Statistical modeling of phenotypic, pedigree and genomic information for improved genetic evaluation in modern plant breeding

A case study with sorghum



Julio G. Velazco

Statistical modeling of phenotypic, pedigree and genomic information for improved genetic evaluation in modern plant breeding:

A case study with sorghum

Julio G. Velazco

Thesis committee

Promotor
Prof. Dr F.A. van Eeuwijk
Professor of Applied Statistics
Wageningen University & Research

Co-promotor
Dr M. Malosetti
Researcher, Mathematical and Statistical Methods Group
Wageningen University & Research

Other members
Prof. Dr D. de Ridder, Wageningen University & Research
Prof. Dr H.P. Piepho, University of Hohenheim, Germany
Prof. Dr C.C. Schön, Technical University of Munich, Germany
Dr Y.C.J. Wientjes, Wageningen University & Research

This research was conducted under the auspices of the Graduate School for Production Ecology & Resource Conservation

Statistical modeling of phenotypic, pedigree and genomic information for improved genetic evaluation in modern plant breeding:

a case study with sorghum

Julio G. Velazco

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board,
to be defended in public
on Tuesday 17 March 2020
at 4 p.m. in the Aula

Julio G. Velazco
Statistical modeling of phenotypic, pedigree and genomic information for improved genetic evaluation in modern plant breeding: a case study with sorghum, 128 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2020)
With references, with summary in English

ISBN: 978-94-6395-279-8
DOI-link: https://doi.org/10.18174/511730

Contents

1 General Introduction 1

2 Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model[1] 11

3 Combining pedigree and genomic information to improve prediction quality: an example in sorghum[2] 39

4 Genomic prediction of grain yield and drought-adaptation capacity in sorghum is enhanced by multi-trait analysis[3] 57

5 Genomic prediction for broad and specific adaptation in sorghum accommodating information on individual marker effects 75

6 General Discussion 95

List of References 109

Summary 123

Acknowledgements 125

[1] Velazco J.G., Rodríguez-Álvarez M.X., Boer M.P., Jordan D.R., Eilers P.H.C. Malosetti M., van Eeuwijk F.A. (2017) *Theoretical and Applied Genetics* 130:1375–1392

[2] Velazco J.G., Malosetti M., Hunt C.H., Mace E.S., Jordan D.R., van Eeuwijk F.A. (2019) *Theoretical and Applied Genetics* 132(7):2055–2067

[3] Velazco J.G., Jordan D.R., Mace E.S., Hunt C.H., Malosetti M., van Eeuwijk F.A. (2019) *Frontiers in Plant Science* 10:997

Chapter 1

General Introduction

Global climate change and food insecurity are major concerns of the 21st century. Agricultural production should increase by 60–110% to meet the projected food demands of the word population by 2050 (McGuire 2012). However, the rates of global crop production are still far below the mentioned requirements and most studies predict a future decline in grain yield of major crops due to climate change (Ray et al. 2013; Wiltshire et al. 2013). Rainfed farming systems are drastically affected by climatic conditions, with water scarcity and increasing temperature being the most important limiting factors for crop productivity and, ultimately, for food security worldwide (Daryanto et al. 2013). Efforts to ensure food supply will require accelerating the development of climate resilient crop varieties. This is particularly necessary for crops that provide staple food grain in developing countries and semi-arid regions of the world.

Plant breeding can play a crucial role in enhancing crop productivity and adaptation to climate change. The main goal of breeding programs is to efficiently identify and select the best-performing genotypes as potential cultivars or as parental material to improve crop performance in future generations (Falconer and Mackay 1996; Bernardo 2010). For this, new selection techniques based on modern approaches to quantitative genetics have to be adopted by breeding programs in order to accelerate genetic progress. Advances in high-throughput genotyping technologies and the increasing cost-effective access to high-density genomic data have facilitated the adoption of a novel form of marker-assisted selection known as genomic selection (GS). This genetic evaluation method has already revolutionized animal breeding over the past decade and is gaining momentum in crop breeding. In GS, phenotypic and genome-wide marker data from a reference (or training) population is used to predict genetic merit of selection candidates that have only been genotyped but not phenotyped (Meuwissen et al. 2001). As a result, selection efficiency can potentially increase, reducing phenotyping costs and generation interval. Moreover, additional opportunities for GS in crops are provided by current developments in high throughput phenotyping technologies. The success in the incorporation of genomics as breeding tool depends on an appropriate statistical analysis of the phenotypic and genetic data generated in crop breeding programs.

1.1. Phenotypic analysis in plant breeding

Adequate phenotyping has been the basis of conventional plant breeding and is also an essential element in the implementation of GS. Crop breeding programs rely on the evaluation of phenotypic performance data typically obtained from series of field experiments known as multi-environment trials (MET). These trials are conducted to assess test genotypes and predict their performances across a target population of environments (TPE), which is assumed to be representative of future production environments where the newly bred varieties will be grown (Comstock 1977). Most agronomic traits with economic importance,

such as grain yield, are quantitative and complex, meaning that phenotypes are determined by numerous genes with small effects and highly influenced by environmental factors. This originates two potential challenges that can complicate selection decisions in plant breeding. These challenges are: spatial variation within trials and differential genotype performance across trials, namely genotype-by-environment interaction (G×E). Both phenomena must be accounted for in MET analysis for a proper assessment of genotypic values.

1.1.1. Spatial analysis of field trials

Environmental or non-genetic effects are firstly manifested at individual-trial level as a consequence of variable growing conditions typically caused by soil heterogeneity and agronomic practices. Ignoring these sources of spatial variability in phenotypic analysis can introduce bias in parameter estimation and prediction of genetic merit. For this reason, adjusting phenotypes for spatial trends has been largely recognized as an important prerequisite for an appropriate assessment of genotypic performance in plant breeding experiments. The classical approach to minimize the impact of field heterogeneity has been based on blocking and randomization by means of using designed experiments. More advanced experimental designs have been specially developed for plant breeding trials, where many genotypes, usually hundreds, are evaluated with low levels of replication (John and Williams 1995; Cullis et al. 2006; Williams et al. 2014). However, relying only on design factors may not be sufficient to capture all sources of environmental effects. Several methods that explicitly model spatial trends *a posteriori* have been proposed to improve the efficiency of field trial evaluations (e.g., Cullis and Gleeson 1991; Zimmerman and Harville 1991; Piepho et al. 2008b). Mixed model methodology has become very popular for the analysis of MET data in plant breeding, mainly due to its flexibility, computational efficiency and ability to easily handle unbalanced data (Smith et al. 2005; Piepho et al. 2008a). Accordingly, spatial analysis based on mixed models has been increasingly adopted as a natural extension. Under this approach, Gilmour et al. (1997) developed an influential and sophisticated spatial method for the analysis of crop breeding trials. They proposed to identify different components of field variation and to account for each one of these by including specific terms in the spatial model. The process of model building requires several steps and is guided by diagnostic graphics of residuals and by comparing the goodness of fit of alternative candidate models using different spatial terms. The approach of Gilmour et al. (1997) and simplified versions of it have been successfully used and reported by several studies (e.g., Qiao et al. 2004; Dutkowski et al. 2006). However, its adoption as a routine method of spatial analysis in series of plant breeding trials may be still hindered by the intensive intervention and specialist skills required. Efficient spatial methods that are easy to understand and simple to implement in routine analysis of plant breeding trials are still required.

1.1.2. G×E analysis

Besides spatial variation within trials, G×E plays a prominent role in determining genotypic performance across trials. The presence of G×E implies that genotypes respond differently to the environmental conditions of a TPE. The effectiveness of selection strategies under these circumstances will depend on the characteristics of G×E patterns. The breeder may attempt to exploit G×E by selecting for genotypes with specific adaptation to particular environments within the TPE. This strategy is likely to be relevant only in presence of repeatable G×E effects, which are typically associated with predictable—or fixed—environmental factors (e.g. soil type). Conversely, if the main divers of G×E patterns are unpredictable environmental components, such as year-to-year climatic variation, the most feasible approach is to select for broadly-adapted genotypes with superior overall performance across the TPE, minimizing in this way the impact of G×E (Atlin et al. 2011).

In order to reduce uncertainty in the identification of genotypes with broad or specific adaptation, the use of efficient and flexible methods for G×E analysis is required. For this task, mixed model analysis is currently the method of choice in plant breeding since it provides a flexible framework to account for differences in the genetic variances across environments as well as the lack of perfect genetic correlation among environments, which are the two sources for G×E. Different forms of G×E can be fitted under the mixed model approach, from those analogous to variance components models (Talbot 1984; Cullis et al. 1996) to more complex alternatives including unstructured and multiplicative mixed models (Piepho 1997; Smith et al. 2001b). Within this framework, explicit environmental information can also be incorporated as covariates to study G×E and to predict genotypic performance (van Eeuwijk et al. 1996; Denis et al. 1997).

1.2. Genomic selection in plant breeding

Plant breeders have relied historically on phenotypic selection, using trait phenotypes as proxies for the genetic merits of selection candidates. This started to change in the last decades with the advances in molecular genetics techniques, which allowed the discovery of several polymorphic markers in the genome. The availability of molecular markers opened the possibility to combine phenotypes and DNA information for selection of superior genotypes through what is generally known as marker-assisted selection (MAS). Initial attempts to implement MAS in plant breeding were based on quantitative trait loci (QTL) mapping and, more recently, on genome-wide association studies (GWAS). Both methods consist in identifying and selecting markers judged to be linked to causative genes or QTLs according to significance tests of their estimated effects. The subset of identified candidate markers is then used to assist prediction of genotypic performance for the target trait. Conventional MAS based on linkage studies has been more suitably implemented for traits

controlled by a small number of major genes (Dekkers and Hospital 2002; Xu and Crouch 2008). However, the gains in genetic progress have been much lower than originally expected for complex quantitative traits of economic importance, which are mostly determined by numerous small-effect genes, with some occasionally having large effects (Bernardo 2008; Heffner et al. 2009). The main conceptual challenge to conventional MAS resides in that QTLs are actually unknown. As a result, the process of finding out which markers are linked to QTLs induces errors, which are generally magnified in analysis of highly polygenic traits using many markers. This is because, in QTL mapping and GWAS, loci with small effects that do not exceed some significance threshold are implicitly set to have zero effect. However, a small effect does not necessarily mean lack of effect; therefore, some useful information may be unfairly discarded due to lack of power to detect small-effect QTLs. At the same time, a proportion of the genetic variance that could be explained by the removed markers is likely to be wrongly attributed to the few loci that passed the formal statistical test, overestimating their effects (Melchinger et al. 1998; Schön et al. 2004). The resulting lack of power and bias may decrease prediction accuracy.

These limitations of conventional MAS were overcome by adopting an innovative approach to MAS, known as genome-wide MAS or simply GS (Meuwissen et al. 2001). The core strategy of GS is to avoid pre-selection of markers based on detection thresholds and to assume that all markers are potentially linked to QTLs controlling the trait of interest. Hence, markers with small effect will be included, while no bias in estimated QTL effects will be induced by cause of the detection process. An additional conceptual difference is that the goal of GS is to predict genotypic or breeding values (BVs) of selection candidates, but without the explicit intention of identifying QTLs or drawing inferences regarding their biological nature. Finally, it is important to note that the GS approach requires many markers covering all the genome in order to work effectively. In that sense, the transition from conventional MAS to GS has been facilitated in practice by recent developments in genotyping technologies, particularly Single Nucleotide Polymorphism (SNP) genotyping, which have made possible the availability of affordable dense marker assays.

1.2.1. Genomic prediction methods

As mentioned above, GS is based on using all markers simultaneously to derive genomic predictions (GP) of unphenotyped genotypes. This creates a statistical challenge for the implementation of GP models since the number of parameters that must be estimated (i.e., the marker effects) is typically far larger than the observations used for estimation (i.e., the phenotyped individuals). The most common approach to address this challenge is to fit marker effects as random and apply some type of regularization. Several GP methods have been proposed for this, including a range of Bayesian regression models varying in the

particular prior distribution assumed for the marker effects (Gianola 2013) as well as non-parametric and semi-parametric approaches (de los Campos et al. 2010; Gonzalez-Recio et al. 2014). Simulation studies have shown that Bayesian methods are advantageous when the trait is assumed to be influenced by few QTLs with large effects (e.g. Meuwissen et al. 2001; Habier et al. 2007; Daetwyler et al. 2010). However, differences in predictive performance among GP methods tend to disappear when the target trait is highly polygenic (e.g., Wimmer et al. 2013; Wang et al. 2015). Under this scenario, a GP method based on *best linear unbiased prediction* (BLUP; Goldberger 1962), known as genomic BLUP or G-BLUP (VanRaden 2008), has been proven to perform similar (e.g. Yang et al., 2010, Clark et al., 2011) or even better than Bayesian approaches (Daetwyler et al. 2010; Heslot et al. 2013). G-BLUP makes use of the classic Mixed Model Equations (Henderson 1973; 1975) to obtain direct predictions of genetic merit by relying on a genetic similarity matrix computed from genome-wide markers. This genomic relationship matrix (**G**) is intended to approximate the additive-genetic resemblance between individuals at QTLs.

In theory, accurate estimates of genetic similarities, and therefore of predictions, can be obtained from genomic information if markers and unknown QTLs are in complete linkage disequilibrium (LD) (Habier et al. 2009). However, in the most frequent situation of incomplete LD, using only relationships from **G** may result in loss of some proportion of the genetic variance explained by the model. When training genotypes and selection candidates are related through pedigree, markers can capture genetic relationships as well as co-segregation of alleles within families regardless of populational LD (Habier et al. 2013). These sources of genetic information are also accounted for by traditional additive-genetic relationships derived from pedigree data. Consequently, incorporating information from pedigree into genomic models may be a useful strategy to assist predictions in structured breeding populations by simultaneously accounting for familial relationships not fully captured by markers. This alternative has not yet been explored extensively in plan breeding applications.

The implicit assumption of G-BLUP is that SNP effects are normally distributed with a constant variance and equal contribution to the trait variation, which is consistent with the infinitesimal model of quantitative genetics (Fisher 1918). The mentioned properties of G-BLUP make it more robust statistically and faster to compute when compared to other GP methods, resulting in a more attractive alternative for routine implementation (Gianola et al. 2014). Moreover, the mixed model formulation enables a straightforward extension of G-BLUP to the multivariate context. One important application of multivariate G-BLUP in plant breeding is to combine information of several correlated traits in a joint multi-trait analysis for increasing the accuracy of GP. In addition, G×E interactions can be explicitly modeled within the G-BLUP framework to improve predictions in multi-environment evaluations. However, a potential limitation of G-BLUP is that assuming equal variance for all marker effects may be

simplistic biologically since, in reality, loci are expected to contribute differently to the genetic control of a trait. An option to account for heterogeneous locus-specific variances is to use a **G** matrix that incorporates weights reflecting the genetic variance captured by each SNP (Zang et al. 2010; Wang et al. 2012; Su et al. 2014). This approach is able to accommodate deviations from the infinitesimal model, as with Bayesian GP methods, while still maintaining the simplicity in implementation of BLUP-based models. The SNP weighting approach has not been studied empirically so far in the context of plant breeding.

1.2.2. Evaluation of genomic predictions

Ideally, the potential success of a GP method for increasing genetic gain will depend on how well predicted genotypic values approximate the true genotypic values. However, in empirical studies using real data, the true genotypic values are unknown. Consequently, the performance of the GP methods is evaluated in practice by comparing predicted values to the realized values, with the latter being typically represented by raw observations or pre-adjusted phenotypes.

Cross-validation technique is usually used in GS studies because it provides a robust and nonparametric method for assessing and comparing GP models. In this method, the data are split into a training data set (TS) and a validation data set (VS). The TS, including genotypes and phenotypes, is used to estimate parameters of the GP model. These parameters are then used to compute predicted values for individuals in the VS based on genotype data only. The most common measure used to summarize the performance GP models is the Pearson's correlation, which indicates the strength of the linear association among predicted and realized values. However, other evaluation criteria can provide additional information on predictive performance, including measures of bias and accuracy (Vitezica et al. 2011; González-Recio et al. 2014).

1.3. Sorghum as a strategic global crop

One of the important and viable strategies to mitigate the impact of climate change is to develop superior genetic resources in crops that are inherently adapted to harsh growing conditions. Because of its drought-tolerance ability, sorghum [*Sorghum bicolor* (L.) Moench] is particularly important as staple cereal crop in semi-arid regions of the world. More than 70% of the world's total production of sorghum comes from small-scale farmers in drier regions of Asia and Africa where the crop is grown with limited input of water and nutrients, making it one of the major food sources in drought-prone environments (Talwar et al. 2013). Sorghum is the world's 5th most important cereal crop and is also produced for feeding and bioenergy purposes all over the world. It expanded partly as an "emergency crop" in times

when drought and erratic rainfall make maize production difficult or not feasible (Paterson 2008). A distinctive feature of sorghum that contributes to its ability to withstand drought stress is known as stay-green character. This trait, manifested as delayed leaf senescence, is the ability of a plant to retain active green tissue during grain-filling stage under water-limiting conditions, improving yield. The wide agro-climatic adaptation of sorghum enabled its expansion from West Africa into temperate regions of China, the Americas and Australia. Important producer countries include the United States and Australia, where grain sorghum is primarily grown for livestock feed. It is also used as a feedstock for producing ethanol, with sorghum being the second most important grain-based bioethanol crop after maize in the US (Dahlberg et al. 2011). Finally, there is also a growing interest in Europe for production of biomass-type sorghum as substitute of maize for biogas plants (Shiringani et al. 2010; Cai et al. 2013).

Sorghum breeding is primarily oriented to the development of hybrid varieties. Hybrids grown by farmers are typically produced and sold by the private industry. This activity is complemented by public breeding programs focused on the development of parental inbred lines, which are then used by seed companies for production of commercial hybrids. A crucial step in sorghum breeding is to assess the genetic value of parental lines on the base of their performance in single-cross combinations with tester genotypes. This progeny testing method, known as test-cross evaluation, is used to estimate the general combining ability or breeding values of inbred parents. Selection of lines based on test-cross performance demands extensive phenotyping resources since breeding programs typically evaluate thousands of lines using several testers in multi-environment field trials. For this reason, the implementation of GP in sorghum breeding is a promising strategy for selection of new untested parental lines, reducing cost and time required in early stages of the hybrid breeding pipeline. Even though GS has been successfully implemented in other major cereals, such as wheat and maize (e.g., Albrecht et al 2014; Sukumaran et al. 2017), empirical studies investigating the efficiency and potential benefits of GS in sorghum are still very limited.

1.4. Aim and outline of this thesis

The overall purpose of this thesis was to assess different statistical modeling strategies using phenotypic, pedigree and genomic information for improving genetic evaluation of complex traits in sorghum. The studies were carried out within the mixed model framework using BLUP methodology by considering large experimental data sets from a public sorghum breeding program for parental lines in Australia.

Chapter 2 presents the evaluation of a novel method that uses two-dimensional smoothing to account for spatial variation in plant breeding field trials. The new approach was assessed in comparison with the conventional and most competitive spatial method by considering the

improvement in precision and the prediction of genotypic values. The study is illustrated with a series of large and partially replicated trials belonging to early-stage breeding testing.

In **Chapter 3**, a first empirical study is provided on the implementation of genomic models for across-environment prediction of parental breeding values in sorghum. The Chapter investigates the impact of combining pedigree and genomic information on the quality of prediction for productivity and adaptability traits. For this, kinship matrices using different combinations of pedigree and marker-based relationships were explored in order to optimize predictive results. BLUP models using a blended pedigree-genomic matrix were applied to an extensive dataset consisting of advanced testing of female parent lines based on multi-year and multi-location testcross evaluation.

Chapter 4 extends the study of Chapter 3 to the multi-trait genomic prediction context. This research evaluates the ability of multi-trait analysis to improve genomic predictions of grain yield and drought-adaptation capacity by exploiting information from correlated secondary traits. The optimal combination of traits was determined empirically by evaluating predictive performances of alternative multi-trait models in different prediction scenarios. In addition, the benefits of including genealogical information into multi-trait genomic models through a combined pedigree-genomic matrix are reported for the first time in plant breeding.

Chapter 5 deals with genomic prediction for broad and specific adaptation in a subdivided TPE. Specifically, grain yield performances in drought-stress and non-stress environments were considered as the breeding targets of adaptation. The effect of modeling G×E on predictive ability and bias of multi-environment genomic prediction was assessed by comparing different structures for the genetic covariance between environments. The Chapter also explores how accounting for heterogeneous variances of SNP effects can affect genomic prediction of grain yield. This was accomplished by using a BLUP method based on individual-SNP weighting.

Chapter 6 concludes presenting a general discussion. The main contributions of this thesis are summarized and discussed in a broader perspective, including practical implications and opportunities for further progress in sorghum and crop breeding.

Chapter 2

Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model

Velazco J.G., Rodríguez-Álvarez M.X., Boer M.P., Jordan D.R., Eilers P.H.C., Malosetti M., van Eeuwijk F.A.

Theor Appl Genet (2017) 130:1375–1392.

Abstract:
Adjustment for spatial trends in plant breeding field trials is essential for efficient evaluation and selection of genotypes. Current mixed model methods of spatial analysis are based on a multi-step modelling process where global and local trends are fitted after trying several candidate spatial models. This paper reports the application of a novel spatial method that accounts for all types of continuous field variation in a single modelling step by fitting a smooth surface. The method uses two-dimensional P-splines with anisotropic smoothing formulated in the mixed model framework, referred to as SpATS model. We applied this methodology to a series of large and partially replicated sorghum breeding trials. The new model was assessed in comparison with the more elaborate standard spatial models that use autoregressive correlation of residuals. The improvements in precision and the predictions of genotypic values produced by the SpATS model were equivalent to those obtained using the best fitting standard spatial models for each trial. One advantage of the approach with SpATS is that all patterns of spatial trend and genetic effects were modelled simultaneously by fitting a single model. Furthermore, we used the same flexible model to adequately adjust for field trends in every site and for different traits. This strategy reduces potential parameter identification problems and simplifies the model selection process. Therefore, the new method should be considered as an efficient and easy-to-use alternative for routine analyses of plant breeding trials.

