each stage and year in the chickpea breeding program with data bands indicated as superscripts. The final columns give the number of lines in each contemporary group (CG19-CG14) for lines under consideration for selection in 2019.

	Number of test lines				Number of 2019 test lines					
Stage	2016	2017	2018	2019	CG19	CG18	CG17	CG16	CG15	CG14
S1	443^{D}	559^{C}	763^{B}	638^{A}	633	5	0	0	0	0
S2	113^{E}	100^{D}	146^{C}	176^{B}	0	176	0	0	0	0
S3	58^{F}	58^{E}	49^{D}	58^{C}	0	0	46	11	0	1

Table 8. MET dataset construction for 2019 selection decisions in chickpea breeding program. Number of lines missing years of data in datasets comprising bands A-C, A-D, A-E and A-F. Total number of lines for selection: 638, 176 and 58 (for S1 to S3).

	# years	Bands in dataset			
Stage	missing	A-C	A-D	A-E	A-F
S1	0	638	638	638	638
S2	0	176	176	176	176
S3	0	46	57	57	58
	1	11	0	1	0
	2	0	1	0	0
	3	1	0	0	0

Table 9. Summary of environments in the chickpea MET dataset: number of trials for each stage of testing (S1, S2, S3) and total number of trials. Number of plots (nplot) and lines (nline); mean yield (t/ha).

	Number of trials						Mean
Environment	S1	S2	S3	Total	nplot	nline	yield
2016-Balaklava	1	0	0	1	504	440	1.78
2016-Horsham	1	0	0	1	504	436	2.17
2016-Melton	1	0	0	1	504	450	2.17
2016-Yenda	1	0	0	1	504	360	2.59
2017-Balaklava	1	1	0	2	744	510	0.97
2017-Curyo	0	1	0	1	216	108	1.78
2017-Horsham	1	1	0	2	756	508	2.22
2017-Melton	1	1	0	2	756	512	1.36
2017-Yenda	1	1	0	2	924	648	1.18
2017-York	1	1	0	2	480	312	1.71
2018-Ardlethan	1	1	1	3	1356	972	0.66
2018-Balaklava	1	1	1	3	1032	676	0.44
2018-Curyo	0	1	1	2	492	210	0.74
2018-Horsham	1	1	1	3	948	599	2.02
2018-Melton	1	1	1	3	1032	676	0.49
2018-Mingenew	1	1	1	3	876	522	1.42
2018-Northampton	0	0	1	1	192	65	1.62
2018-Wagga Wagga	0	0	1	1	192	65	1.31
2019-Ardlethan	1	1	1	3	1260	858	0.35
2019-Curyo	0	1	1	2	444	180	1.63
2019-Dalwallinu	0	1	1	2	456	182	0.23
2019-Horsham	1	1	1	3	984	594	1.26
2019-Melton	1	1	1	3	816	414	0.59
2019-Mingenew	1	1	1	3	948	570	0.99
2019-Narrabri	1	0	0	1	780	642	2.11
2019-Pinery	1	1	1	3	756	368	0.78
2019-Wagga Wagga	0	0	1	1	240	81	0.58
2019-Yenda	0	0	1	1	240	80	1.38

Appendices

A CODE FOR COMPUTING A-VALUES

We consider S4 selection decisions for the final MET dataset for the durum motivating example. The data-frame is "final.df" and the key fields in this data-frame are "Environment" (factor with 30 levels); Gkeep (factor with 6,951 levels corresponding to the lines grown in the trials) and "Year" (factor with 6 levels). The inverse of the NRM is stored in sparse form in the object "durum.giv" and corresponds to 7,628 lines (lines grown in the trials and their ancestors). The mean of the diagonal elements of the NRM is $\bar{a} = 1.9$. The steps in ASReml-R (Butler et al., 2009) are as follows:

1. run the model with start.values=T to be able to set the values of the variance parameters

```
sv <- asreml(Yield ~ Environment, random =~ vm(Gkeep,durum.giv) +
idv(Year):vm(Gkeep,durum.giv) + idv(Environment):vm(Gkeep,durum.giv) +
ide(Gkeep) + Year:ide(Gkeep) + Environment:ide(Gkeep),
residual=~idv(units), data=final.df, start.values=TRUE,
na.action = na.method(x='include'))
```

2. set the variance parameter values as given in Table 3. These are ordered as additive V, VY and VE followed by non-additive V, VY and VE, then Error. Also set constraints so the parameters are fixed at these values in the next run of the model

```
gam.all <- sv$vparameters.table
gam.all$Value <- c(0.15,0.09,0.40,0.07,0.04,0.19,1,1)
gam.all$Constraint <- 'F'
```

3. run the model with the (fixed) pre-specified variance parameter values

```
temp.asr <- asreml(Yield ~ Environment, random =~ vm(Gkeep,durum.giv) +
idv(Year):vm(Gkeep,durum.giv) + idv(Environment):vm(Gkeep,durum.giv) +
ide(Gkeep) + Year:ide(Gkeep) + Environment:ide(Gkeep),
residual=~idv(units), data=final.df, R.param=gam.all, G.param=gam.all,
na.action = na.method(x='include'), maxit=1)
```

4. use the ASReml-R predict function to predict the total (additive + non-additive) line effects for the test lines in S4 in 2018 (there are 56 of these with names stored in the vector gs4.2018). Then obtain the A-value as the average pairwise variance (square of average sed)

```
temp.pvs <- predict(temp.asr, classify='Gkeep', maxit=1,
only=c("vm(Gkeep, durum.giv)","ide(Gkeep)"), levels=gs4.2018)
Avalue <- temp.pvs$avsed^2
```