2.1. Introduction

Efficient phenotypic and genomic selection schemes in plant breeding programs rely on accurate assessment of the phenotypic performance of genotypes in field experiments (Qiao et al. 2004; Lado et al. 2013; Bernal-Vasquez et al. 2014; Sarker and Singh 2015). Plant breeding trials usually involve a large number of test entries covering large areas where spatial variation is likely to be an obstacle to reliable prediction of genetic values. This is particularly challenging in early generation variety trials conditioned by the use of limited replication of genetic material.

A number of sophisticated experimental designs, such as those enabling the recovery of inter-block information (Yates 1940; Patterson et al. 1978; John and Williams 1995) or partially replicated designs (Cullis et al. 2006; Williams et al. 2014), have been developed to correct for part of the field trend. However, efficient approaches to account for more complex environmental variation require complementing experimental designs with appropriate models of analysis (Basford et al. 1996; Qiao et al. 2000; Smith et al. 2002). Several spatial methods have been suggested to improve the precision of phenotyping. The most commonly used spatial models consider the correlation between residuals from neighboring plots to adjust for local trend or small-scale variation. These spatial methods include nearest neighbor analyses (Bartlett 1978; Wilkinson et al. 1983), and mixed model analyses using first-order autoregressive (AR1) functions (Cullis and Gleeson 1991) or other spatial covariance structures (e.g. Zimmerman and Harville 1991; Piepho and Williams 2010). Polynomials have been also used on top of experimental design features to account for additive and non-additive trends along row and column directions (Edmondson 1993; Federer 1998). Fertility trends in early generation variety trials have been modelled by fitting one-dimensional cubic smoothing splines within blocks (Durbán et al. 2001). Durbán et al. (2003) applied semiparametric models for spatial analysis of field experiments and presented graphical and analytical model selection criteria.

Within the mixed model framework, Gilmour et al. (1997) proposed an elaborate procedure for spatial analysis of agricultural variety trials. Their approach starts by fitting a two-dimensional separable AR1 model by default to account for local trend. Eventually, extraneous variation resulting from trial management practices may be accommodated with additional model terms, while global trends reflecting large-scale variation across the field are modelled by one-dimensional polynomials or splines in the direction of rows and/or columns. The authors suggested a sequential model-fitting scheme to identify the most suitable spatial model. The procedure relies on graphical diagnostic tools and requires several modeling choices to be tried. Stefanova et al. (2009) extended this modeling process by including more formal diagnostics to facilitate model selection. However, the above mentioned approach is not without limitations. First and foremost, the proposed multi-step procedures may not be

attractive for routine analysis of large series of trials since it requires a high level of hands-on intervention. Furthermore, there exists a risk of over-fitting the spatial data when the number of candidate models involved in the model selection process increases. Finally, convergence failures due to parameter identification problems may occur when trying to fit different spatial terms simultaneously (Dutkowski et al. 2006; Müller et al. 2010; Piepho et al. 2015).

Multidimensional regression spline methods represent a flexible alternative to account for complex variation structures. They allow the modeling of smooth multidimensional (or interaction) surfaces (e.g. Ruppert et al. 2003; Currie et al. 2006; Wood 2006). Regression splines are efficient curve-fitting functions composed of polynomial pieces, generally quadratic or cubic, that are joined at points called "knots". An interesting method using splines is based on two-dimensional P-splines (2D P-splines) as proposed by Eilers and Marx (1996, 2003), and its formulation in the linear mixed model framework (Eilers 1999; Currie and Durbán 2002). P-splines combine regression splines and a roughness penalty, which is the key component. This penalization is tuned by one or more smoothing parameters that control the degree of smoothness of the fitted spatial surface to prevent over-fitting. The connection between P-splines and mixed models provides attractive advantages. It enables the use of efficient algorithms for inference and prediction. Furthermore, the optimal smoothing parameters are automatically estimated by restricted maximum likelihood (REML; Patterson and Thompson 1971) as ratios of variance components.

Some applications of 2D P-spline models have been reported for spatial analysis of field trials. Cappa and Cantet (2007) and Cappa et al. (2011, 2015) used these models within a Bayesian approach to account for global trends in forest genetic trials. These studies considered a single smoothing parameter that controls the smoothness of the spatial effects in the direction of both rows and columns, imposing isotropic smoothing. In agricultural experiments, Taye and Njuho (2008) proposed using P-splines in two dimensions to adjust for global trend and to model local variation with Papadakis and kriged covariates. The authors compared P-spline models assuming additive trends or interaction between trends and emphasized the importance of choosing between both model settings. A different approach to spatial analysis of field trials using 2D P-spline mixed models was recently proposed by Rodríguez-Álvarez et al. (2016a). They introduced a novel spatial model that adjusts for both global and local trends simultaneously. The authors called this model SpATS, an acronym for *Spatial Analysis of field Trials with Splines*. The new spatial method makes use of the P-spline ANOVA representation of the smooth surface according to Lee et al. (2013). The distinctive feature of the SpATS model is an attractive decomposition of the spatial surface into additive one-dimensional trends and two-dimensional interaction trends. Furthermore, the model assigns a different smoothing parameter to each spatial component, allowing for anisotropic smoothing. This parametrization enables a flexible modeling of the spatial surface, where each component has a straightforward interpretation.

For the present research, we considered a series of multi-environmental trials from a sorghum [*Sorghum bicolor* (L.) Moench] breeding program in eastern Australia. These trials belong to initial stages of evaluations, where a large number of breeding lines (approximately 1000) were tested in each experiment using partially replicated designs. Furthermore, studies regarding the implications of performing spatial analysis in sorghum genetic trials are limited in the literature. Consequently, this dataset serves to illustrate a situation when a flexible and efficient spatial analysis tool is specially required.

This paper reports an application of the SpATS mixed model to adjust for all types of field trend in early generation sorghum breeding trials. We use a one-step modeling approach to spatial analysis by fitting a general SpATS model to analyze the whole series of trials. This approach is assessed in comparison with more elaborate and trial-specific spatial models identified according to the method of Gilmour et al. (1997). Both methods are compared in terms of variance component estimates, the improvement of precision and correlation of predicted genotypic effects. The new spatial model has been fitted using a tailor-made R package (R Development Core Team, 2016) called SpATS (Rodríguez-Álvarez et al. 2016b), which is publicly available from CRAN (https://cran.r-project.org/package=SpATS)

2.2. Materials and methods

2.2.1. Data set

In this study, we used data from 21 sorghum breeding trials conducted at 12 different locations in eastern Australia between 2005 and 2008. The dataset is part of the public germplasm enhancement program managed by the University of Queensland and Queensland's Department of Agriculture and Fisheries. A total of 3947 backcross recombinant inbred lines (BC-RILs) were evaluated as male parents in test-cross hybrid combinations with a single female tester. The BC-RILs were derived from crosses between an elite inbred line and a range of exotic sorghum lines. Detailed descriptions of the breeding population used in this paper can be found in Jordan et al. (2011) and Mace et al. (2013). The set of trials is considered to represent the target population of environments in the Australian sorghum cropping region.

Each trial was laid out as a rectangular array using resolvable *p*-rep designs (Cullis et al. 2006). Table 1 summarizes information related to the individual trials, including the field layout and the number of genotypes per location. Plots were 5 m wide along rows by 1.5 or 2 m long down the columns, with two rows of plants in each plot. The *p*-rep designs consisted of 30% of the test-cross hybrids having two replicates (p = 30%) while the remaining 70% of the genotypes were unreplicated. Across all trials, a total of 10 commercial varieties were included as check entries with additional levels of replication. Allocation of the replicated test

genotypes was based on an optimality measure determined by the average pairwise prediction error variance and assuming a pre-specified spatial model (Cullis et al. 2006). The search algorithm is constrained so that the replicated hybrids occurred once in each half of the trial, which established two resolvable blocks in all the designs.

We illustrate the spatial analyses with two traits: grain yield (t/ha) and plant height (cm). Data were not available for grain yield at trial HER08 and for plant height at trial SPR08 (Table 1). The proportion of missing plots ranged between 3% and 29%.

Table 1 Description of experimental layout and mean values of grain yield (GY) and plant height (PH) for each trial in the sorghum breeding dataset

Trial	Year	Location	Rows	Columns	Plots	Genotypes	Mean GY (t/ha)	Mean PH (cm)
BIL05	2005	Biloela	55	28	1540	1136	4.10	104
DAB05	2005	Dalby Box	76	20	1520	1167	3.05	91
DYS05	2005	Dysart	77	20	1540	1079	1.31	95
HER05	2005	Hermitage	77	20	1540	1202	7.65	113
JIM05	2005	Jimbour	44	20	880	682	4.60	105
BIL06	2006	Biloela	81	20	1620	1060	6.70	115
CEP06	2006	Cecil Plains	48	30	1440	953	3.24	95
DAB06	2006	Dalby	62	20	1240	823	2.00	101
GON06	2006	Goondiwindi	72	20	1440	957	6.58	117
HER06	2006	Hermitage	74	20	1480	1075	8.75	109
BIL07	2007	Biloela	86	20	1720	998	2.78	104
CLE07	2007	Clermont	34	40	1360	768	3.03	112
DYS07	2007	Dysart	44	40	1760	938	2.82	113
HER07	2007	Hermitage	70	25	1750	1012	5.39	104
BIL08	2008	Biloela	80	20	1600	1010	4.33	123
DAB08	2008	Dalby Box	64	20	1280	947	6.67	133
DAL08	2008	Dalby	62	20	1240	903	6.48	132
HER08	2008	Hermitage	80	20	1600	1012	-	133
KIL08	2008	Kilcummin	75	20	1500	899	3.36	131
LIV08	2008	Liverpool Plains	66	20	1320	980	10.06	125
SPR08	2008	Springsure	46	20	920	753	3.87	-

2.2.2. The SpATS model

In this section we present a brief description of the SpATS model; for a thorough treatment of the model specifications we refer to the original study by Rodríguez-Álvarez et al. (2016a).

Consider that observations in each sorghum breeding trial were obtained from plots arranged as a rectangular grid, where plot positions are collected in vectors of row (***r***) and

column ($\boldsymbol{c}$) coordinates. Under the SpATS model, field trends are modelled by a smooth bivariate function of the spatial coordinates, $f(\boldsymbol{r}, \boldsymbol{c})$, represented by 2D P-splines. As said, this technique optimizes the fitted surface by penalizing or shrinking the spatial effects. The magnitude of the penalization over the fitted trend is determined by the smoothing parameters. These terms control the balance between smoothness of the fitted surface and fidelity to the spatial data. For instance, larger values of the smoothing parameters result in smoother spatial gradients, while smaller values produce rougher fitted trends. Following the approach of Rodríguez-Álvarez et al. (2016a), additional terms were included in the SpATS model to account for other sources of environmental variation and genotype effects in our sorghum breeding trials.

Thus, the SpATS mixed model for each trial is given by

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{Z}_s\boldsymbol{s} + \mathbf{Z}_u\boldsymbol{u} + \mathbf{Z}_g\boldsymbol{g} + \boldsymbol{e}, \tag{1}$$

where the vector $\boldsymbol{y}$ contains the phenotypic observations (grain yield or plant height) arrayed as rows within columns, $\boldsymbol{\beta}$ is a vector of fixed terms including the intercept, a check variety effect and a resolvable block effect, and $\mathbf{X}$ is the associated design matrix. The fixed (unpenalized) term $\mathbf{X}_s\boldsymbol{\beta}_s$ and the random (penalized) component $\mathbf{Z}_s\boldsymbol{s}$ form the mixed model expression of the smooth spatial surface, i.e., $f(\boldsymbol{r}, \boldsymbol{c}) = \mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{Z}_s\boldsymbol{s}$, where the vector of random spatial effects $\boldsymbol{s}$ has a covariance matrix $\mathbf{S}$. The vector $\boldsymbol{u}$ comprises the mutually independent sub-vectors of random row and column effects accounting for discontinuous field variation, with design matrix $\mathbf{Z}_u = [\mathbf{Z}_r | \mathbf{Z}_c]$ and covariance matrix $\mathbf{U} = \text{diag}(\sigma^2_r\mathbf{I}_r, \sigma^2_c\mathbf{I}_c)$. The vector $\boldsymbol{g}$ contains the random genotypic effects of test-cross hybrids and $\mathbf{Z}_g$ is the associated design matrix. We assumed independent genotypic variance, i.e., $\boldsymbol{g} \sim N(\mathbf{0}, \mathbf{G})$, with $\mathbf{G} = \sigma^2_g\mathbf{I}_g$. The vector $\boldsymbol{e}$ consist of spatially independent residuals with distribution $\boldsymbol{e} \sim N(\mathbf{0}, \sigma^2_e\mathbf{I})$. This term, also called *nugget*, represents the measurement error from each plot.

The SpATS model adopts the P-spline ANOVA (*PS*-ANOVA) formulation proposed by Lee et al. (2013) to represent the 2D P-splines in the mixed model framework. Detailed descriptions of the design matrices $\mathbf{X}_s$ and $\mathbf{Z}_s$ and the covariance matrix $\mathbf{S}$ under this formulation are given in Lee et al. (2013) and Rodríguez-Álvarez et al. (2016a). In this paper, we present the main result of the *PS*-ANOVA parameterization, which is the decomposition of the smooth surface into a sum of linear components and univariate and bivariate smooth functions, such that

$$f(\boldsymbol{r}, \boldsymbol{c}) = \underbrace{\beta_{s1}\boldsymbol{r} + \beta_{s2}\boldsymbol{c} + \beta_{s3}\boldsymbol{rc}}_{\mathbf{X}_s\boldsymbol{\beta}_s} + \underbrace{f_1(\boldsymbol{r}) + f_2(\boldsymbol{c}) + h_3(\boldsymbol{r})\boldsymbol{c} + \boldsymbol{r}h_4(\boldsymbol{c}) + f_5(\boldsymbol{r}, \boldsymbol{c})}_{\mathbf{Z}_s\boldsymbol{s}}, \tag{2}$$

where the spatial surface is represented by: linear trends across the row (β_{s1}) and column (β_{s2}) positions and a linear interaction trend (β_{s3}); two main smooth trends across rows [$f_1(\boldsymbol{r})$] and

columns [$f_2(\boldsymbol{c})$]; two linear-by-smooth interaction terms, where the slope of a linear trend along one covariate ($\boldsymbol{c}$ or $\boldsymbol{r}$) is allowed to vary smoothly as function of the other covariate [$h_3(\boldsymbol{r})$ or $h_4(\boldsymbol{c})$, respectively]; and $f_5(\boldsymbol{r}, \boldsymbol{c})$ is the *pure* smooth-by-smooth interaction between column and row trends.

Under this representation, the vector of random spatial effects $\boldsymbol{s}$ contains five mutually independent sub-vectors $\boldsymbol{s}_k$, with k = 1, ..., 5 referring to the additive and interaction random components in [2]. Then, the spatial covariance matrix $\mathbf{S}$ is a direct sum of matrices $\mathbf{S}_k$, that is $\mathbf{S} = \text{blockdiag}(\mathbf{S}_1, \ldots, \mathbf{S}_5)$, where each block $\mathbf{S}_k$ depends on a specific smoothing parameter λ_{s_k} (see Rodríguez-Álvarez et al. 2016a for details). Within the mixed model framework, each smoothing parameter is determined by REML as the ratio between the residual variance and the corresponding variance of spatial effects, i.e., $\lambda_{s_k} = \sigma_e^2/\sigma_{s_k}^2$. Therefore, the smoothness of the spatial surface is tuned by five distinct parameters, applying anisotropic smoothing. The parameterization provides the SpATS model with flexibility to account for both global trends and local variation in the field. Furthermore, the decomposition of $f(\boldsymbol{r}, \boldsymbol{c})$ enables a more explicit interpretation of the main patterns of spatial variation.

2.2.2.1. Implementation of the model

The SpATS model with anisotropic smoothing based on the *PS*-ANOVA approach by Lee et al. (2013) was fitted with the R package (R Development Core Team, 2016) SpATS (Rodríguez-Álvarez et al. 2016b), which is publicly available from CRAN (https://cran.r-project.org/package=SpATS). The spatial surface in model [1] was fitted using cubic B-spline bases and second-order penalties, which are commonly used settings in the P-spline framework. Across trials, we used 11 and 31 equally-spaced knots for the P-splines in the column and row directions, respectively. In this way, we set approximately one knot for every two rows or columns. Then, the spatial surface contains a total of 425 model parameters to be estimated. These quantities were chosen in order to provide enough flexibility to the spatial surface. Within the penalized smoothing context, the exact choice of the number of knots is not critical once a certain minimum number of knots is exceeded (Ruppert et al. 2003; Eilers et al. 2015). This number can be equal to the number of rows and columns, i.e., the number of data points in each dimension, or even more. The only limiting factor would be the computational time: the larger the number of knots, the larger the computational effort. It is important to remark that the use of a large number of knots provides flexibility, but in practice, the smoothing parameters are responsible for optimizing the fit to the data.

The estimation procedure implemented in the R-package SpATS provides REML-based variance components and computes the empirical best linear unbiased estimates (BLUEs) of fixed effects and the empirical best linear unbiased predictors (BLUPs) of random effects. An important by-product of the procedure is that, for each random effect of the model, an

associated *effective dimension* is computed. The practical implications of the latter concept are considered in the following sections.

2.2.2.2. The effective dimension of the fitted spatial surface

The effective dimension (ED) or effective number of parameters of the model is a central concept within the P-spline methodology. It is a measure of complexity of the model components and is mainly a function of the smoothing parameters (Eilers et al. 2015). The effective dimension of a model is computed as the trace of the *hat* matrix **H**. If we focus on the spatial part of the SpATS model [1], we have that

$$\tilde{f}(\boldsymbol{r}, \boldsymbol{c}) = \mathbf{X}_s\widehat{\boldsymbol{\beta}}_s + \mathbf{Z}_s\tilde{\mathbf{s}} = \mathbf{H}_\beta \boldsymbol{y} + \mathbf{H}_s \boldsymbol{y},$$

where $\mathbf{H}_\beta$ is hat matrix of the fixed component, with effective dimension $ED_\beta = \text{trace}(\mathbf{H}_\beta) = \text{rank}(\mathbf{H}_\beta)$, which is always a constant. More importantly, the total effective dimension of the random (penalized) component of the spatial surface is $ED_s = \text{trace}(\mathbf{H}_s)$, where $\mathbf{H}_s$ is known as the *smoother matrix*. In this context, the sum of the diagonal elements of $\mathbf{H}_s$ expresses the number of parameters effectively involved in the modelling of the spatial surface. From the *PS*-ANOVA decomposition used in the SpATS model, we have that $\mathbf{H}_s = \sum_{k=1}^{5} \mathbf{H}_{s_k}$, with $k = 1$, ..., 5 referring to the additive and interaction smooth components of the spatial trend as detailed in [2]. Thus, we can decompose ED_s as the sum of partial effective dimensions associated to each spatial component

$$ED_s = \text{trace}(\mathbf{H}_s) = \sum_{k=1}^{5} \text{trace}(\mathbf{H}_{s_k}) = \sum_{k=1}^{5} ED_{s_k}.$$

Specifically, when a smoothing parameter $\lambda_{s_k} = \sigma_e^2/\sigma_{s_k}^2 \rightarrow \infty$, then $ED_{s_k} \rightarrow 0$; while for a value of $\lambda_{s_k} = \sigma_e^2/\sigma_{s_k}^2 \rightarrow 0$, ED_{s_k} approaches the maximum value. The upper bound for ED_{s_k} is determined by the number of knots used to fit the smooth surface. Therefore, ED_{s_k} serves as a reverse indicator of the smoothness of the corresponding component, i.e., the higher the degree of smoothness (larger value of λ_{s_k}), the smaller the number of ED_{s_k} (see Rodríguez-Álvarez et al. 2016a for details).

Consequently, the total effective dimension ED_s can be interpreted as a measure of the magnitude of field variation, with larger values indicating more intense spatial patterns. Additionally, the partial effective dimensions ED_{s_k} are indicative of the relative importance of each spatial component in [2]. In this case, the magnitudes of specific ED_{s_k} will quantify the contribution of the main and interaction spatial trends to the fitted surface, reflecting the complexity of the spatial patterns.

2.2.2.3. Generalized heritability based on the genetic effective dimension

As previously mentioned, an effective dimension connected to each variance component of the SpATS model is computed. The effective dimension associated to the genotypic effects (ED_g) is particularly interesting for plant breeding. ED_g = trace($\mathbf{H}_g$) is a measure of the degree of shrinkage imposed on the genotypic effects, where $\mathbf{H}_g$ is the hat matrix for genotypes. In this case, $\mathbf{H}_g$ depends on the regularization parameter $\lambda_g = \sigma_e^2/\sigma_g^2$ and transforms the observations into predicted genotypic values, such that $\mathbf{H}_g\boldsymbol{y} = \mathbf{Z}_g\tilde{\boldsymbol{g}}$ (see Rodríguez-Álvarez et al. 2016a for details). Therefore, ED_g decreases as shrinkage of genotypic effects increases. Given the properties of the genetic effective dimension, Rodríguez-Álvarez et al. (2016a) proposed a novel expression of heritability:

$$H^2 = \frac{ED_g}{n_g - l}$$

where n_g is the number of genotypes and l is the number of zero eigenvalues of $\mathbf{H}_g$. The authors showed that this definition corresponds to the generalized heritability introduced by Oakey et al. (2006). Furthermore, in the specific situation when genotypic effects are assumed independent (i.e., ignoring pedigree/marker information), and by ignoring the zero eigenvalues, the following equivalence can be stablished:

$$H^2 = \frac{ED_g}{n_g} = 1 - \frac{\overline{PEV}}{\sigma_g^2},$$

where $\overline{PEV}$ stands for average prediction error variance of genotype BLUPs.

Note that the right hand term corresponds to the generalized heritability developed by Welham et al. (2010) and is also equivalent to the heritability given by Cullis et al. (2006). Given that our study does not incorporate a genetic relationship matrix, we can profit from the latter equivalence to perform a straightforward comparison between the heritability estimated by the SpATS model and that obtained from the standard mixed models.

2.2.3. Standard models

Under the standard mixed model framework, we started by fitting a *non-spatial* model. This baseline model included a random and independent genotypic effect for the test-cross hybrids, a fixed effect for check varieties, a fixed resolvable block effect accounting for the randomization design and the spatially independent error term $\boldsymbol{e} \sim N(\mathbf{0}, \sigma_e^2\mathbf{I})$. Then, the non-spatial model was extended by searching for the most appropriate spatial model for each case following the approach of Gilmour et al. (1997). The latter model is referred to as the *best standard spatial* (BSS) model.

The general representation of the BSS model can be formulated as

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{Z}_s\boldsymbol{s} + \mathbf{Z}_u\boldsymbol{u} + \mathbf{Z}_g\boldsymbol{g} + \boldsymbol{\xi} + \boldsymbol{e}, \quad (3)$$

where $\mathbf{X}\boldsymbol{\beta}$ contains the same fixed terms as the non-spatial model. The term $\mathbf{X}_s\boldsymbol{\beta}_s$, in this case, may include linear trends aligned with rows and/or columns to account for global variation, while $\mathbf{Z}_s\boldsymbol{s}$ contains the random part of one or two one-dimensional cubic smoothing splines indexed by row or column positions (see Verbyla et al. 1999 for details). This latter term accounts for non-linear additive global trends. $\mathbf{Z}_u\boldsymbol{u}$, $\mathbf{Z}_g\boldsymbol{g}$, and $\boldsymbol{e}$ are defined as in the SpATS model [1]. Finally, $\boldsymbol{\xi}$ is the vector of spatially correlated residuals modelling local trend, with distribution $\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{R})$. The matrix $\mathbf{R} = \sigma^2_{\xi}[\text{AR1}(\rho_c) \otimes \text{AR1}(\rho_r)]$ represents the Kronecker product of first-order autoregressive processes across columns and rows, respectively, and σ^2_{ξ} is the spatial residual variance.

Following Gilmour et al. (1997), the search for the BSS model was based on diagnostic graphics such as the sample variogram and related plots of residuals. Comparisons between candidate models with the same fixed effects were assessed by the REML likelihood ratio test (REML-LRT). Fixed spatial terms were included in the BSS model when judged significant according to Wald-F test. It is important to note that the BSS model for each trial and trait may represent a simplified version of the full model [3], where the reduced model results from omitting one or more superfluous spatial components.

The standard mixed models were fitted using the ASReml-R package (Butler et al. 2009).

2.2.4. Comparison of spatial methods

The SpATS model was compared with the non-spatial and the BSS models in terms of meaningful parameters for plant breeding application. The following estimates were considered for comparison:

- Genetic variance (σ^2_g) and spatially-independent residual variance (σ^2_e).
- Generalized heritability. Estimated following Rodríguez-Álvarez et al. (2016a) for the SpATS model and according to Cullis et al. (2006) for the standard models. These measures are interpreted as broad-sense heritability, which serves as a descriptive measure of the precision of a trial, i.e., of the ability to detect genotypic differences among test-cross hybrids.
- Pearson correlations of predicted genotypic values between environments. Given that genotype-by-environment interaction has the same effect on the magnitudes of these correlations for the three models, any increase in their values relative to the non-spatial model will indicate the improvement of precision caused by the spatial models

(Qiao et al. 2004; Müller et al. 2010). Only correlations between pairs of environments presenting at least 30 common genotypes were considered.

- Spearman rank correlations between predicted genotypic values from the different models in the same environment. Calculated to compare whether the rankings of genotypes obtained from SpATS and from the standard models differed.

2.3. Results

2.3.1. Spatial analysis with SpATS

We start with a detailed treatment of the spatial analysis using the SpATS model, illustrated with two contrasting trials regarding the intensity and structure of spatial variability. Table 2 presents the ED_s of the univariate and bivariate spatial smooth components (see equation [2]), and their relative contribution to the fitted surface for grain yield in trials DYS05 and DAB08. The magnitudes of the total ED_s indicate that the spatial variation was more intense in DYS05. This is reflected by the higher ED_s or fitted parameters required to model the underlying field trend (111.2 ED in DYS05 vs 2.1 ED in DAB08). According to the partial ED_s, DYS05 also exhibited a higher complexity in the structure of the spatial surface, where the smooth-by-smooth interaction between trends accounted for most of the field variation (87% of the total ED_s). In contrast, the environmental trend at DAB08 was smoother and less complex as it presented a lower total ED_s and was mostly captured by main smooth effects across row positions. The zero values of ED_s associated to the linear-by-smooth interactions in DAB08 indicate that these terms were not necessary to model the spatial surface.

Table 2 Spatial effective dimensions (ED_s) of the smooth surface components fitted by the SpATS model and its relative contribution (%) for grain yield in two example trials

Spatial smooth terms		DYS05		DAB08	
		ED_s	%	ED_s	%
Additive trends	$f_1(\boldsymbol{r})$	3.0	3	1.4	67
	$f_2(\boldsymbol{c})$	4.2	4	0.2	10
Interaction trends	$h_3(\boldsymbol{r})\boldsymbol{c}$	1.9	2	0.0	0
	$\boldsymbol{r}h_4(\boldsymbol{c})$	5.5	5	0.0	0
	$f_5(\boldsymbol{r}, \boldsymbol{c})$	96.6	87	0.5	24
	Total	111.2	100	2.1	100

Fig. 1 shows the graphical representations of the fitted spatial trend $f(\boldsymbol{r}, \boldsymbol{c})$ and the spatially independent residuals $\boldsymbol{e}$ for the two example trials, as obtained from the SpATS package. Note that the pictures of the spatial trend use a finer grid than that of the field plots; the P-splines makes their computation possible. The spatial surfaces display an irregular patchy pattern in

DYS05 and a rather smooth gradient across the field in DAB08. The shape of an evident patch of fertility present in DYS05 was best modelled by considering interactions between column and row trends, as indicated by the partial ED_s (Table 2). Likewise, the previous interpretation of the spatial trend based on the ED_s in DAB08 coincides with the plot of the fitted surface, which essentially exhibits a one-dimensional gradient across rows.

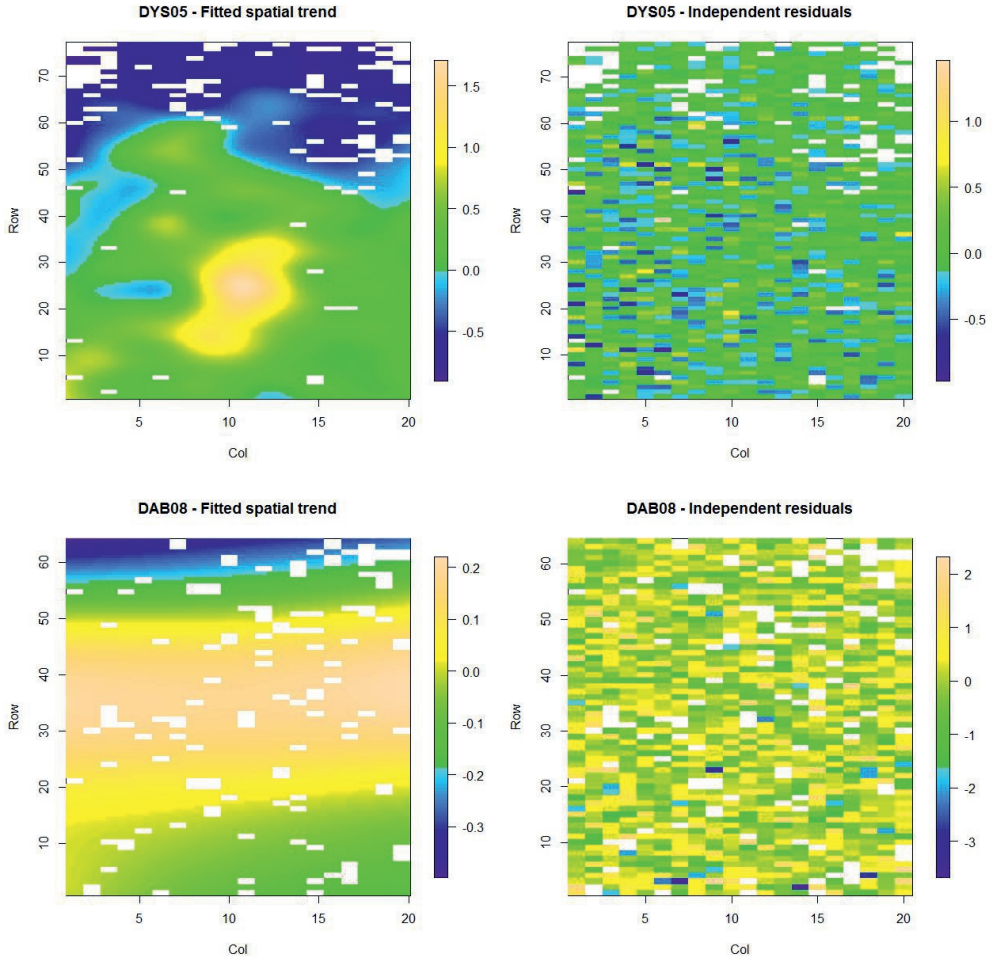


Fig. 1 Fitted spatial trend and spatially independent residuals from the SpATS model for grain yield in trials DYS05 (top) and DAB08 (bottom) plotted against row and column positions. Scales of grain yield variation expressed in t/ha

The inspection of the plots of residuals in Fig. 1 suggests that the spatial patterns have been effectively removed in both trials by the 2D P-spline surface; hence, these residuals could be

considered as true random noise. Other plots of residuals and formal tests could also be used to diagnose outliers, model assumptions or remaining spatial trends after fitting the spatial model. For the latter purpose, an interesting alternative is the variogram computed from the independent residuals, as proposed by Piepho and Williams (2010). This nugget-based variogram can also be obtained with the R-package SpATS. The ranges of grain yield variation (in t/ha) explained by the fitted trends reflect the magnitude of spatial effects in each trial. The comparison between the scales of spatial and residual site variations provides a clear idea of the relative importance of field trends in these trials. For instance, the range of yield variability due to spatial trends in DYS05 was of similar magnitude to that caused by the spatially independent error, while the amount of variation resulting from the latter term was about tenfold the spatial variability in DAB08 (Fig. 1). Again, the higher relevance of spatial trends for trial DYS05, was also indicated by the total ED_s presented in Table 2.

In Table 3, we specify the spatial terms of the BSS models for grain yield in DYS05 and DAB08. The connection between these results and those from the analysis with the SpATS model (Table 2) is not straightforward since the parameterization of both spatial models is different. Assuming that extraneous variations have been adjusted by both models, here we stress the differences in modelling global and local trends. For instance, according to the standard spatial analysis, in DYS05 there was only a main global trend in the direction of rows while the column and interaction trends detected by SpATS were apparently modelled as two-dimensional autocorrelated residuals by the BSS model. The main trend across row positions in DAB08 (see Table 2 and Fig. 1) seems to be modelled, under the standard approach, by a small autocorrelation across rows and by a value of ρ_c close to 1. The latter autocorrelation suggests that the trend across columns is actually confounded with the random row effects (Piepho and Williams 2010; Piepho et al. 2015). Finally, the ratios of spatial variance to residual variance (σ^2_ξ/σ^2_e) were 2.0 for DYS05 and 0.3 for DAB08, indicating a higher intensity of spatial variation in the former trial (Dutkowski et al. 2002; Zas 2006). The latter results coincide with the interpretation based on the total effective dimensions of the spatial surfaces given in Table 2.

Table 3 Spatial terms, estimates of autocorrelations and variance components for spatially dependent (σ^2_ξ) and independent residuals (σ^2_e) from the *best standard spatial* (BSS) models fitted to grain yield data in two example trials

Trial	BSS model[a]	ρ_r	ρ_c	σ^2_ξ	σ^2_e
DYS05	R+Spl(*r*)+AR1xAR1+*n*	0.87	0.67	0.103	0.064
DAB08	R+AR1xAR1+*n*	0.24	0.96	0.201	0.611

[a] R: random row effects; Spl(*r*): cubic smoothing spline indexed by row positions; AR1xAR1: correlated residual modelled as two-dimensional first-order autoregressive process; *n*: spatially independent residual (or *nugget* variance).

The effective dimensions associated to the fitted spatial trends (ED_s) for all trials and both traits are given in Fig. 2. For simplicity, the partial ED_s for the five smoothing terms of the SpATS model are grouped as: ED_s of the additive smooth trends and ED_s of the interactions between trends.

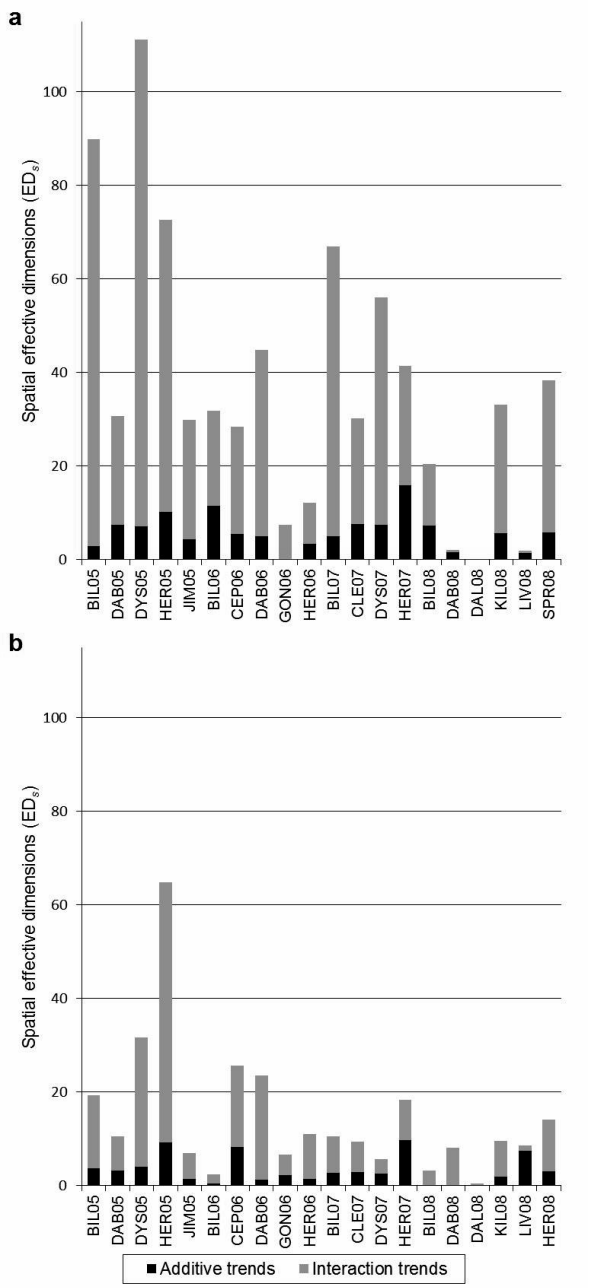


Fig. 2 Effective dimensions (ED_s) associated to additive and interaction trends of the spatial surfaces fitted by the SpATS model for grain yield (**a**) and plant height (**b**) in the sorghum breeding trials. Note: data were not available for grain yield at HER08 and for plant height at SPR08

The intensity of spatial variation and the complexity of the fitted surfaces were highly variable across sites and traits. For instance, the environmental trends for grain yield at DYS05 and BIL05 or HER05 for plant height present a large number of ED_s and a significant contribution of the trend interaction terms, indicating strong and complex patterns of field variation. Others cases, such as DAB08 for grain yield and LIV08 for yield and plant height, show lower total ED_s, reflecting smoother spatial surfaces that were mainly described by additive one-dimensional trends. In general, the intensity of spatial variation for grain yield was higher than for plant height, with median total ED_s of 31 and 10, respectively. In most instances, the smooth trend interactions represented the major components of the spatial surface. This is reflected by the median ED_s associated to interaction effects, which were 82% and 79% of the total ED_s for yield and plant height, respectively. The latter results highlight the importance of modelling interactions between row and column trends and reveals complex structures of field variation in the sorghum dataset.

2.3.2. Standard spatial analysis

A summary of the main features of the BSS models fitted to the sorghum dataset is reported in Table 4. Details of the BSS model identified in each of the 20 trials for both traits are presented in Table 5.

Table 4 Number of times the *best standard spatial* (BSS) models for the 20 trials included terms accounting for global and local trends, and median of estimated spatial parameters

	Grain yield	Plant height
Number of trials including:		
Global trend terms	15	11
Correlated residuals (AR1xAR1)	17	17
Nugget effect	17	14
Median of spatial parameters:		
ρ_r	0.82	0.87
ρ_c	0.73	0.69
Proportion (%) of correlated error[a]	52	25

[a] Relative to the sum of correlated and independent residual variances

The results in Table 4 show that most of the trials required terms accounting for global trends, local variation and nugget effect. Autocorrelations (ρ) along rows and columns were predominantly large and similar for both traits, as reflected by their median. Over 80% of the autocorrelation coefficients were larger than 0.60, indicating strong spatial variation that could be interpreted as a combination of large-scale gradients and patchy patterns according to the standard approach. When considering the models with nugget, the importance of the

spatial variance relative to the spatially independent residual variance was generally higher for grain yield. The predominance of random noise in plant height measurements indicates that this trait was less influenced by spatial effects in the field. This is consistent with the generally lower effective dimensions of the spatial surfaces estimated for plant height (see Fig. 2). Furthermore, there was a strong positive correlation between the σ^2_ξ/σ^2_e ratio and the total ED$_s$ across the whole dataset (r = 0.69).

Table 5 Details of the *best standard spatial* (BSS) models in each trial for grain yield (GY) and plant height (PH)

Trial	BSS model for GY[a]	BSS model for PH[a]
BIL05	Lin(*r*)+AR1xAR1+*n*	Lin(*r*)+AR1xAR1+*n*
DAB05	R+AR1xAR1+*n*	Lin(*c*)+AR1xAR1+*n*
DYS05	R+Spl(*r*)+AR1xAR1+*n*	Lin(*r*)+AR1xAR1+*n*
HER05	Lin(*r*)+Spl(*c*)+AR1xAR1+*n*	Spl(*r*)+Lin(*c*)+AR1xAR1+*n*
JIM05	C+AR1xAR1+*n*	R+AR1xAR1+*n*
BIL06	Spl(*c*)+AR1xAR1+*n*	C+AR1xAR1
CEP06	R+AR1xAR1+*n*	Spl(*c*)+AR1xAR1+*n*
DAB06	C+Spl(*r*)+AR1xAR1+*n*	AR1xAR1+*n*
GON06	Spl(*c*)+AR1+*n*	R+C+AR1
HER06	R+Spl(*r*)+Lin(*c*)	R+C+Lin(*c*)+AR1xAR1+*n*
BIL07	Lin(*c*)+AR1xAR1+*n*	R+Spl(*c*)
CLE07	R+Lin(*r*)+AR1xAR1+*n*	Lin(*c*)+AR1xAR1+*n*
DYS07	Spl(*c*)+AR1xAR1+*n*	Lin(*c*)+AR1xAR1+*n*
HER07	R+Spl(*r*)+Spl(*c*)+AR1xAR1+*n*	Lin(*r*)+AR1xAR1+*n*
BIL08	R+C+Spl(*c*)	C
DAB08	R+AR1xAR1+*n*	Lin(*c*)+AR1+*n*
DAL08	R+C+Lin(*c*)	C
HER08	–	AR1
KIL08	R+Lin(*r*)+AR1xAR1+*n*	AR1xAR1+*n*
LIV08	R+C+AR1xAR1+*n*	AR1xAR1+*n*
SPR08	Spl(*c*)+AR1xAR1+*n*	–

[a] Spl(·): cubic smoothing spline indexed by row (*r*) or column (*c*) positions; Lin(·): linear regression on row (*r*) or column (*c*) positions. R: random row effects; C: random column effects; AR1 and AR1xAR1: correlated residuals modelled as one- and two-dimensional first-order autoregressive process, respectively; *n*: spatially independent residual (nugget effect). Note that all the models included a fixed block effect

2.3.3. Comparison of SpATS and the standard method

The estimates of trial genetic variability from SpATS and the BSS models were generally similar for both traits (Fig. 3). Small differences were evident for grain yield at some environments, where the estimates increased or decreased from one model to the other

without a clear tendency. More marked discrepancies were observed between the genetic variances from the non-spatial model and those from both spatial models for grain yield (not shown). This suggests that ignoring the adjustment for spatial trends in yield data can lead to either overestimating or underestimating the genetic variability.

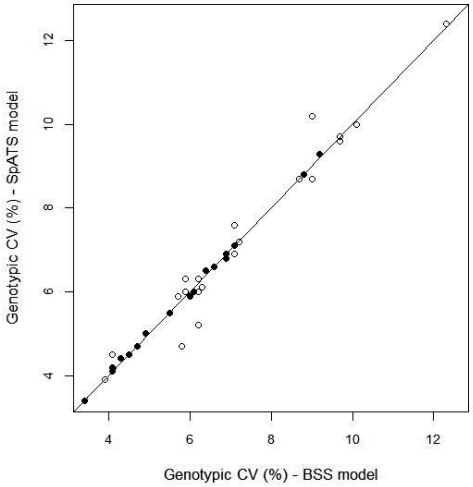


Fig. 3 Comparison of genotypic variability estimated by the BSS and the SpATS models, expressed as coefficient of variation (CV), for grain yield (○) and plant height (●). The diagonal line indicates identical values

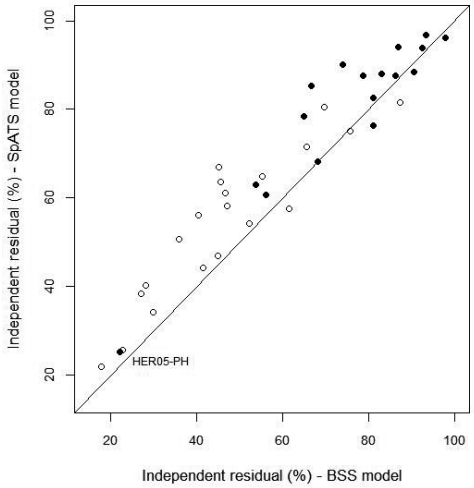


Fig. 4 Comparison of spatially independent residual variance (σ_e^2) from the BSS and the SpATS models, expressed as percentage (%) of the residual variance in the non-spatial model, for grain yield (○) and plant height (●). The diagonal line indicates identical values. The labelled data point corresponding to plant height at trial HER05 (HER05-PH) is mentioned in the text

The SpATS model and the BSS models reduced the spatially independent residual variance compared with error variance of the non-spatial model in both traits (Fig. 4). In general, the relative decreases in σ_e^2 were larger for grain yield, with the spatial models achieving a mean reduction by 49% for grain yield and by 22% for plant height. These reductions reflect the ability of both methodologies to account for field variation not adjusted by the randomization-based model. Exceptionally, the adjustment of spatial trend for plant height caused a large decrease in σ_e^2 at trial HER05. Note that field trend in this case was particularly important, presenting the highest total ED_s for plant height and a major contribution of interaction effects (see Fig. 2). In general, the BSS models estimated smaller values of σ_e^2 compared to the SpATS model. The spatially independent component from SpATS and the BSS models represented, on average, 66% and 60% of the residual variance from the non-spatial model, respectively.

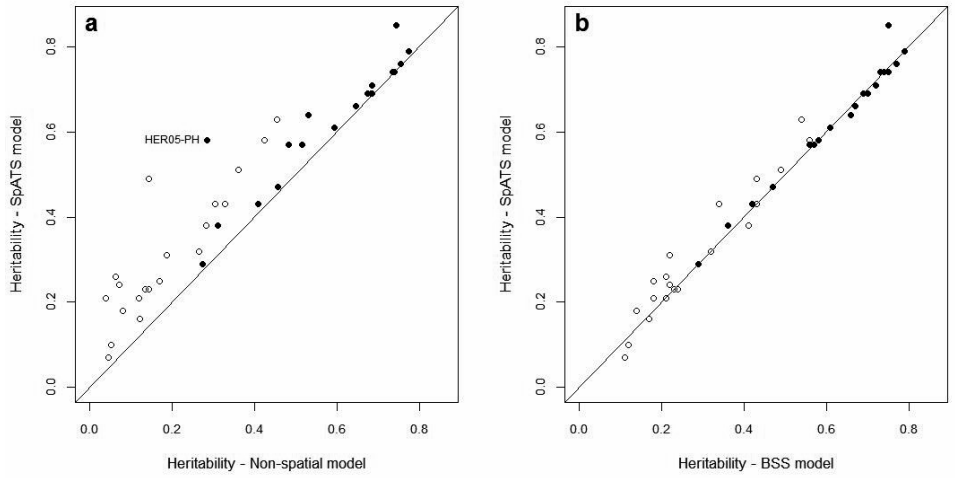


Fig. 5 Comparison between estimates of heritability from the non-spatial and SpATS models (**a**), and from SpATS and the BSS models (**b**) for grain yield (○) and plant height (●). The diagonal lines indicate identical values. The labelled data point corresponding to plant height at trial HER05 (HER05-PH) is mentioned in the text

Figure 5a shows the changes in the estimates of trial heritability from the non-spatial model to the SpATS model. The spatial method increased the heritability in most instances, with levels of improvement in precision being generally higher for grain yield. Not surprisingly, a remarkable increase in heritability was also achieved for plant height in HER05 after fitting trends with SpATS. Trial heritabilities estimated by both spatial methods were very consistent for plant height (Fig. 5b). However, more variation in the estimates was observed for grain yield, where similar or slightly higher heritabilities were obtained with the

SpATS model in most trials. Finally, notice that heritabilities were, in general, lower for grain yield, which was the trait affected by stronger spatial variation (as inferred from the total ED_s in Fig. 2).

The Pearson correlations of genotype BLUPs between environments obtained from the two spatial methods were, on average, slightly higher than those obtained from the non-spatial model in both traits (Fig. 6). The mean correlations for grain yield increased from 0.04 to 0.10 and 0.09 after applying the BSS models and SpATS, respectively (Fig. 6a). For plant height, both spatial models caused a mean increase of 0.05 in the correlations, changing from 0.46 to 0.51 (Fig. 6b). At the same time, the spatial methods reduced the variation of estimated correlations for the latter trait. The higher mean correlations between environments in plant height reflect a lower influence of genotype-environment interaction.

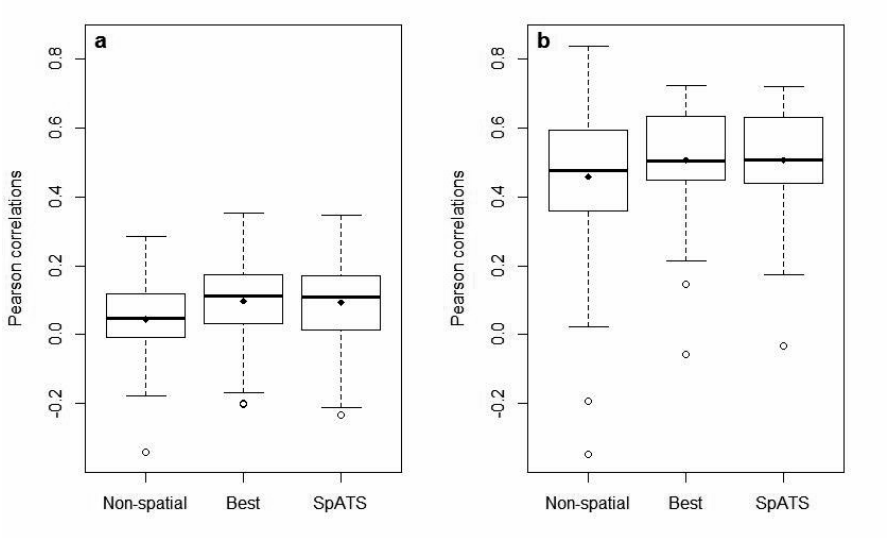


Fig. 6 Correlations of genotype BLUPs between environments from each model for grain yield (**a**) and plant height (**b**)

For illustration purpose, Fig. 7 presents the BLUPs of genotype effects from SpATS and the BSS models for grain yield in the example trials DYS05 and DAB08. Differences in the rankings were small for both environments. However, changes in the order of genotypes were more evident at DYS05 (Fig. 7a), an environment where, as previously noted, the nature of spatial variation was more complex. The predicted rankings established by both spatial methods were also consistent for the rest of the dataset, with mean Spearman correlations across trials of 0.970 for grain yield and 0.989 for plant height. As expected, the rankings of genotype were more dissimilar between SpATS and the non-spatial models. Rank correlations

for yield between these models ranged from 0.500 at DYS05 (where ED_s = 111.2) to 0.926 at DAL08 (where ED_s = 0.0), with a mean value of 0.802. In the case of plant height, correlations were generally higher, varying from 0.767 at HER05 (where ED_s = 64.8) to 0.988 at DAL08 (where ED_s = 0.4) and a mean value of 0.944.

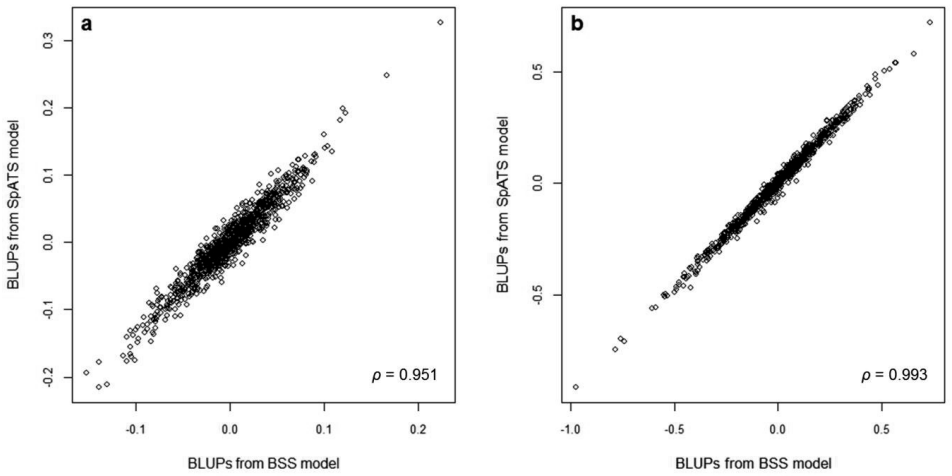


Fig. 7 Genotype BLUPs from the BSS model and the SpATS model and Spearman rank correlations (ρ) for grain yield in trials DYS05 (**a**) and DAB08 (**b**)

As suggested by one of the reviewers, we tried to implement a single-step model fitting strategy with the standard method by using the full model [3] across our dataset. Convergence problems were evident in 8 and 10 trials out of 20 for grain yield and plant height, respectively. It was possible to decrease the rate of failure by relaxing convergence criteria. However, given that we know that the full model was a misspecified one, tuning strategies should not be used to get convergence. The failures to converge reflected identifiability problems for the situation where the AR1xAR1 structure, global trend terms and the nugget are included in the same model. In contrast, the SpATS model did not suffer from this confounding difficulty; the three types of field variation were fitted in a stable way.

2.4. Discussion

2.4.1. Spatial analysis with SpATS

This study presented the SpATS model as a suitable alternative to the standard spatial models for the adjustment of field trends in sorghum genetic trials. We reported a first application of the new spatial model to a real and extensive plant breeding field testing. This

method fits a smooth surface to account for all sources of continuous environmental variation. The mixed model representation of SpATS features the joint modelling of additive one-dimensional trends plus interactions between trends in the row and column directions. Moreover, the model specification assigns different degrees of smoothing to each additive and interaction effect by means of specific smoothing parameters. These weighting terms, which are automatically tuned by REML-based variance components, shrink irrelevant effects to optimize the fit of the spatial surface.

We have stressed the practical importance of the effective dimension of the model as an integral part of spatial analysis with SpATS. This study highlights how ED_s can be used to interpret the intensity and the structure of spatial variation. The ED_s is a very appealing tool to quantify the magnitude of spatial effects, reflecting the amount of smoothing of the spatial surface and allowing an easy identification of the main patterns of field heterogeneity. Furthermore, the genetic effective dimension was used to compute a generalized heritability in the context of the analysis with SpATS (Rodríguez-Álvarez et al. 2016a). This novel expression of heritability is valid for more general situations commonly found in plant genetic trials, *e.g.*, when data is unbalanced and/or when residuals are spatially correlated. Equivalent definitions of generalized heritability were also proposed by Cullis et al. (2006) and Welham et al. (2010) in the context of standard mixed model analyses.

In order to subject the new method to a hard evaluation, we analysed large-scale sorghum breeding trials arrayed as partially replicated (*p*-rep) designs (Cullis et al. 2006). These experiments are characterised by the absence of traditional blocking factors, allowing very few or no design features to be retained in the randomization-based model. Moreover, the use of partially replicated experiments assumes that field trend affecting unreplicated genotypes can be properly predicted by the spatial model (Payne 2006). Consequently, the analysis of *p*-rep designs requires the inclusion of spatial parameters as an essential add-on component for an efficient testing of genetic material. The results of our study demonstrate the effectiveness of SpATS to account for spatial trend and predict adjusted genotypic values under these circumstances.

The SpATS model adjusts a continuous surface across the whole field. A more refined modelling could consider discontinuity in the spatial trends by fitting a different surface within each block. Even though the former approach is more conservative, we consider it reflects the structure of trial design and should be a realistic model for most commonly used experiments in plant breeding. Furthermore, the smoothness of the spatial surface fitted by SpATS is controlled by five different terms, providing enough flexibility for an appropriate fit of the spatial trend.

2.4.2. Comparison of SpATS and the standard method: parametrization

The SpATS model presents similarities with the full formulation of the standard spatial model (see eq. 3). Both models contains two one-dimensional spline terms, each one fitted as the sum of a fixed linear trend and a random non-linear component. In addition, discontinuous spatial trends are accounted for by random row and column effects in both cases. However, there is a major difference between both models when accounting for the remaining spatial variation. The standard model fits a separable AR1 process, whereas SpATS uses P-spline interaction terms.

This difference in parametrization affects the way in which SpATS and the standard methods model field variation. Under the standard mixed model approach, gradients across the field can be adjusted by blocking factors and by one-dimensional polynomials or splines along rows and columns. However, more complex two-dimensional gradients that do not align well with row and column directions are expected to affect field trials as well. The structure of spatial variation found in our dataset (Figure 2) and in previous studies of agricultural and forest field trials demonstrate that fitting only additive gradients in one dimension may result in insufficient modelling of global trend (Federer 1998; Fu et al. 1999; Taye and Njuho 2008). It is in principle possible to extend the standard spatial model with additional fixed terms, like a linear-by-linear interaction term (Federer 1998), and random terms, like a smoothing spline interaction term, but these extensions were never used under the standard approach and are prone to cause problems (Gilmour 2000). In this research, we showed that the SpATS model is able to account for intricate patterns of large-scale variation by explicitly modelling the interactions between global trends along rows and columns.

A common practice in spatial analysis is to fit an autoregressive model, originally proposed to adjust for local trend, and assume that it is flexible enough to also account for global trend (Zimmerman and Harville 1991; Dutkowski et al. 2006; Piepho et al. 2008b). A more conservative approach considers that the underlying spatial correlation is likely to hold only within blocks and that large-scale trend is partially accounted by blocking factors (Williams et al. 2006; Piepho and Williams 2010). However, continuous non-stationary trends across the whole field can be better fitted by specific spatial terms in the model, as was suggested in the seminal paper by Gilmour et al. (1997). Following their approach, we found that additional terms accounting for global variation could not have been ignored in most of the sorghum trials and for both traits (see Table 4). Several studies using real and simulated data have shown that underfitting global trend may cause the variance of treatment differences to be underestimated (Zimmerman and Harville 1991; Brownie et al. 1993; Brownie and Gumpertz 1997). This false improvement in precision can be particularly negative in plant breeding trials as it reduces the efficiency of selection decisions. Furthermore, Brownie and Gumpertz (1997) reported that local trend is overestimated in presence of unaccounted large scale trend.

Given that global and local model terms are actually "competing" to fit part of the same spatial variation, the estimated covariance parameters will vary according to the global terms included in the spatial model. This inconsistency in the estimates of autocorrelations was also observed in our study during the search of the BSS models (not shown). The aforementioned situation raises the issue of parameter identification when both global and local trend are trying to be fitted. Therefore, spatial parameters should be interpreted with special care under the standard approach. Conversely, the new spatial method based on 2D P-splines simplifies the problem of spatial model identification by always modelling all types of field trend as a single continuous process. This unified modelling avoids the necessity of distinguishing between global and local trend. Both forms of continuous variation are simultaneously fitted by the flexible interaction surface with anisotropic smoothing. As a result, SpATS provides a straightforward representation of the spatial trend that is easy to interpret. Moreover, the ANOVA-type decomposition of the smooth surface facilitates the characterization of the spatial trend, providing additional insight into the structure of field variation.

Another difference between both spatial methods was evident regarding the estimation of the residual variance. In our dataset, the standard spatial models exhibited a clear tendency to estimate smaller spatially independent components than the SpATS model (Figure 4). The same discrepancy was reported by Rodríguez-Álvarez et al. (2016a) in a simulation study where they analysed data generated according to different autoregressive models with nugget. These authors showed that, when the autocorrelations are large ($\rho_r = \rho_c = 0.9$), the SpATS model provides relatively accurate estimates of the random error variance, whereas the autoregressive model tends to underestimate this term. The possibility of confounding the spatial component with the nugget variance when fitting autoregressive models in field trials was also reported by Cullis et al. (1998) and extensively discussed in Piepho et al. (2015). Given the large autocorrelations estimated in most of our sorghum trials, we may suggest that SpATS performed generally better in identifying the true spatially independent residuals, while the BSS models were actually modelling part of the random error as spatially correlated data.

This potential confounding of parameters in autoregressive and other non-linear spatial models with nugget causes frequent convergence problems (e.g., Dutkowski et al. 2006; Müller et al. 2010; Liu et al. 2015; Rodríguez-Álvarez et al. 2016a). When convergence cannot be reached, one could fall back to alternative models without nugget effects (Müller et al. 2010; Leiser et al. 2012). This strategy is far from attractive given that the potential best fitting model would be deliberately ignored. Furthermore, our research (Table 3) and other studies demonstrated that a spatially independent component accounting for measurement error is frequently required (e.g., Cullis et al. 1998; Qiao et al., 2000; Liu et al. 2015). In contrast to the standard spatial modelling approach, the SpATS model always fits a random error variance on top of the spatial surface and, in our experience, it always converges readily;

see also Rodríguez-Álvarez et al. (2016a). As a reviewer suggested, in addition to the identifiability issues mentioned above for the standard approach, it cannot be excluded that the difference in convergence performance between SpATS and the standard models may be related to the standard method using a covariance structure that is non-linear in the variance parameters, while the covariance structure of SpATS is linear in the parameters. Further study is required here.

2.4.3. Comparison of SpATS and the standard method: performance

The comparison between SpATS and the best fitting standard spatial models revealed a similar performance for the evaluation criteria considered in this paper. Besides the differences discussed above, both methods caused similar reductions in the spatially independent residual variance compared with the error of the non-spatial model. These changes indicate the magnitudes of spatial variation adjusted by the spatial models for both traits. The generally large decreases in the random error component (>30%) obtained for grain yield reflect that strong spatial trends affected this trait in most trials (Stroup et al. 1994; Yang et al. 2004). The lower reductions observed for plant height could be related to the dominant presence of random environmental variation (see Table 3). Interestingly, the same inferences can be drawn by considering the higher ED_s that were usually associated with grain yield trends (Fig. 1). The larger number of parameters effectively estimated by SpATS to better approximate the underlying spatial surface reflected the higher intensity of field trends for grain yield data. The ability of the ED_s to indicate the relative importance of spatial variation was evidenced by the strong positive association between the number of ED_s and the ratio of spatial to spatially independent variance from the standard models.

In general, the estimates of genetic variance from the SpATS model were comparable to those obtained by the BSS models. The inconsistencies between both models observed in some cases may result from the impossibility to clearly identify the genetic and the environmental variation in presence of spatial correlation. Several simulation studies have shown that unadjusted patchiness in the field may inflate the genetic variance (e.g., Loo-Dinkins 1990; Magnussen 1993, 1994). This identification problem was apparent across candidate BSS models, where the autoregressive models ignoring the nugget estimated higher trial genetic variances than the better-fitting models using nugget (data not shown). The overestimation of genetic variation when adjusting autoregressive models without nugget was also reported by Dutkowski et al. (2002) in tree breeding trials and by Rodríguez-Álvarez et al. (2016a) using simulated data. In addition, the latter authors showed that SpATS produced more accurate estimates of genetic variance, which were highly consistent with those obtained from the best fitting standard model including the nugget. However, more extensive

assessments of the SpATS model would be still necessary with respect to the validity of estimates when spatial variation is present.

Several studies considered the changes in heritability to measure the impact of alternative models on the efficiency of plant breeding evaluations (e.g. Smith et al. 2001b; Welham et al. 2010; Sarker and Singh 2015). Following this approach, we used the generalized heritability to compare the performance of the SpATS model and the standard spatial models. The adjustment of spatial trends with the new spatial model led to levels of heritability equivalent to the standard models in all the sorghum trials. The increases in grain yield heritability compared to the randomization-based model were broadly consistent with the results from standard spatial analysis of sorghum breeding trials in West Africa (Leiser et al. 2012). The improvement in precision, measured as the increase in the correlation of genotype predictions between environments, were generally the same for both spatial methods. Similar magnitudes of improvements through standard spatial analysis were previously reported by Leiser et al. (2012) in sorghum, but smaller increases have been achieved for wheat, sugar beet and barley breeding trials (Qiao et al. 2004; Müller et al. 2010).

The analysis with SpATS affected the predictions of genotypic values, since the ranking of genotypes changed after modelling spatial trends. A bigger impact on genotype ranks was usually observed in cases where the fit of a smooth surface produced larger increases in broad-sense heritability. Our results showed high consistency in the ranking of genotypes predicted by the SpATS model and the BSS models for all cases. This indicates that the use of the new spatial method would hardly produce changes in selection decisions compared to the more refined spatial models. The consistent but small changes in predicted rankings may be a consequence of the differences discussed above related to how both spatial methods accommodate global and local trends.

2.4.4. Comparison of SpATS and the standard method: modelling strategy

In this paper, we used a single-step modelling strategy to perform the definite spatial analysis in each trial. Furthermore, the same SpATS model was applied for individual-trial analysis across the whole dataset. This approach differs from the common multi-step modelling procedure based on sequential fitting of alternative spatial models for each trial. The latter practice may be a limitation for efficient routine application given that several model selection steps are required to arrive at a final spatial model. To perform the standard spatial analysis in the present paper, we inspected alternative AR1 models. However, the number of potential candidate models increases if other spatial covariance structures are also considered. A strategy to simplify the model selection process may be to restrict the number of candidate spatial models, potentially reducing the efficiency of analysis. A remarkable attempt to maximize efficiency of plant breeding trials through standard spatial analysis was reported by

Leiser et al. (2012), who fitted 91 different models for each trial to identify the best models in 17 environments. Unfortunately, these efforts for further modelling usually result in modest benefits relative to simpler models. Alternative spatial methods based on kriging are also time-consuming and difficult to apply in practice (Zas 2006; de la Mata and Zas 2010).

Our approach using SpATS accounted for all types of spatial variation by fitting a single model rather than using a multi-step modelling procedure. Under this simplified strategy, model selection steps required to identify the appropriate spatial correlation and/or global trend terms are not needed; both local and global trends are automatically modelled in a single step by the smooth surface. The SpATS approach relies on the estimation procedure to effectively reduce the influence of the smooth surface components that are not needed. This implicit model selection is automatically tuned by specific smoothing parameters (or penalties) and is reflected in the ED_s. Accordingly, after convergence, the ED_s of unimportant components will tend to zero, meaning that these terms are not contributing to the complexity of the spatial model. The new method also simplifies the practice of using diagnostic graphics, such as variograms, to guide model selection. The reason is that the selections steps required to fit global and local trends under the standard method are reduced to one with SpATS, and thus the diagnostic plots associated to those steps are essentially skipped. Random row and column effects were fitted by default in our SpATS model, as discontinuous spatial effects were also present in most cases (data not shown). The inclusion of these effects in a default spatial model is justified by the frequent existence of non-smooth effects caused by blocking factors or extraneous variation (e.g., Piepho and Williams 2010; Liu et al. 2015). We set the same number of equally-spaced knots in each dimension of the 2D P-spline for every trial. These quantities were chosen to be so many as to ensure ample flexibility to the smoother. Other studies on spatial analysis with P-splines reported that models using different numbers of knots produced similar fits and results (Cappa and Cantet 2007; Cappa et al. 2010). Moreover, Eilers et al. (2015) demonstrated that, once a sufficient number of knots has been chosen, optimizing their quantity is not worthwhile because the smoothing parameters will regulate the smoothness of the fit to optimize the bias-variance trade-off.

For the present research, we have used a general SpATS model considering the design and treatment factors of our dataset. However, it is noteworthy that the mixed model formulation of SpATS enables more refined model building/selection according to specific situations. For instance, having an ED of zero is equivalent to an associated variance component being zero. It implies that we could use any tests that evaluate the relative fit of a variance model (e.g., REML-LRT, AIC) to perform model selection.

The results from this study showed that the SpATS model performed comparably to more refined and site-specific spatial models. One advantage of the novel method is that all types of continuous spatial variation and genetic effects can be modelled simultaneously in a single modelling step. As Dutkowski et al. (2006) pointed out, this approach should be superior to

fitting all terms in a multi-step process as it will avoid parameter identification problems derived from confounding spatial heterogeneity with genetic heterogeneity due to aggregation of related genotypes. An additional benefit is that the SpATS model may be useful to improve the efficiency of two-stage analysis of multi-environment trials (MET). The reason is that the same flexible model can be fitted in the first stage to account for the spatial surfaces of all the trials, obtaining adjusted genotype means to be used in the second stage. The gain in speed of analysis using the new method results from the fact that less computational steps would be needed to identify an appropriate spatial model for each trial.

2.5. Conclusion

The SpATS model provided a flexible and efficient alternative to account for spatial patterns in the sorghum breeding field trials. The performance of the new model was equivalent to the more elaborate standard spatial models when considering the improvement in precision and the predictions of genotypic values. The suitability of SpATS was consistent across trials and traits exhibiting different magnitudes of heritability and complexity of spatial variation. A major advantage of the new model over existing techniques is that global and local trends are jointly modelled by the smooth surface. Moreover, we used a general SpATS model to adequately fit all experiments, which avoids the examination of several candidate models for each trial. Given the results of this study, the use of the new method should be considered as a simple and effective strategy to optimize the practical application of spatial analysis in plant breeding trials.

Author Contribution Statement:
FvE conceived the research. JV, DJ, MM and FvE designed the research. JV performed statistical analyses and wrote the manuscript. MXRA, MB and PE supported the application and understanding of the SpATS methodology and corresponding R-package. DJ coordinated the field trials and the data collection. MXRA, MB, MM and FvE edited the manuscript. All authors read and approved the final manuscript.

Acknowledgments:
We thank two anonymous reviewers for insightful comments and suggestions on earlier versions of the manuscript. The trial data used in this research was generated with funding from the Grains Research and Development Corporation (GRDC) of Australia. J.G. Velazco acknowledges financial support from the National Institute of Agricultural Technology (INTA) of Argentina, Res. DN 1126/13. M.X. Rodríguez-Álvarez was supported by the Spanish Ministry of Economy and Competitiveness MINECO grant MTM2014-55966-P and BCAM Severo Ochoa excellence accreditation SEV-2013-0323, and by the Basque

Government through the BERC 360 2014-2017. M. Malosetti and F.A. van Eeuwijk worked on this paper as part of the Integrated Breeding Program.

Chapter 3

Combining pedigree and genomic information to improve prediction quality: an example in sorghum

Velazco J.G., Malosetti M., Hunt C.H., Mace E.S., Jordan D.R., van Eeuwijk F.A.
Theor Appl Genet (2019) 132(7):2055–2067.

Abstract:
Selection based on genome-wide markers has become an active breeding strategy in crops. Genomic prediction models can make use of pedigree information to account for the residual polygenic effects not captured by markers. Our aim was to evaluate the impact of using pedigree and genomic information on prediction quality of breeding values for different traits in sorghum. We explored BLUP models that use weighted combinations of pedigree and genomic relationship matrices. The optimal weighting factor was empirically determined in order to maximize predictive ability after evaluating a range of candidate weights. The phenotypic data consisted of testcross evaluations of sorghum parental lines across multiple environments. All lines were genotyped and full pedigree information was available. The performance of the best-predictive combined matrix was compared to that of models fitting the component matrices independently. Model performance was assessed using cross-validation technique. Fitting a combined pedigree-genomic matrix with the optimal weight always yielded the largest increases in predictive ability and the largest reductions in prediction bias relative to the simple G-BLUP. Under this approach, the weight that optimized prediction varied across traits. The benefits of including pedigree information in the genomic model were more relevant for traits with lower heritability, such as grain yield and stay-green. Our results suggest that the combination of pedigree and genomic relatedness can be used to optimize predictions of complex traits in crops when the additive variation is not fully explained by markers.

3.1. Introduction:

Selection based on dense genome-wide markers has become a revolutionary alternative to traditional genetic evaluations for improving quantitative traits in crops (Jannink et al. 2010; Crossa et al. 2017). This selection technique exploits the association between high-density markers and unknown causative genes to predict genetic merit or breeding value (BV). Genomic prediction (GP) is expected to increase accuracy of evaluations by capturing large and small allelic effects across the genome simultaneously (Meuwissen et al. 2001). These effects are estimated using phenotypic and genotypic data from a reference breeding population, and then integrated to predict genome-assisted BVs of untested selection candidates that have been only genotyped. The implementation of this prediction method can potentially lead to higher rates of genetic gain and lower phenotyping costs compared to classical phenotypic or pedigree-based selection.

The underlying requirement for obtaining optimal genomic predictions is that available markers or haplotypes of markers are in complete linkage disequilibrium (LD) with quantitative trait loci (QTLs) of interest. Consequently, a major assumption is that the full additive genetic variance can be accurately explained by markers (Goddard 2009). When this condition is not met due to incomplete LD with causative genes, accuracy of prediction is expected to decline (Habier et al. 2007; Goddard et al. 2011). In such cases, pedigree information may be incorporated into GP models to account for the residual polygenic variance not captured by markers, and to reduce empirical bias of predictions. Pedigree and genome-wide markers can offer different, yet complementary, information on genetic relatedness among individuals. While pedigrees represent expected average relationships describing potential transmission of genes, genomic data provides observed realized relationships. The latter are expected to be more accurate because markers can trace alleles, capturing random Mendelian sampling and unknown ancestral relationships not considered in the pedigree. Nevertheless, the inclusion of a residual genealogical effect in prediction models might account for potential LD patterns not explained by markers at population and family levels. Consequently, the joint use of pedigree and genomic information may provide better estimates of genetic similarities between genotypes, affecting predictive performance.

The benefits of exploiting pedigree and marker data in plants have been previously reported in the context of QTL and association mapping (Bink et al. 2002; Parisseaux and Bernardo 2004; Malosetti et al. 2007). In sorghum, Jordan et al. (2004) demonstrated that combining pedigree information with mapped markers facilitated the identification of genetic regions under selection in a breeding population. Within the GP framework, several approaches to combine genealogy with genomic data have been reported. Bayesian regression and semiparametric models were firstly implemented in crops (de los Campos et al. 2009; Crossa et al. 2010). Alternative models based on best linear unbiased prediction (BLUP) have

been used in maize and wheat (Albrecht et al. 2011; Burgueño et al. 2012; Sukumaran et al. 2017). These BLUP models include two mutually independent genetic effects, one depending on a pedigree-based relationship matrix (**A**) and one depending on a genomic relationship matrix (**G**). A similar approach based on BLUP integrates pedigree and genomic relatedness into a single matrix. This strategy has been applied in the context of animal genetic evaluations (VanRaden 2008; Goddard et al. 2011; Gao et al. 2012), but it has not been implemented so far in plants. Different models combining pedigree and genome-wide markers have also been proposed for specific situations where not all individuals in the reference population are genotyped (Legarra et al. 2014; Liu et al. 2014; Fernando et al. 2016).

Compatibility between **A** and **G** may be an issue when jointly used in prediction models. It is important to consider that genetic relationships in both matrices are in different scales with reference to the base population (Legarra et al. 2014). Therefore, the use of a compatible scale for pedigree and genomic relationships should be considered when estimating and interpreting genetic variances and derived measures, such as heritability and expected response to selection. Different scaling methods have been implemented to achieve the same base population for **A** and **G** (Vitezica et al. 2011; Christensen 2012). These corrections are aimed at accounting for the non-randomness of genotyping due to previous selection decisions.

Sorghum [*Sorghum bicolor* (L.) Moench] is the 5th most important cereal crop worldwide after wheat, maize, rice and barley. Its drought-tolerance ability makes it a strategic crop for sustainable grain production in the perspective of climate change and increasing food demand. Early-stage breeding of hybrid sorghum involves the development of elite inbred lines, which will be subsequently used as parents of commercial hybrids. Initial selection of superior parental lines is typically based on their additive genetic values estimated from testcross performance trials. Genomic selection could be beneficial for reducing time and field testing resources during the testcross evaluation step, increasing efficiency of hybrid development. Despite the new opportunities offered by GP methods to accelerate genetic progress in sorghum, empirical studies are still limited compared to other crops (Kulwal 2016). A first implementation of GP in sorghum was reported for biomass traits in a global germplasm collection (Yu et al. 2016). In a more recent study by Hunt et al. (2018), genomic models were applied for prediction of testcross yield performance in the context of individual trial analysis.

For the present research, we considered a multi-year and multi-location testcross evaluation of sorghum parental lines using several testers and including different production and adaptability traits. The dataset belongs to a public breeding program in Australia where identification of superior parental lines is typically based on extensive phenotyping of progeny performance. This interesting case study was used to explore the potential of combining pedigree, markers, and phenotypic data for optimization of genomic prediction in sorghum.

The objectives of this article were to explore the impact of combining pedigree and genomic information on the quality of BV predictions, and to determine the combinations of information that optimize predictive performance for different traits. For these purposes we applied BLUP models using a blended kinship matrix constructed as a weighted combination of matrices **A** and **G**, where different weights were tried in search of improved prediction quality.

3.2. Materials and Methods:

3.2.1. Phenotypic data

The dataset is part of the sorghum breeding program for female parental lines conducted by the University of Queensland and the Department of Agriculture and Fisheries in Queensland, Australia. Female lines are evaluated in hybrid combination with different male testers across multiple environments. Testcross performance is then used to estimate the general combining ability (GCA) or BV of parental lines.

The phenotypic records used in this study consisted of 26 testcross performance trials where a total of 646 female lines were tested across 12 locations over a period of 7 years between 2008 and 2014. This trial series comprises a representative sample from a target population of environments covering the main sorghum cropping region in Australia. Phenotypes of 2645 testcross hybrids were used to assess female lines in crosses with one to five different testers. These male parents were chosen to express contrasting levels of yield potential and stay-green capacity (Jordan et al. 2012). In each trial, between 110 and 315 lines were evaluated and 3 to 5 testers were used. Across the dataset, more than 50% of the lines were crossed with at least 3 different testers and grown in at least 9 environments. The sets of testcrosses entering evaluation were designed to provide a degree of connectivity between lines and testers across trials. Each experiment was laid out as a resolvable partially replicated design (Cullis et al. 2006), where 30% of the testcross hybrids had two replicates and commercial varieties were included with additional replication. The number of testcrosses in each trial varied between 247 and 858. We considered four productivity and adaptability traits routinely measured by the program: grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT). Stay-green is a drought resistance trait that expresses as delayed leaf senescence in environments where water-stress conditions occur (Borrell et al. 2014). In this dataset, the stay-green trait was expressed in 9 trials and observations were available for 603 lines.

3.2.2. Pedigree and genotypic data

Inbred parent lines were derived from pedigree breeding methods resulting in a highly structured breeding population. The 646 female lines are basically grouped into 74 full-sib families including different numbers of siblings. Genealogical information on the tested lines and 499 ancestors tracing back 28 generations was available to compute the pedigree-based relationship matrix **A**.

All the female lines were genotyped using an integrated DArT and genotyping-by-sequencing (GBS) methodology involving complexity reduction of the genomic DNA to remove repetitive sequences using methylation sensitive restriction enzymes prior to sequencing on Next Generation sequencing platforms (DArT, www.diversityarrays.com). The sequence data generated were then aligned to the most recent version (v3.1.1) of the sorghum reference genome sequence (Paterson et al. 2009) to identify SNP (Single Nucleotide Polymorphism) markers. SNPs with minor allele frequency lower than 2.5% or more than 20% of missing values were discarded. Missing genotypes were imputed based on random sampling from marginal allele distributions using the R package synbreed (Wimmer et al. 2012). After quality filtering, 4781 evenly-spaced SNPs were retained to compute the genomic relationship matrix **G**.

3.2.3. Phenotypic analysis

A weighted two-stage approach was used for the analysis of phenotype data. In the first stage, each trial was individually analyzed to account for design factors and spatial field variation. In the second stage, spatially-adjusted testcross means from the first stage were used to compute adjusted line means across testers and environments.

For the analysis of each trial we applied a novel spatial method that adjusts for all types of field trends in a single modelling step by fitting a smoothed surface (Rodríguez-Álvarez et al. 2018a). We used the same flexible spatial model to analyze the whole series of trials and all traits. Velazco et al. (2017) showed that this approach performs as well as the more elaborate spatial methods, which are typically based on a specific multi-step modelling for each trial and trait. The general spatial model used across trial-trait combinations is defined as

$$y_{ijkl} = \mu + (L \times T)_i + B_j + R_k + C_l + f(r, c)_{kl} + e_{ijkl}, \quad (1)$$

where the plot observation y_{ijkl} was modeled by fitting fixed effects for: the overall mean (μ), the i-th line × tester hybrid ($L \times T$) and the j-th block (B); and random effects for: the k-th row (R) and the l-th column (C). The term $f(r, c)$ is a smooth function of row (r) and column (c) plot coordinates representing the fitted spatial surface, which simultaneously accounts for global and local trends (see Rodríguez-Álvarez et al. 2018a; Velazco et al. 2017 for details).

Finally, e is the random spatially independent residual representing measurement error in each plot. All random effects were assumed independent, homoscedastic and normally distributed with zero mean.

The spatial analyses were implemented within the REML-based mixed model framework using the R-package SpATS (Rodriguez-Álvarez et al. 2018b).

In the second stage, spatially-adjusted testcross means from all trials where jointly modelled by

$$y_{ijk} = \mu + L_i + T_j + E_k + LT_{ij} + LE_{ik} + TE_{jk} + LTE_{ijk} + e_{ijk}, \quad (2)$$

where in this case y_{ijk} represents the adjusted mean estimated by best linear unbiased estimation (BLUE) of the i-th female line crossed with the j-th tester in the k-th environment, which was fitted by a main line genetic effect (L), a main tester effect (T), a main environmental effect (E) and all possible interactions between these effects. The residual followed $e_{ijk} \sim N(0, \mathbf{R})$, where $\mathbf{R}$ is a diagonal matrix with elements equal to the squared standard errors of each mean y_{ijk} estimated by model (1) in the first stage (Frensham et al. 1997). The analysis using diagonal weights instead of the full genotypic covariance matrix from each trial is preferable in practice because of its computational efficiency and comparable results (Möhring and Piepho 2009; Welham et al. 2010). Since trials were considered random, all the interactions involving E were random. Heterogeneous variances for the latter effects were allowed to improve goodness of fit of the model in each trait, as evaluated by the Akaike Information Criterion. In this stage, the effects L and T were also taken as fixed. However, given that not all lines were testcrossed with all testers, the interaction effect LT was considered random to estimate line means across testers (Bernal-Vazquez et al. 2014).

3.2.4. Prediction models

Different parental or GCA models based on BLUP were applied to predict breeding values of female lines from progeny performance. These models differed in the amount of pedigree and genomic information used for predictions.

The prediction models used in our study assume that available SNPs may not explain all additive genetic variance. The general model formulation can be defined as:

$$\mathbf{y_L} = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e}, \quad (3)$$

where the vector $\mathbf{y_L}$ contains the BLUEs of line effects ($\hat{L}_i$) from model (2), $\mathbf{1}$ is a vector of ones with associated general mean μ, $\mathbf{Z}$ is a design matrix allocating line BLUEs to unknown genetic effects, $\mathbf{g}$ is the vector of total additive genetic effects and $\mathbf{e}$ is the vector of residuals.

Random residuals were assumed $\mathbf{e} \sim N(0, \mathbf{R})$, where $\mathbf{R}$ is a diagonal matrix with elements equal to the squared standard errors of genotypic BLUEs from the second stage of phenotypic analysis. This matrix accounts for differences in precision of estimated line means.

Total additive genetic effects were assumed $\mathbf{g} \sim N(0, \mathbf{K}\sigma_g^2)$, where $\mathbf{K}$ is a combined kinship matrix exploiting pedigree and genomic information. This matrix is constructed as $\mathbf{K} = w\mathbf{A} + (1 - w)\mathbf{G_s}$, where $\mathbf{A}$ is the numerator relationship matrix among lines computed from the full pedigree and $\mathbf{G_s}$ is a scaled genomic relationship matrix based on the SNP data (see details below). The weighting factor w represents the fraction of total additive variance (σ_g^2) that is not captured by markers, such that $w\sigma_g^2$ is the amount of residual polygenic variance explained by genealogical relationships. In this model, referred here as K-BLUP, the weight can take any value between 0 and 1. It should be noted that $\mathbf{G}$ and $\mathbf{A}$ are not orthogonal; therefore, the interpretation of w should be considered with caution.

For our study, a sequence of eight values of w from 0.1 to 0.8, with increments of 0.1, was explored to assess the impact on predictions. For the sake of comparison, the extreme cases $w = 0$ and $w = 1$ were also considered. Note that the latter case corresponds to the traditional pedigree-based model, A-BLUP, which relies only on familial information. Alternatively, assuming $w = 0$ results in the conventional G-BLUP model (VanRaden 2008), where predictions are exclusively conditional on marker-based similarities.

An equivalent formulation of the K-BLUP model is as follows:

$$\mathbf{y_L} = \mathbf{1}\mu + \mathbf{Zm} + \mathbf{Za} + \mathbf{e},$$

where the total additive genetic effects in (3) are decomposed into a vector of genomic additive effects ($\mathbf{m}$) and a vector of residual polygenic effects ($\mathbf{a}$), such that $\mathbf{g} = \mathbf{m} + \mathbf{a}$, with respective variances $\sigma_g^2 = \sigma_m^2 + \sigma_a^2$. In this model both types of genetic effects are assumed mutually independent with distributions $\mathbf{m} \sim N(0, \mathbf{G_s}\sigma_m^2)$ and $\mathbf{a} \sim N(0, \mathbf{A}\sigma_a^2)$, where $\sigma_m^2 = (1 - w)\sigma_g^2$ and $\sigma_a^2 = w\sigma_g^2$. This parametrization, referred in the rest of the article as AG-BLUP, presents two differences compared to the K-BLUP model: first, the AG-BLUP model requires the fitting of two relationship matrices ($\mathbf{A}$ and $\mathbf{G}$) to combine pedigree and genomic information, while this information is condensed into a single matrix in K-BLUP; second, under AG-BLUP, the magnitude of w is driven by the data in order to maximize the (restricted) likelihood of the model, as opposed to pre-specified weights used within the K-BLUP method. The performance of AG-BLUP was also compared to that of the best-predictive K-BLUP model.

The $\mathbf{G_s}$ matrix used in our models was rescaled to make it compatible with $\mathbf{A}$ in reference to the same base breeding population. The adjustment was based on fitting $\mathbf{G_s}$ to $\mathbf{A}$ by applying $\mathbf{G_s} = a + b\mathbf{G}$, where $\mathbf{G}$ is the unscaled genomic matrix as computed with the first

method of VanRaden (2008), and the parameters *a* and *b* are estimated by equating the average levels of inbreeding and the overall relationships in **A** and **G** (Vitezica et al. 2011; Christensen et al. 2012). The added constant *a* accounts for old relationships among non-genotyped ancestral lines in the pedigree, while *b* is a scaling term accounting for the reduction in the genetic variance of genotyped lines relative to the pedigreed base population (see Legarra et al. 2014 for details).

Given the variance components from prediction models, the narrow-sense heritability (h^2) of line means was obtained as: $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$, where σ_e^2 is the residual variance comprising non-additive genetic effects and true errors associated with mean line estimates.

Models in the second stage of phenotypic analysis as well as prediction models were fitted using the average information REML (AI-REML) algorithm as implemented in the mixed model package ASReml-R (Butler et al. 2017).

3.2.5. Model validation

The quality of predictions from each model was evaluated using a five-fold cross-validation technique. We considered two different strategies for splitting the data into training (TS) and validation sets (VS). These strategies emulate different selection schemes: one based on within-family predictions (W-fam) and one based on among-family predictions (A-fam). In W-fam, 20% of lines from each full-sib family formed the VS, i.e., predicted lines belong to families that were (tested) in the TS. In A-fam, we sampled 20% of whole families to construct the VS, i.e., predicted lines belong to full-sibs families that were not present in the TS. With this setting we examined contrasting levels of genetic relatedness between TS and VS: high for W-fam and low for A-fam. Each splitting scenario was repeated 20 times using the same random seed throughout all models. The line BLUEs from phenotypic analysis ($\mathbf{y_L}$) were regarded as the realized genetic values of lines and used for validation of prediction models.

3.2.6. Prediction quality evaluation

Model performances were assessed using multiple criteria. In this research, predictive ability was taken as the primary criterion for identifying the best prediction model. However, measures of empirical bias and accuracy were also considered, as they offer complementary information that should be taken into account when evaluating the predictive performance of models. Predictive ability was measured as the Pearson's correlation (r_{PA}) between predicted additive values $\hat{\mathbf{g}}$ and realized values $\mathbf{y_L}$ in the validation set. Bias of predictions was investigated by regressing $\mathbf{y_L}$ on $\hat{\mathbf{g}}$, where a coefficient of regression $b = 1$ designates an

empirically unbiased predictor, while $b < 1$ indicates inflation of the variance of genotype predictions or over-predictions. The mean squared error of prediction (MSEP) from the linear regression was also measured as an indicator of prediction accuracy, which incorporates concepts of both bias and precision. The evaluation of models was based on average values over the 20 replicates when using within-family or across-family relatedness for prediction. Significance of pairwise differences in predictive ability among prediction models was assessed by the Hotelling-Williams t-test (Steiger 1980), which is the appropriate test when comparing two correlations that are not independent as a result of sharing a common variable ($\mathbf{y_L}$ in our case).

3.3. Results

3.3.1. Heritability and effect of scaling genomic relationships

Table 1 presents the narrow-sense heritabilities (h^2) for all traits estimated from BLUP models using only pedigree relationships, using only genomic relationships or combining both sources of information. We also show the influence of scaling the genomic matrix on the estimates of h^2. In general, heritabilities varied from relatively low in grain yield to high in the case of plant height. Models using the combined matrix **K** increased h^2 for most traits compared to A-BLUP and G-BLUP. These changes in heritability reflect the ability of the models to capture genetic variation in the breeding population. Note that the weight w used to construct **K** in each case was defined to maximize the likelihood, being the resulting K-BLUP model equivalent to fitting the model AG-BLUP. Finally, using the scaled genomic matrix $\mathbf{G_s}$ in G-BLUP and K-BLUP increased heritability estimates for all traits relative to the heritabilities obtained using the conventional genomic matrix.

Table 1 Estimates of narrow-sense heritability for grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT) from prediction models using only the pedigree-based matrix (A-BLUP), using only the unscaled (**G**) or the scaled ($\mathbf{G_s}$) genomic matrix, and using the combined **K** matrix (K-BLUP)

Model	GY		SG		PH		FT	
	G	$\mathbf{G_s}$	**G**	$\mathbf{G_s}$	**G**	$\mathbf{G_S}$	**G**	$\mathbf{G_S}$
A-BLUP ($w = 1$)	0.40	0.40	0.61	0.61	0.70	0.70	0.63	0.63
G-BLUP ($w = 0$)	0.28	0.34	0.41	0.46	0.72	0.77	0.62	0.67
K-BLUP (w = maxLL)[a]	0.43	0.46	0.56	0.59	0.78	0.82	0.72	0.76

[a]maxLL: weight that maximized the REML log-likelihood in each case; for GY: $w = 0.54$ with **G** and $w = 0.48$ with $\mathbf{G_s}$, for SG: $w = 0.43$ with **G** and $w = 0.38$ with $\mathbf{G_s}$, for PH: $w = 0.19$ with **G** and $w = 0.16$ with $\mathbf{G_s}$, for FT: $w = 0.34$ with **G** and $w = 0.29$ with $\mathbf{G_s}$.

3.3.2. Impact of combining pedigree and genomic relationships on prediction quality

Fig. 1 shows the patterns of variation in prediction quality measures of BLUP models when using different weights (w) to construct the combined relationship matrix **K**.

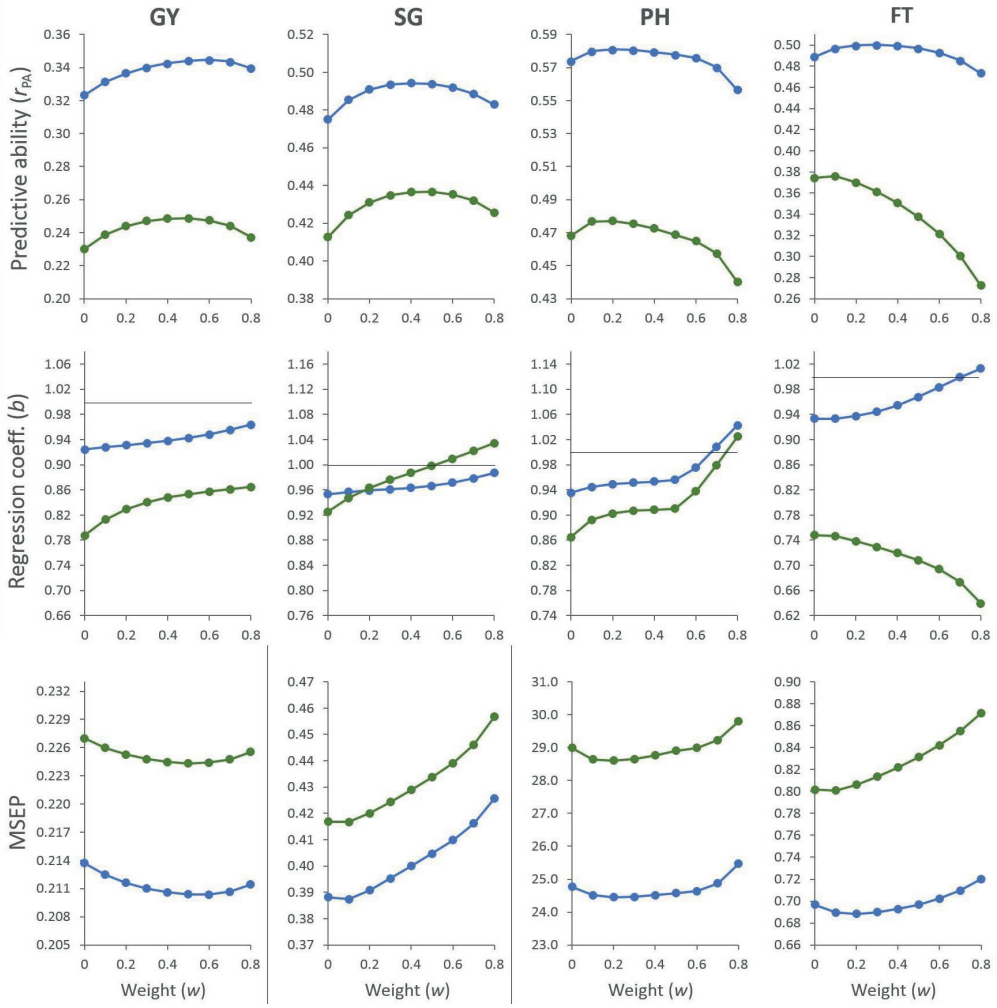


Fig. 1 Predictive abilities, regression coefficients and MSEP from BLUP models using different weights (w) to construct the combined matrix **K** for grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT) predictions within (*blue*) and among (*green*) families. The weight $w = 0$ correspond to the simple G-BLUP model. The *horizontal lines* indicate a regression coefficient $b = 1$

In both prediction scenarios, predictive ability for GY and SG improved as the weight placed on pedigree information increased from $w = 0$ up to $0.4 \leq w \leq 0.6$. The use of higher

weights caused decreasing trends in r_{PA} for both traits. The largest relative increases in predictive abilities were observed for GY, with K-BLUP achieving 7% and 8% of improvement compared to G-BLUP (w = 0) for within- and among-family predictions, respectively. For PH and FT, r_{PA} was generally maximized by using lower weights ($w \leq 0.2$) in both prediction scenarios. Any further increase in w slightly changed r_{PA} within families and was clearly detrimental for among-family predictions. Using a combined pedigree-genomic matrix provided more limited benefits for PH and FT, representing less than 2% of relative gains in r_{PA} over using only marker information.

In most cases, placing increasing weights on pedigree relationships reduced bias of predictions relative to predictions based on genomic relationships alone (w = 0). Moreover, the addition of genealogy information was more effective in reducing inflation of predictions when these were based on among-family relatedness. The only exception was observed for FT, where higher values of w gave more biased among-family predictions.

The use of increasing weights to form the **K** matrix reduced MSEP for GY, reaching the lowest values at w = 0.5 in both prediction scenarios. The opposite tendency was observed for the other traits, with higher accuracies obtained when more weight was assigned to genomic relationships and with accuracy of PH predictions being less sensitive to changes in w.

The optimal weighting factors for each trait and prediction scenario are given in Table 2. These weights were determined in order to maximize predictive ability of K-BLUP after examining the variation in r_{PA} over the set of candidate weights, as shown in Fig. 1. The best-predictive weights based on cross-validation were generally similar to the weights that maximized the log-likelihood of the model for each trait in the entire breeding population (see footnotes to Table 1).

Table 2 Optimal weights (w) used to construct the combined matrix **K** for prediction of grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT) under within- and among-family prediction scenarios

Prediction scenario	GY	SG	PH	FT
Within families	0.6	0.4	0.2	0.2
Among families	0.5	0.5	0.2	0.1

3.3.3. Performance of prediction models

Table 3 presents the measures of prediction quality from the BLUP models using pedigree and/or genomic information for all traits and both cross-validation scenarios. Independent of the prediction model, predictive abilities were significantly lower and predictions tended to be more biased when among-family information was used. The pedigree-based BLUP model, considered here as the benchmark, always gave the lowest r_{PA} and the highest MSEP.

Table 3 Mean values (and SD of 20 replicates) for predictive ability (r_{PA}), relative increment of r_{PA} (Δr_{PA}), regression coefficient (Bias) and mean squared error of predictions (MSEP) from BLUP models using different relationship matrices for grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT) prediction within and among families. The best values for each evaluation criterion are boldfaced

Trait	Quality criterion	A-BLUP	G-BLUP	AG-BLUP	K-BLUP[a]
Within-family prediction					
GY	r_{PA}	0.299 (0.011)	0.323 (0.013)	0.339 (0.014)	**0.345** (0.014)
	Δr_{PA}(%)	0	8.1	13.4	**15.3**
	Bias (*b*)	**0.963** (0.040)	0.924 (0.045)	0.924 (0.043)	0.948 (0.045)
	MSEP	0.217 (0.002)	0.214 (0.002)	0.211 (0.002)	**0.210** (0.002)
SG	r_{PA}	0.437 (0.010)	0.475 (0.007)	0.490 (0.009)	**0.494** (0.009)
	Δr_{PA}(%)	0	8.6	12.1	**13.0**
	Bias (*b*)	**0.988** (0.030)	0.953 (0.024)	0.952 (0.025)	0.963 (0.025)
	MSEP	0.514 (0.005)	**0.388** (0.004)	0.401 (0.005)	0.400 (0.005)
PH	r_{PA}	0.420 (0.011)	0.574 (0.011)	0.579 (0.010)	**0.581** (0.009)
	Δr_{PA}(%)	0	36.6	37.8	**38.3**
	Bias (*b*)	**0.997** (0.032)	0.935 (0.024)	0.944 (0.023)	0.949 (0.021)
	MSEP	30.4 (0.4)	24.8 (0.5)	24.6 (0.5)	**24.4** (0.4)
FT	r_{PA}	0.394 (0.015)	0.489 (0.011)	0.497 (0.014)	**0.500** (0.013)
	Δr_{PA}(%)	0	24.0	26.1	**26.8**
	Bias (*b*)	**0.964** (0.045)	0.933 (0.029)	0.937 (0.030)	0.944 (0.027)
	MSEP	0.774 (0.011)	0.697 (0.011)	0.693 (0.013)	**0.690** (0.013)
Among-family prediction					
GY	r_{PA}	0.184 (0.037)	0.230 (0.027)	0.243 (0.028)	**0.249** (0.030)
	Δr_{PA}(%)	0	25.0	31.9	**35.1**
	Bias (*b*)	**0.858** (0.181)	0.788 (0.094)	0.828 (0.094)	0.853 (0.104)
	MSEP	0.231 (0.004)	0.227 (0.004)	0.225 (0.004)	**0.224** (0.004)
SG	r_{PA}	0.365 (0.030)	0.413 (0.022)	0.426 (0.019)	**0.437** (0.016)
	Δr_{PA}(%)	0	12.9	16.7	**19.5**
	Bias (*b*)	1.007 (0.121)	0.925 (0.055)	0.958 (0.061)	**0.998** (0.050)
	MSEP	0.552 (0.014)	**0.417** (0.010)	0.432 (0.009)	0.434 (0.007)
PH	r_{PA}	0.235 (0.044)	0.468 (0.022)	0.469 (0.021)	**0.477** (0.022)
	Δr_{PA}(%)	0	99.1	99.4	**102.9**
	Bias (*b*)	0.791 (0.155)	0.865 (0.051)	0.884 (0.056)	**0.902** (0.055)
	MSEP	35.0 (1.0)	29.0 (0.9)	28.9 (0.9)	**28.6** (0.9)
FT	r_{PA}	0.156 (0.064)	0.374 (0.020)	0.352 (0.031)	**0.376** (0.023)
	Δr_{PA}(%)	0	139.6	125.2	**140.6**
	Bias (*b*)	0.448 (0.196)	**0.748** (0.047)	0.708 (0.066)	0.747 (0.051)
	MSEP	0.929 (0.042)	0.802 (0.019)	0.823 (0.030)	**0.801** (0.021)

[a] Using **K** matrices constructed with the specific optimal weights given in Table 2

Statistically significant increases of r_{PA} were obtained by models exploiting genomic relationships alone or combined with pedigree in most cases (Table 3). Exceptionally, the differences with A-BLUP became significant for GY predictions within families only when

genealogy information was included in the genomic model. In general, the use of genome-wide information caused larger relative gains in r_{PA} for among-family predictions. The inclusion of marker-based relationships had a higher impact on improving r_{PA} for PH and FT, which were the traits with higher heritabilities. Models combining information from **A** and **G** consistently outperformed the basic G-BLUP in predictive ability. These improvements were statistically significant only for SG in both prediction scenarios. The highest predictive abilities were always achieved with K-BLUP and this model exhibited the lowest MSEP in most situations. Moreover, including a residual polygenic component through the optimal **K** matrix produced the largest reductions of bias relative to G-BLUP in most cases. Although differences between AG-BLUP and K-BLUP were significant only for FT predictions among-families, the latter model was slightly superior in predictive ability and prediction bias across all traits and both cross-validation schemes. The relative benefit of using the best-predictive **K** matrix was more evident when predictions relied on among-family information.

3.4. Discussion

Advances in genotyping technology have facilitated the implementation of genomic selection for several plant species. However, for some strategic cereal crops, especially sorghum, efforts are still needed to attain a full insight into the prospects of this genetic evaluation method. Here we present a first comprehensive study addressing the potentialities of exploiting pedigree and genome-wide marker information to enhance prediction of parental BVs in sorghum. The idea of combining different kinship matrices for genomic prediction has also been introduced in the context of reproducing kernel Hibert spaces regression using multiple kernels (de los Campos et al. 2010; Gianola and Schön 2016). This method provides a flexible framework, including the possibility of using non-linear combinations of kinship matrices (Corrada Bravo et al. 2009; Gianola and de los Campos 2008). Our research is based on the BLUP method as it is easy to understand, compute and implement in available mixed model softwares. Moreover, several studies have shown that this method performs comparably to other models, such as the Bayesian alternatives, for prediction of complex quantitative traits in crops (e.g. Heslot et al. 2012; Wimmer et al. 2013).

Results showed that the use of genomic information consistently improved predictive ability and accuracy of prediction in sorghum relative to the classical pedigree-based method. This is expected because the **G** matrix accounts for Mendelian segregation of alleles, distinguishing between full-sib lines that are more or less related than expected due to random chance. Therefore, genomic-estimated BVs of unphenotyped full-sibs will reflect genetic differences caused by Mendelian sampling, while these sibs will have identical pedigree-based BVs reflecting only mid-parent genetic contributions. Similarly, specific pairs of lines from unrelated families may share more alleles than expected by chance and will have

realized relationships different from zero. These features make the **G** matrix potentially more informative than **A** to better approximate genetic relationships between parental lines. The advantage of replacing pedigree by genomic-based similarities in prediction models has also been reported for other crops (e.g., Albrecht et al. 2011; Burgueño et al. 2012; Auinger et al 2016). Less conclusive results regarding the relative value of **A** and **G** for prediction in sorghum were reported by Hunt et al. (2018). However, their study is not strictly comparable to the present research due to different predictive contexts. Hunt et al. (2018) used phenotypes of testcross progenies derived from crosses with a common tester. Consequently, the total genetic value of hybrids was predicted since general and specific combining ability effects could not be distinguished. In contrast, our study aimed to predict the additive genetic merit of parental lines given that the availability of testcross data using several testers allowed averaging out most dominance deviations. Moreover, predictions reported by Hunt et al. (2018) were based on separate analyses of individual trials, whereas we obtained across-environments predictions using multi-environment trial analysis.

Before combining pedigree and genomic information in our prediction models, marker-based relationships were adjusted to take into account the difference in scale between **A** and **G**. While relationships in **A** are defined in relation to the founder population of the pedigree, the reference population for relationships in **G** is automatically set to the genotyped individuals when current allele frequencies are used (VanRaden 2008; Hayes et al. 2009). Therefore, if the population genotyped has undergone drift or strong selection, which is usually the case in plant breeding programs, its average breeding value will be different and the genetic variance would be expected to be reduced relative to the founder breeding population (Legarra et al. 2014). Several studies using real and simulated data have shown that re-scaling the **G** matrix improved predictions to a mild degree when not all individuals in the reference population were genotyped (Forni et al. 2011; Vitezica et al. 2011; Christensen et al. 2012). In contrast, the adjustment of **G** in our work, where genotypes were available for the entire training population, did not affect predictions (not shown). However, the rescaling did affect the magnitude of genomic variances, and thus changed heritability estimates (Table 1). These heritabilities derived from the adjusted **G** are thought to reflect a compatible scale for genomic and pedigree-based estimates of genetic variability. Consequently, the correction of **G** provided a clearer theoretical framework for interpretation of parameter estimates when pedigree and SNP information are used simultaneously. Even though inconsistencies between the theory underlying classical polygenic models and recent genomic approaches have received increasing attention in animal breeding, this topic has been largely overlooked in crop applications.

In order to investigate the benefits of adding pedigree information into G-BLUP, we used a kinship matrix based on a weighted linear combination of pedigree and marker-based relationships, constructed as $\mathbf{K} = w\mathbf{A} + (1 - w)\mathbf{G}$. VanRaden (2008) and Goddard et al. (2011)

proposed deterministic methods to predict the appropriate *w* based on the error variance of the true genomic relationships or on the effective number of independent chromosomal segments, respectively. These methods and subsequent suggested alternatives provide variable theoretical estimates, and these can be substantially different from optimal data-dependent weights (Ilska et al. 2017). Here, we adopted an analytical approach where the best-predictive *w* was empirically defined for each trait and cross-validation scenario after evaluating the changes in prediction quality over a sequence of candidate weights. The same approach was used applying a semi-parametric Bayesian method (Rodríguez-Ramilo et al. 2014) and in the context of multiple-trait prediction (Momen et al. 2017) in animal species. Our results in sorghum showed a general agreement between the best-predictive weights based on cross-validation and the weights that maximized the fit of the model to the entire breeding population. This suggests that the goodness of fit can be used as guiding tool to attain an optimal prediction model. Our study demonstrates, however, that the best-fitting model may not produce the best predictive performance. This topic will be specifically addressed later in the present section when discussing the differences between K-BLUP and AG-BLUP.

In our research, the optimal weighting factor *w* varied across traits. This implies that the optimal similarity matrix, from a predictive perspective, is actually trait-specific. According to the optimal weights for PH and FT (Table 2), a large proportion of the total additive variation in these traits was captured by SNPs (between 80% and 90%). This is expected since our sorghum breeding population is highly structured, with strong family relationships and small effective population size. Consequently, the genetic variance explained by markers might not only be a result of SNPs located on causative genes or in LD at population level, but it might mainly depend on SNPs capturing familial relationships between lines (Habier et al. 2007). For GY and SG, lower amounts of additive genetic variance were explained by markers (between 40% and 60%) and higher weights on pedigree were generally required to optimize predictions. The lower levels of variance accounted for by markers may be due to the more complex genetic information driving these low-heritability traits, which was not totally captured by imperfect coverage of available SNPs. The higher importance of including pedigree information for GY and SG is reflected by the larger predictive improvements achieved from using K-BLUP instead of G-BLUP (Table 3). One reason may be that, when some markers are not in LD with QTLs, the addition of pedigree information contributes to capturing associations between causative alleles due to common ancestral identity, improving predictions. Furthermore, besides residual LD patterns explained by family structure, Jensen et al. (2012) pointed out that an additional polygenic component can also take into account potential LD across chromosomes. In agreement with our results, Liu et al. (2011) found that higher weights placed on pedigree information were required to optimize predictions of traits with lower heritabilities in dairy cattle. The use of trait-specific optimal weights has not only been recommended for genomic selection in animals (Liu et al. 2011; Gao et al. 2012), but

also in wheat (Ashraf et al. 2016). In the latter study, identification of the optimal weighting terms was based on likelihood curves. The authors pointed out that differing weights had relatively small effect on improving the likelihood. We arrived at similar deductions when the weights were tuned to maximize predictive ability. Our results show that, despite specific values of *w* were identified for each trait, a weighting factor between 0.4 and 0.5 would generally perform well across traits, with predictive performance being clearly sub-optimal only for FT predictions across families (Fig. 1). It should be considered, however, that these results are partly dependent on the marker density used.

In this paper we compare a BLUP model blending pedigree and genomic information into a single matrix **K** with an equivalent model that fits **A** and **G** separately. While the latter approach has been previously used for genomic prediction in crops, the K-BLUP method had not been explored until now for plant breeding applications. Results show that the benefits of using K-BLUP instead of AG-BLUP were generally marginal but fairly consistent across traits and prediction scenarios. The predictive performance of AG-BLUP is expected to be sub-optimal compared to that of K-BLUP. The reason is that the weight used by AG-BLUP for prediction is the one that best fits the genotypes and phenotypes of the TS. However, this *w* is not necessarily an accurate estimator of the weight that optimizes prediction of phenotypes in the VS. Therefore, the likelihood-based *w* may contain some information that is only relevant for the reference lines, but with little predictive value for the validation lines. On the other hand, within the K-BLUP approach, *w* is empirically derived to optimize prediction of lines in the VS. Then, the resulting best-predictive weight makes no direct reference to the TS data, but it seems to be closer to the *w* that best approximates the genetic variability among validation lines. This would explain why K-BLUP slightly increased predictive ability and reduced over-prediction relative to AG-BLUP in all cases. Our results are consistent with those obtained by Ilska et al. (2017) in chicken, who found that the increased goodness of fit in the training set was accompanied by decreased accuracy and higher bias of predictions in the validation set.

Besides the difference in predictive performance, the fact that information from the two sources is conveyed by fitting a single relationship matrix provides additional benefits regarding model applicability. In our study, computational time was reduced by more than 35% when fitting **K** instead of **A** plus **G** (not shown). In addition, K-BLUP is likely to produce more stable results when using small training sets since fewer variance components have to be estimated. We applied relatively simple prediction models, but differences in computational burden and stability between AG-BLUP and K-BLUP are prone to increase when more elaborate models are used. For instance, when prediction models include interactions of genotypes with environmental factors or when multi-trait prediction is aimed. Finally, the use of a blended matrix prevents from potential convergence problems resulting from collinearity between pedigree-estimated genetic effects and genomic effects when **A** and

G are fitted separately, which may also deteriorate the quality of predictions (Legarra et al. 2008).

In our study, the main criterion used to define the optimal weighting factor for K-BLUP was the predictive ability. Given that all selection candidates belong to the same generation, we considered that this was the criterion to maximize, since prediction bias should not be too strong. It is noteworthy that, in this research, the highest predictive ability and the smallest empirical bias were rarely achieved by the same model (Table 3). However, inflation or over-prediction ($b < 1$) should also be considered when searching for the best predictor. For instance, inflation of genomic predictions can be detrimental for genetic gain since the genetic merit of new genotyped lines would be overestimated when compared to older lines that have undergone testcross field evaluations. Our results revealed that including pedigree information through the blended matrix **K** always caused the largest reductions in the inflation produced by G-BLUP. Reduced bias of predictions obtained by increasing the weight on familial relationships has also been reported in animal genomic evaluations (Liu et al. 2011; Gao et al. 2012). In addition, the lowest MSEP were mostly achieved by K-BLUP in our research. The minimization of MSEP has been recommended as an appropriate evaluation criterion for the comparison of prediction models since it considers both bias and precision (Vitezica et al. 2011; González-Recio et al. 2014).

We evaluated model performance when within-family or across-family information was used for prediction. For all models, predictions were clearly better when full-sib relationships were exploited. The value of using information from close relatives for prediction in structured populations is consistent with previous studies in animal and plant breeding (e.g. Wientjes et al. 2013; Schopp et al. 2017). Across traits, the decline in predictive ability was less marked for models including **G** when predictions were based on relationships among families (a mean decrease of 21% vs 40% for A-BLUP). This reflects the capacity of marker information to capture populational LD, which becomes particularly relevant when predictions rely on more distant genetic relatedness (Habier et al. 2007). As shown in Table 3, however, the relative benefit of using genomic information for among-family prediction was less evident for GY and SG. The latter finding emphasizes the contribution of additive genetic relationships and tempers that of LD for predicting traits with lower heritability.

3.5. Conclusion

This paper provides a first empirical evidence based on sorghum breeding data that the use of genomic relationships alone, even with relatively low marker density, can give better predictions of parental BVs than the pedigree-based model. We also investigated how the use of different combinations of pedigree and genomic information affected prediction quality. Our results showed that using a kinship matrix integrating both sources of information yielded

better predictive performance than G-BLUP for different traits and prediction scenarios. Identification of the optimal weighting factor used to combine familial and marker-based relationships was driven by the search for maximizing predictive ability. Under this approach, the weight that optimizes predictions differed between traits. These weights were generally consistent with the weights that optimized model fitting to the entire dataset. The impact of including genealogy information to improve genomic predictions was stronger for traits with lower heritability, such as grain yield and stay-green. Findings of this paper might be relevant for other breeding programs with limited genotyping resources and when lowly heritable traits are the main targets of selection.

Author Contribution Statement

JV, MM, DJ and FvE designed the research. JV performed statistical analyses and wrote the paper. MM, EM, DJ and FvE edited the manuscript. DJ coordinated the experiments and data collection. CH and EM processed and prepared the dataset. All authors read and approved the final manuscript.

Acknowledgments

We thank self-identified reviewer Daniel Gianola and an anonymous reviewer for valuable suggestions to improve the content of this manuscript. The data used in this research was generated with funding from the Grains Research and Development Corporation (GRDC) of Australia. JV acknowledges financial support from the National Institute of Agricultural Technology (INTA) of Argentina, Res. DN 1126/13.

Chapter 4

Genomic prediction of grain yield and drought-adaptation capacity in sorghum is enhanced by multi-trait analysis

Velazco J.G., Jordan D.R., Mace E.S., Hunt C.H., Malosetti M., van Eeuwijk F.A.
Front Plant Sci (2019) 10:997.

Abstract:
Grain yield and stay-green drought adaptation trait are important targets of selection in grain sorghum breeding for broad adaptation to a range of environments. Genomic prediction for these traits may be enhanced by joint multi-trait analysis. The objectives of this study were to assess the capacity of multi-trait models to improve genomic prediction of parental breeding values for grain yield and stay-green in sorghum by using information from correlated auxiliary traits, and to determine the combinations of traits that optimize predictive results in specific scenarios. The dataset included phenotypic performance of 2645 testcross hybrids across 26 environments as well as genomic and pedigree information on their female parental lines. The traits considered were grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT). We evaluated the improvement in predictive performance of multi-trait G-BLUP models relative to single-trait G-BLUP. The use of a blended kinship matrix exploiting pedigree and genomic information was also explored to optimize multi-trait predictions. Predictive ability for GY increased up to 16% when PH information on the training population was exploited through multi-trait genomic analysis. For SG prediction, full advantage from multi-trait G-BLUP was obtained only when GY information was also available on the predicted lines per se, with predictive ability improvements of up to 19%. Predictive ability, unbiasedness and accuracy of predictions from conventional multi-trait G-BLUP were further optimized by using a combined pedigree-genomic relationship matrix. Results of this study suggest that multi-trait genomic evaluation combining routinely measured traits may be used to improve prediction of crop productivity and drought adaptability in grain sorghum.

4.1. Introduction

Water scarcity in rain-fed cropping systems is a major challenge to a world of increasing food demand (UNCTAD, 2011). In this context, grain sorghum (*Sorghum bicolor* L. Moench) is a cereal crop that can play an important role for sustainable farming, as it is particularly resilient to stress conditions caused by drought and erratic rainfall (Paterson et al. 2009). This crop is a staple food in semi-arid regions of the world and is used for feed globally (Acquaah 2012). Grain yield is the primary trait in sorghum, as it is a key measure of crop productivity and profitability of farmers. Another important trait is stay-green, a complex drought-adaptation mechanism associated with increased yield in environments where post-flowering drought occurs frequently (Borrell et al. 2000; Jordan et al. 2003). Accordingly, most efforts for increasing genetic progress in grain sorghum should concern both characteristics (Jordan et al. 2012; Borrell et al. 2014). Further improvement in productivity and drought adaptability requires knowledge-based selection strategies that efficiently exploit available phenotypic and genotypic information in sorghum breeding programs.

Selection of complex quantitative traits can be based on statistical methods that combine phenotypes and high-density marker data to predict genetic merit. This form of marker-assisted selection, known as genomic selection (Meuwissen et al. 2001), has been successfully implemented in animal and plant breeding (Meuwissen et al. 2016; Crossa et al. 2017). Several genomic prediction methods have been developed, including Bayesian regression (Gianola et al. 2009) and semiparametric approaches (de los Campos et al. 2010). An alternative method for genomic prediction within the linear mixed model framework is termed genomic best linear unbiased prediction or G-BLUP (VanRaden 2008). This method is usually preferred in practice because it is simple to implement and computationally less demanding than competing procedures (Gianola et al. 2014). Moreover, G-BLUP is expected to perform similar to other models for prediction of complex agronomic traits such as grain yield (Heslot et al. 2012; Wimmer et al. 2013), which are typically affected by a large number of small-effect genes (Schön et al. 2004). Independent of the model used, one of the main features of genomic selection is that genetic merit can be predicted for selection candidates that have not yet been phenotyped. This application is particularly promising for reducing evaluation cost and generation interval in the sorghum breeding pipeline, where parental lines of commercial hybrids are currently selected on the base of extensive field progeny testing. Moreover, development of female lines as well as hybrid seed production is based on the use of cytoplasmic-genetic male sterility, which requires extra time and human resources. Despite the potential of genomic selection to increase rates of genetic gain in sorghum, studies on the application of genomic models are limited compared to other cereal crops, such as maize, wheat and rice (Kulwal 2016). A first genomic selection study in grain sorghum was reported by Hunt et al. (2018) for prediction of test-cross performance in individual trials. Velazco et

al. (2019b) investigated different genomic models including pedigree information for across-environment prediction of parental breeding values in productivity and adaptability traits.

Most of genomic selection studies, including the ones mentioned above, have been based on separate analysis of individual traits. However, selection decisions in plant breeding programs typically rely on several measured characters. The joint analysis of multiple traits (MT) can increase the accuracy of genetic evaluations relative to single-trait (ST) analysis by exploiting information from correlated characters (Henderson and Quaas 1976). The potential gain in accuracy depends on the strength of genetic and environmental correlations between traits. The benefit is expected to increase for lowly heritable traits, when analyzed together with strongly correlated traits of higher heritability (Thompson and Meyer 1986). Additionally, MT models are able to reduce selection bias or culling bias introduced by contemporary or sequential selection on correlated traits, which are ignored by ST analysis (Mrode 2005). The extension of MT analysis into the context of genomic prediction methods has been studied using real and simulated data (e.g., Calus and Veerkamp 2011; Jia and Jannink 2012). MT genomic models can be used to combine information from correlated traits and from relatives in an efficient way. When the breeder is interested in response to selection of a single target trait, other secondary or auxiliary traits can be incorporated in the prediction model to provide additional information on the primary character. Consequently, more phenotypes recorded on the reference population can be potentially exploited to assist predictions of the target trait in the testing population.

Genome-based MT analysis has been applied for breeding in hybrid crops of other major cereals like maize, rice and wheat (e.g., dos Santos et al. 2016; Wang et al. 2017; Schulthess et al. 2018). In sorghum, MT genomic prediction has been implemented only in biomass-type genotypes using a pre-breeding population (Fernandes et al. 2018). Here, we present a first study on the potential of exploiting trait associations for genomic prediction in advanced breeding material of grain sorghum. Our research is developed in the context of prediction for broad adaptation using testcross performance data across dryland sorghum production environments in Australia.

The aims of this study were to investigate if multi-trait analysis improves across-environment genomic predictions for grain yield and stay-green in sorghum, and to identify the combinations of traits that optimize results in different prediction scenarios. To attain these objectives, the optimal combination of traits was empirically determined for each scenario by evaluating the gain in prediction quality of alternative MT models relative to the ST model. In addition, we explored if the performance of best-predictive MT genomic models can be further enhanced by incorporating pedigree information.

4.2. Material and Methods

4.2.1. Data

The dataset used in this study is part of the sorghum breeding program of female parental lines conducted by the University of Queensland and the Department of Agriculture and Fisheries in Queensland, Australia. The phenotypic records consisted of 26 testcross performance trials where a total of 646 female lines were tested in hybrid combinations across 12 locations between 2008 and 2014. Phenotypes of 2645 testcross hybrids were used to assess female lines performance across a target population of environments (TPE) covering the main sorghum cropping region of Australia. The series of trials belongs to an advanced stage of yield testing (AYT), where measurements are taken from relatively large plots. More details on field layout and structure of the dataset are given in Velazco et al. (2019b). Four productivity and adaptability traits routinely measured in advanced testing were considered for this study: grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT). GY is the main target trait with direct economic value driving selection. SG is an integrated drought-adaptation trait that is expressed as delayed leaf senescence, which is a consequence of improved water balance in the plant under post-flowering drought stress (Borrell et al. 2014). This functional SG phenotype is also considered an important trait since it is associated with enhanced crop productivity in water-limited seasons (Jordan et al. 2003; 2012). Given that SG expression depends on the occurrence of terminal drought conditions, records of this trait were available at 9 trials and for 603 lines in the present dataset. While PH and FT are mainly selected in earlier breeding stages to reduce extreme variation, these traits are considered in advanced testing to ensure appropriate agronomic type for commercial production (Jordan et al. 2011).

All the female lines were genotyped using an integrated DArT and genotyping-by-sequencing (GBS) methodology involving complexity reduction of the genomic DNA to remove repetitive sequences using methylation sensitive restriction enzymes prior to sequencing on Next Generation sequencing platforms (DArT, www.diversityarrays.com). The sequence data generated were then aligned to the most recent version (v3.1.1) of the sorghum reference genome sequence (Paterson et al. 2009; McCormick et al. 2018) to identify SNP (Single Nucleotide Polymorphism) markers. SNPs with minor allele frequency lower than 2.5% or more than 20% of missing values were discarded. Missing genotypes were imputed based on random sampling from marginal allele distributions using the synbreed package (Wimmer et al. 2012) in R (R Core Team, 2018). After quality filtering, 4781 evenly-spaced SNPs remained for analysis.

Inbred parent lines were derived from pedigree breeding methods resulting in a highly structured breeding population. Pedigree data was available for the female lines and 499 ancestors tracing back 28 generations.

4.2.2. Single-trait analysis

Univariate analysis of each trait was performed within the REML-based mixed model framework using a stage-wise approach. In the first stage, adjusted testcross means were computed per trial after correcting for design factors and spatial field variation. For this, we used a novel flexible method for spatial analysis of individual trials based on P-splines (see Rodríguez-Álvarez et al. 2018a; Velazco et al. 2017 for details). The specific model applied in the first stage is described in Velazco et al. (2019b).

In the second stage, spatially-adjusted testcross means from each trial were jointly analyzed to estimate line means across testers and environments. The model is as follows:

$$y_{ijk} = \mu + L_i + M_j + E_k + LM_{ij} + LE_{ik} + ME_{jk} + LME_{ijk}, \quad (1)$$

where y_{ijk} represents the best linear unbiased estimation (BLUE) of the i-th female line crossed with the j-th male tester in the k-th environment, which was fitted by a random line effect (L_i), a fixed male tester effect (M_j), a fixed environmental effect (E_k), and all possible interactions between these factors. Since line effects were considered random, all the interactions involving L_i were random, while ME_{jk} was fixed. Note that LME_{ijk} includes the error of genotype mean. All random effects were assumed independent homoscedastic and normally distributed with zero mean.

Due to shrinkage properties of BLUP (Robinson 1991), random line effects are contracted towards the mean of the line population. Given that not all lines were crossed with the same number of testers or evaluated in the same number of trials, the amount of shrinkage is different for each BLUP of L_i. Moreover, using BLUP(L_i) as response variable in the genomic prediction model is problematic because it would result in double shrinkage of predicted breeding values. Therefore, to eliminate shrinkage in line BLUPs before the genomic prediction stage, we applied the deregression procedure of Garrick el al. (2009). This correction relies on individual reliabilities of BLUP(L_i), as obtained by inverting the coefficient matrix of the mixed model equation (Meyer 1989).

The following G-BLUP model was used to predict parental breeding values of female lines from progeny performance:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e}, \quad (2)$$

where the vector **y** contains deregressed BLUPs of L_i derived from the second stage; **1** is a vector of ones with associated general mean μ; **Z** is a design matrix allocating deregressed BLUP(L_i) to genomic effects; **g** is the vector of additive genomic effects with distribution $\mathbf{g} \sim N(0, \mathbf{G}\sigma^2_g)$, where σ^2_g is the additive genomic variance and **G** is the genomic relationship matrix as computed with the first method of VanRaden (2008); and **e** is the vector of residuals

assuming $\mathbf{e} \sim N(0, \mathbf{D}\sigma^2_e)$, where σ^2_e is the residual variance and $\mathbf{D}$ is a diagonal weighting matrix accounting for heterogeneous residual variances due to differences in individual reliabilities of deregressed BLUP(L_i) (Garrick et al. 2009; VanRaden 2008).

Given the variance components from the single-trait prediction model, the narrow-sense heritabilities (h^2) of line means were obtained as: $h^2 = \sigma^2_g/(\sigma^2_g + \sigma^2_e)$. Note that σ^2_e comprises non-additive genetic effects and true errors associated with line mean estimates.

4.2.3. Multi-trait analysis

For joint analysis of multiple traits, models (1) and (2) were extended using multivariate mixed models. We present a general formulation for any combination among traits T = (GY, SG, PH, FT). The multi-trait case of model (1) can be represented in vector notation as:

$$\mathbf{y}_{ijk} = \boldsymbol{\mu} + \mathbf{L}_i + \mathbf{M}_j + \mathbf{E}_k + \mathbf{LM}_{ij} + \mathbf{LE}_{ik} + \mathbf{ME}_{jk} + \mathbf{LME}_{ijk}, \quad (3)$$

where in this case $\mathbf{y}_{ijk}$ is a vector collecting spatially-adjusted BLUEs of multiple traits T from separate univariate analyses in stage one; and $\mathbf{L}_i$, $\mathbf{LM}_{ij}$, $\mathbf{LE}_{ik}$, $\mathbf{LME}_{ijk}$ are the vectors of multi-trait random effects with respective assumed distributions $\mathbf{L}_i \sim MVN(\mathbf{0}, \mathbf{I}_L \otimes \boldsymbol{\Sigma}_L)$, $\mathbf{LM}_{ij} \sim MVN(\mathbf{0}, \mathbf{I}_{LM} \otimes \boldsymbol{\Sigma}_{LM})$, $\mathbf{LE}_{ik} \sim MVN(\mathbf{0}, \mathbf{I}_{LE} \otimes \boldsymbol{\Sigma}_{LE})$, $\mathbf{LME}_{ijk} \sim MVN(\mathbf{0}, \mathbf{I}_{LME} \otimes \boldsymbol{\Sigma}_{LME})$, where $\boldsymbol{\Sigma}_q$, for $q = L, LM, LE, LME$, is a covariance matrix among traits and $\otimes$ is the Kronecker product operator. For all random effects, the matrix $\boldsymbol{\Sigma}_q$ was modelled as unstructured, allowing for unequal variances across traits and specific covariances for each pair of traits.

The multi-trait G-BLUP model was defined as:

$$\mathbf{y}_T = \mathbf{1}_T\mu_T + \mathbf{Z}_T\mathbf{g}_T + \mathbf{e}_T,$$

where the subscript refers to traits T; $\mathbf{y}_T$ is now a multi-trait vector of deregressed BLUP(L_i), ordered as lines within traits, obtained from joint multivariate analysis using model (3); $\mathbf{g}_T$ is the vector of multi-trait additive genomic effects with distribution $\mathbf{g}_T \sim MVN(\mathbf{0}, \mathbf{G} \otimes \boldsymbol{\Sigma}_g)$; and the multivariate residual effects were assumed $\mathbf{e}_T \sim MVN[\mathbf{0}, (\mathbf{I}_e \otimes \boldsymbol{\Sigma}_e)\mathbf{D}_T]$, where matrix $\mathbf{D}_T$ has diagonal elements containing weights based on individual reliabilities of deregressed BLUP(L_i) for each trait, as obtained from the second stage of multi-trait analysis (model 3). The covariance matrices among traits for additive genomic effects ($\boldsymbol{\Sigma}_g$) and residuals effects ($\boldsymbol{\Sigma}_e$) were assumed unstructured.

We also considered multi-trait prediction models fitting a kinship matrix that combines pedigree and genomic information. For this, we extended the BLUP method advocated by Velazco et al. (2019b) to the multi-trait context. The method, referred as K-BLUP, uses a

combined kinship matrix formed as $\mathbf{K} = w\mathbf{A} + (1 - w)\mathbf{G_s}$, where $\mathbf{A}$ is the numerator relationship matrix among lines computed from the full pedigree data and $\mathbf{G_s}$ is a scaled matrix $\mathbf{G}$ that is compatible with $\mathbf{A}$ in reference to the base breeding population (Vitezica et al. 2011; Christensen 2012). The weighting term w can be interpreted as the fraction of additive genetic variance that is not captured by SNPs and is explained by familial relationships. In this case, the vector $\mathbf{g}_T$ has distribution $MVN(\mathbf{0}, \mathbf{K} \otimes \mathbf{\Sigma}_g)$, and collects additive genomic effects as well as residual polygenic effects. Under our approach, the optimal w is empirically determined in order to maximize predictive ability after evaluating a sequence of candidate weights within the range $0 < w < 1$. To assess the benefits of including additional pedigree information, the best-predictive multi-trait model using conventional G-BLUP (i.e., setting $w = 0$) was compared to the equivalent optimized multi-trait K-BLUP model.

4.2.4. Model fitting

Spatial analyses in the first stage were performed with the R package SpATS (Rodríguez-Álvarez et al. 2018b), which is publicly available from CRAN (https://cran.r-project.org/package=SpATS). Uni- and multivariate models in the second and the genomic prediction stages were fitted by REML using the average information algorithm as implemented in ASReml-R (Butler et al. 2017). We used the correlation form parameterization of the unstructured covariance matrix available in ASReml-R to obtain direct estimates of trait correlations and corresponding standard errors.

4.2.5. Prediction scenarios

To assess the value of multi-trait genomic analysis in sorghum, we considered the general predictive strategy where a single trait is the primary target of prediction (selection) and other auxiliary characters are used to potentially improve predictions of the target trait. Predictive performance of MT models was evaluated through five-fold cross-validation, where 80% of lines were randomly assigned to the training lines set (TLS) and the remaining 20% formed the validation lines set (VLS). We explored two potential scenarios for MT prediction. One scenario assuming that (new) lines in the VLS have not been evaluated in the field for any trait; therefore, phenotypes of auxiliary traits (and the target trait) were available only for the TLS (Aux@TL). A second MT prediction scenario assumes that lines of the VLS have been evaluated in the field, but for traits other than the target; therefore, phenotypes of auxiliary traits were available for both TLS and VLS (Aux@TL+VL). For comparison, we also considered ST prediction, where only records of the target trait in VLS are used for prediction (Figure 1).

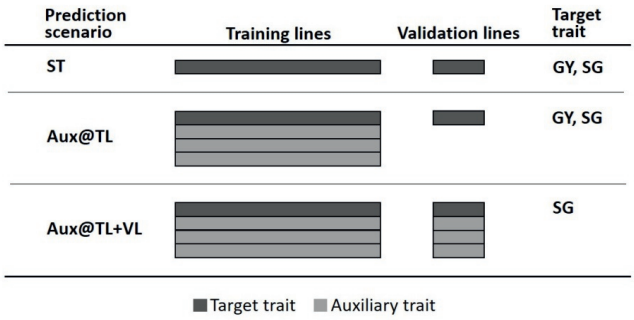


Fig. 1 Five-fold cross-validation schemes representing three prediction scenarios: ST, where only data of the target trait in the training lines set (TLS) are used for prediction in the validation lines set (VLS); Aux@TL, where data of auxiliary traits in TLS are included for prediction of the target trait in VLS; and Aux@TL+VL, where data of auxiliary traits in both TLS and VLS are included for prediction of the target trait in VLS.

The first MT prediction scenario was studied for two cases: when GY or SG was the trait of interest. However, the Aux@TL+VL scenario was considered only for SG as target trait. Given that one of the main purposes of the trial series used in this study is to obtain representative measurements of GY productivity, the fact that lines are field-tested for other traits, but not for GY, would imply atypical circumstances seriously affecting GY during an experiment. Moreover, extremely reduced or aborted grain production affects normal canopy development; hence, data of traits such as PH or SG should be discarded or considered with extreme caution. For these reasons, the Aux@TL+VL scenario was not considered when GY was the target of prediction, as it would represent an unrealistic or exceptional situation in the context of an advanced stage of sorghum testing.

4.2.6. Validation of genomic prediction

In each prediction scenario, the performances of all possible MT models combining two, three and the four traits were compared to that of the ST model, with the latter taken as benchmark. The response variables used for validation of genomic predictions were consistent with the models applied for analysis: $\mathbf{y}_T$ and $\mathbf{y}$ were used to validate $\tilde{\mathbf{g}}_T$ and $\tilde{\mathbf{g}}$, respectively. Model validation was evaluated by considering the predictive ability, unbiasedness and accuracy of predictions in the VLS. Predictive ability was measured as the Pearson's correlation (r_{PA}) between $\tilde{\mathbf{g}}_T$ ($\tilde{\mathbf{g}}$) and $\mathbf{y}_T$ ($\mathbf{y}$). Unbiasedness of genomic predictions was measured by the regression coefficient (b) of $\mathbf{y}_T$ ($\mathbf{y}$) on $\tilde{\mathbf{g}}_T$ ($\tilde{\mathbf{g}}$), where $b = 1$ indicates an empirically unbiased predictor. Accuracy of predictions was assessed by computing the mean squared error of prediction (MSEP) from the linear regression. The evaluation of prediction models was based

on average values over 20 replicates of each cross-validation scenario using the same random seed for all models.

4.3. Results

The narrow-sense heritability estimates using ST G-BLUP are given in Table 1 along with the additive genetic and residual correlations between traits estimated by MT G-BLUP. Heritability estimates ranged from 0.36 to 0.76, with GY presenting the lowest h^2. The heritability of SG was higher than that of GY, but lower than those of PH and FT. Additive genetic correlations were significant only for GY with SG and PH, while FT was statistically uncorrelated with the other traits. Significant residual correlations were estimated between GY and the rest of the traits as well as between PH and FT.

Table 1 Heritabilities (diagonal; in parentheses), additive genetic (above diagonal) and residual (below diagonal) correlations[a] for grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT).

Trait	GY	SG	PH	FT
GY	(0.36)	**0.52**	**0.66**	-0.01
SG	**0.36**	(0.50)	0.03	-0.02
PH	**0.71**	0.04	(0.76)	-0.07
FT	**-0.48**	0.04	**-0.29**	(0.65)

[a] Boldfaced correlations are significant based on approximate 95% confidence interval (Holland 2006).

Figure 2 shows measures of predictive performance from MT G-BLUP when additional information from auxiliary traits in the TLS was used to predict GY or SG (Aux@TL scenario). MT models that included PH data alone or combined with other auxiliary traits increased predictive ability for GY, compared with the ST model. The highest predictive ability represented a 16% increase and it was achieved by combining GY, SG and PH information. Regression coefficients (b) for all models were below 1 for GY, which indicates overestimation of genomic predictions. The average empirical bias of GY predictions from the ST model was slightly reduced only by the MT model including PH alone as auxiliary trait. The relative decrease (%) in MSEP by MT models was consistent with their increases in predictive ability; MT G-BLUP exploiting PH information had a greater impact on improving accuracy of genomic prediction for GY. When SG was the target of prediction, ST G-BLUP outperformed MT G-BLUP models in predictive ability. Even though prediction models incorporating GY data in the TLS reduced the absolute deviation of b from 1, these models tended to over-predict additive genetic values for SG ($b < 1$). The inclusion of auxiliary traits in TLS produced higher MSEP relative to using only SG data for prediction.

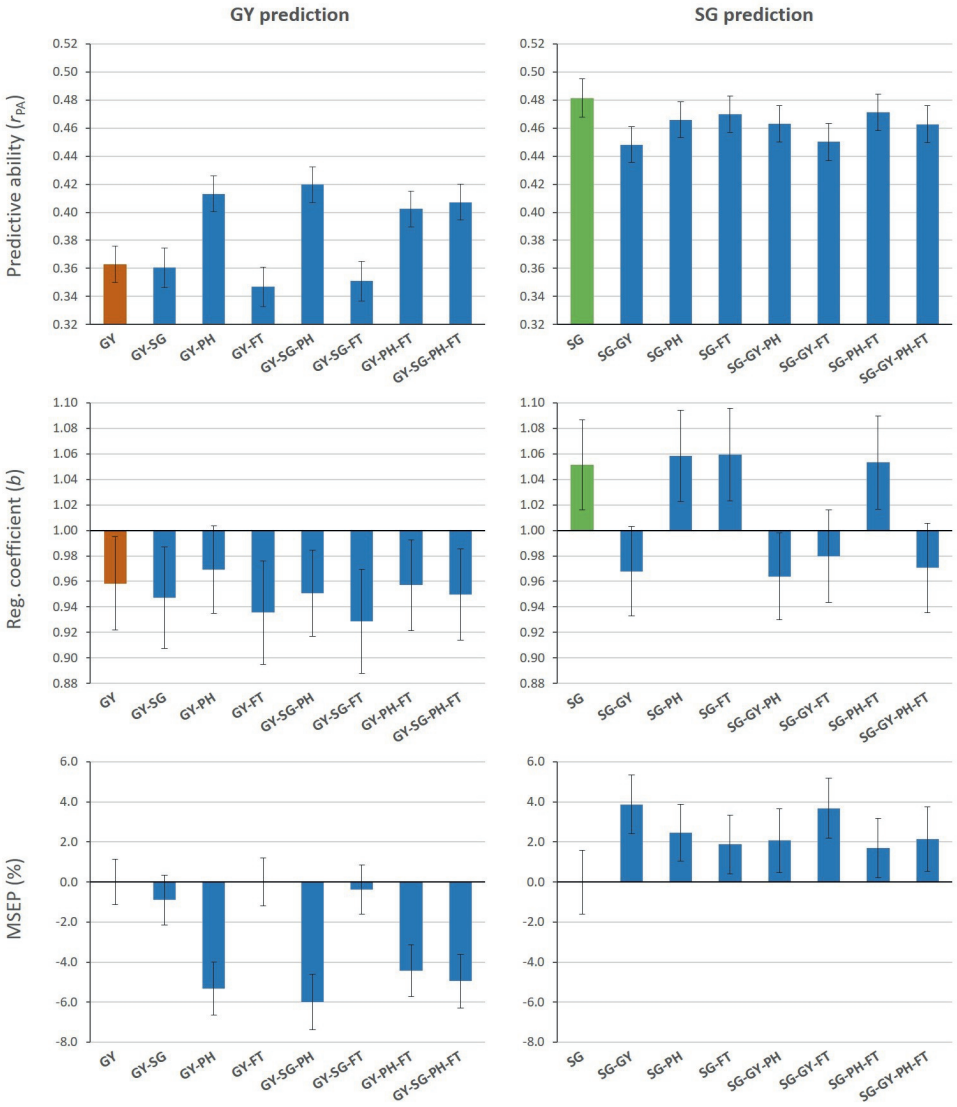


Fig. 2 Mean values (and SD of 20 replicates) for predictive ability, regression coefficient and relative MSEP from single- and multi-trait G-BLUP models using different combinations of grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT) data in the training lines set for prediction of GY (*left*) or SG (*right*).

Different results were found for SG predictions when MT G-BLUP models used auxiliary trait records on both TLS and VLS (Figure 3). Under this prediction scenario, all MT models including GY as auxiliary trait improved predictive abilities for SG. Relative increases of up to 19% in predictive ability were observed for MT models adding combined GY-PH or GY-PH-FT information from TLS and VLS. Moreover, predictions from these models tended to be less biased than those from ST G-BLUP and over-prediction was eliminated. Finally,

MSEP for SG was consistently reduced when GY information on VLS was incorporated by MT G-BLUP models.

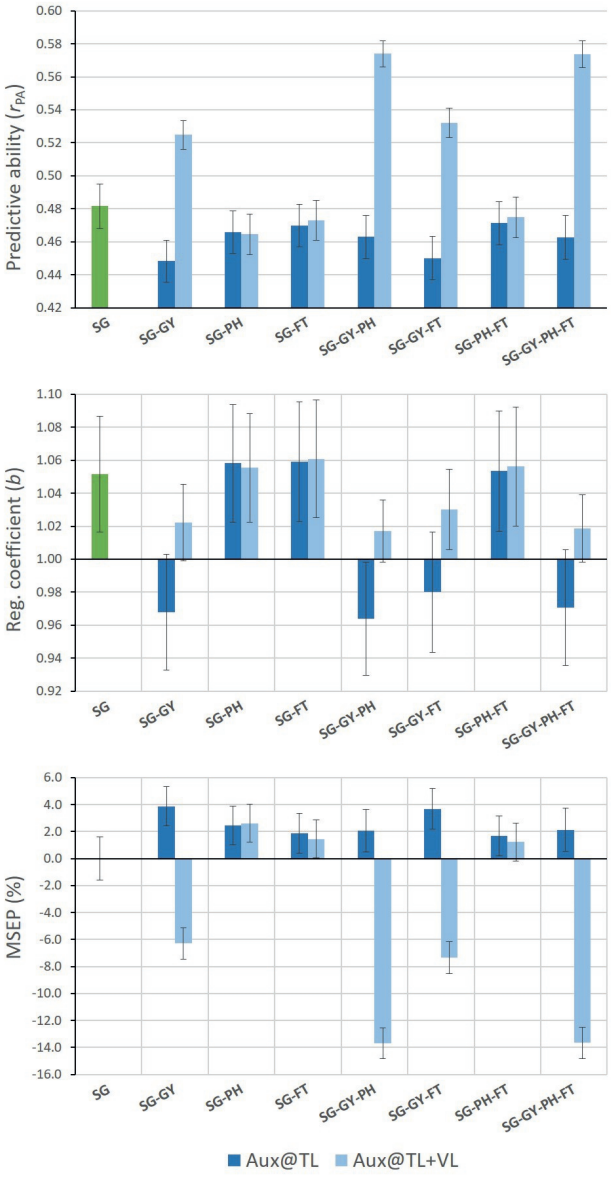


Fig. 3 Mean values (and SD of 20 replicates) for predictive ability, regression coefficient and relative MSEP for SG predictions from single- and multi-trait G-BLUP models using different combinations of grain yield (GY), stay-green (SG), plant height (PH) and flowering time (FT) data only in the training lines set (Aux@TL), and in both the training and validation lines sets (Aux@TL+VL).

After evaluating predictive performance of G-BLUP method, we explored the use of a combined pedigree-genomic matrix **K** for optimization of multi-trait genomic prediction.

Table 2 presents results from the best-predictive MT G-BLUP models and from the optimized MT K-BLUP models for the scenarios where MT analysis outperformed ST analysis (Aux@TL for GY and Aux@TL+VL for SG). In these prediction scenarios, the optimal use of combined trait information, from a predictive perspective, resulted from exploiting genetic correlations among GY, SG and PH (Figs. 2 and 3). When GY was the target trait, validation results of MT G-BLUP were improved by MT K-BLUP (with **K** using $w = 0.25$) for all evaluation criteria to a marginal extent. Under the prediction scenario for SG, bias and MSEP from multi-trait models were slightly higher when **K** (using $w = 0.30$) was used instead of **G**. However, an additional 10% gain in predictive ability was obtained by MT K-BLUP in this scenario.

Table 2 Predictive abilities (r_{PA}), regression coefficients (Bias) and relative MSEP from single-trait G-BLUP and K-BLUP, and from the best-predictive multi-trait G-BLUP and K-BLUP models for GY prediction using auxiliary traits data on training lines (Aux@TL) and for SG prediction using auxiliary traits data on both training and validation lines (Aux@TL+VL). The best value for each evaluation criterion is boldfaced.

Prediction scenario	Quality criterion	Single-trait		Multi-trait[a]	
		G-BLUP	K-BLUP	G-BLUP	K-BLUP[b]
GY: (Aux@TL)	r_{PA}	0.363 (0.013)	0.373 (0.014)	0.420 (0.013)	**0.429** (0.015)
	Bias (b)	0.958 (0.037)	0.965 (0.037)	0.951 (0.034)	**0.967** (0.039)
	MSEP (%)	0 (1.1)	−0.7 (1.1)	−6.0 (1.3)	−**6.6** (1.6)
SG: (Aux@TL+VL)	r_{PA}	0.482 (0.013)	0.508 (0.015)	0.574 (0.008)	**0.630** (0.009)
	Bias (b)	1.052 (0.035)	1.074 (0.036)	**1.017** (0.019)	1.038 (0.022)
	MSEP (%)	0 (1.6)	2.7 (1.6)	−**13.7** (1.2)	−12.0 (1.2)

[a]Multi-trait models combining GY, SG and PH information.
[b]Using $\mathbf{K} = w\mathbf{A} + (1 - w)\mathbf{G_s}$ with optimal weights $w = 0.25$ and $w = 0.30$ for GY and SG prediction, respectively.

4.4. Discussion

This study intended to establish the value of using multi-trait analysis to improve across-environment genomic predictions for grain yield and stay-green in sorghum by exploiting information from auxiliary traits. Our approach to predict is consistent with selection for broad adaptation, where the set of trials is considered to be representative of the TPE. To determine the efficiency of this predictive method, multi-trait BLUP models were evaluated in terms of prediction quality measures for two cases: when the target trait and auxiliary traits were recorded only on relatives of the predicted lines; and when records of auxiliary traits were also available for the predicted lines *per se*. Multi-trait BLUP can be seen as a generalized (linear) selection index, with the additional abilities to account for unbalanced information from any set of relatives while properly adjusting for fixed effects in the data

(Lynch and Walsh 1998; Mrode 2005). In our index, and from mixed model theory, data of several traits were optimally weighted based on the multi-trait genetic and residual covariance matrices in order to maximize accuracy of predicted genetic merit by BLUP estimation. An important decision is which auxiliary traits to include in a multi-trait model. In this article, we examined all possible combinations of available traits, and the optimal multi-trait BLUP model for each prediction scenario was empirically determined on the base of predictive outcomes.

The present study considered the general situation where the breeding objective is to improve a single character with economic value (GY or SG), while other traits would contribute to that genetic goal. This is equivalent to assigning a relative economic weight of 1 to the target trait and 0 weights to auxiliary traits in a selection index. It is worth mentioning that, in a real breeding program, SG is unlikely to be considered as an independent target of selection since its economic value is linked to its positive effect on GY and reduced lodging under stressing conditions. Our research could be easily extended to the case of simultaneous improvement for GY and SG by constructing an index with non-zero economic weights for both traits. This would give predictive results that are intermediate of those presented in our research, with variations depending on the relative economic weights assigned (not shown). Given that defining the appropriate relative economic value of GY and SG is beyond the scope of the present study, here we presented the evaluation of multi-trait prediction for a target trait at a time, without losing generality of results.

4.4.1. Multi-trait genomic prediction for grain yield

Genetic correlations between traits estimated in this research are consistent with previous results showing strong associations of GY with SG, PH and FT in sorghum hybrids (Jordan et al. 2003). In our case, additive parental effects of GY were only correlated with additive effect of SG and PH, whereas associations between GY and FT were basically explained by residual effects. In addition, the estimated heritability for GY in our study was lower than for the rest of the traits. Based on these findings, and according to multivariate BLUP and selection index theories, prediction of breeding values for GY should mainly benefit by utilizing additional information from SG and PH, while introducing an uncorrelated trait, such as FT, would reduce prediction efficiency due to incorporation of sampling error. These expectations were corroborated by our predictive results from cross-validation when records of auxiliary traits were available for the training lines (Figure 2). Multi-trait models combining GY-SG-PH information produced the highest improvements in predictive ability for GY, representing up to 18% increase relative to ST G-BLUP (Table 2). These results are somewhat more promising than those previously reported for GY prediction in wheat

(Rutkoski et al. 2016; Sun et al. 2017), maize (Lyra et al. 2017) and rye (Schulthess et al. 2016), which found null increases in predictive ability under equivalent prediction scenarios.

The differences in response from using MT analysis are highly dependent of the experimental data and the genetic target of prediction. In our research, the inclusion of PH information in multi-trait models appeared to be essential for improving predictive ability, unbiasedness and accuracy of GY predictions. This may be attributed to the fact that PH presented the strongest additive genetic correlation and the largest difference in heritability with GY. The key role of these two factors for benefiting from multi-trait prediction of low-heritability traits has been demonstrated by simulation studies in plant and animal breeding contexts (Jia and Jannink 2012; Guo et al. 2014). In addition, the present study assumes that the genetic target of selection is across-environment breeding value, which depends on overall heritabilities and trait correlations across the TPE. It should be considered, however, that the levels of heritability and correlations are variable among experiments due to the influence genotype-by-environment interaction, which may induce variability of predictive results in specific trials.

4.4.2. Multi-trait genomic prediction for stay-green

Our validation results were different for SG prediction when only MT information on the training set was used. As expected from considering heritabilities and additive genetic trait correlations, predictive performance of the ST model was generally better than those of MT models. This may be explained because SG was genetically correlated only with GY, but this trait had the lowest heritability (Table 1). Under these circumstances, information from PH and FT could not be borrowed to predict SG, while GY data would be genetically less informative than the target trait itself. On the other hand, predictive results for SG were favoured by multi-trait genomic models when GY information was available for both the training lines and the predicted lines *per se*. This was reflected by increases in predictive ability and reductions in bias and MSEP of SG predictions (Figure 3).

The comparison of both prediction scenarios for SG allowed an empirical assessment of the value of two different sources of information from correlated traits. When auxiliary traits are only available for the training set, the estimates of covariance parameters among traits in $\mathbf{\Sigma}_g$, which are then used for prediction, contain information from the training lines exclusively. Alternatively, if records of auxiliary traits are also available for the validation set, better parameter estimates in $\mathbf{\Sigma}_g$ can be obtained for prediction of the target trait. This is not only because more data is used for parameter estimation, but mainly because the extra information contained in estimated trait correlations is directly sourced from the predicted lines. Our results for SG suggest that, when this extra information is exploited, the strength of correlations among traits is more important than the relative levels of trait heritability. This

would explain why SG predictions were improved by multi-trait G-BLUP using GY data, even when this trait had lower heritability than the target trait. Moreover, predictive performance of the MT model combining SG, GY and PH indicates that, despite SG and PH were uncorrelated, additional information from PH could be transmitted via GY to further enhance predictions of SG (Figure 3).

Similar benefits of exploiting information from secondary traits in the validation set have been reported in crops for other traits, but not for SG (Ruskoski et al. 2016; Lado et al. 2018; Fernandes et al. 2018). Even though the link between stay-green trait and yield stability has been demonstrated in sorghum, maize and wheat (Jordan et al. 2012; Christopher et al. 2008; Gregersen et al. 2013), the predictive use of SG-GY association in cereals had not been studied yet in the context of genomic selection.

4.4.3. Combining pedigree and genomic information for multi-trait prediction

This study explored if additional genealogical information from pedigree could further improve multi-trait genomic prediction. Merging pedigree and marker-based relationships has been shown to be beneficial for single-trait prediction in animals (e.g., Rodríguez-Ramilo et al. 2014; Ilska et al. 2017) and more recently in plant breeding (Velazco et al. 2019b). In the context of multi-trait analysis, models relying exclusively on SNP information may give distorted estimates of genetic variances and correlations between traits, mainly due to incomplete linkage disequilibrium (LD) of markers with causal loci (de los Campos et al. 2015; Gianola et al. 2015). Pedigree information can be used to account for residual polygenic effects not traced by SNPs, capturing LD patterns between loci that are due to common ancestral identity. Our study showed that better predictions of GY and SG were obtained by multi-trait models using a weighted combination of **A** and **G** instead of **G** alone (Table 2). Multi-trait K-BLUP produced the highest increase in predictive ability (of about 30%) for the SG prediction scenario, where trait correlations seemed to play a more important role in model performance. This result suggests that the **K** matrix optimized estimates of genetic correlation between traits, from a predictive perspective. The same approach has been applied to infer trait correlations from multi-trait prediction models in chicken (Momen et al. 2017).

Here, we used a different optimal weight w to construct **K** according to the trait of interest for prediction ($w = 0.25$ and $w = 0.30$ to predict GY and SG, respectively). This is in line with the idea that genetic (genomic) similarities between relatives are actually trait-specific (Fernando and Gianola 2018), which contrasts with the assumption of a common relationship matrix used in standard multi-trait G-BLUP. A more elaborate multi-trait G-BLUP model using trait-dependent weights has been recently proposed by Karaman et al. (2018) as a computationally less demanding alternative to multi-trait Bayesian methods. In this BLUP model, weights used to compute **G** are derived from posterior trait-specific (co)variance

estimates of SNP effects, which are obtained from a previous Bayesian analysis. The authors found, based on simulations, that the benefits of their weighting method were generally not significant for the low-heritability trait that more closely fitted an infinitesimal model (using 500 QTLs). Alternatively, our multi-trait K-BLUP uses a more straightforward approach that assumes a common weight across the genome, while still relaxing the assumptions of conventional multi-trait G-BLUP by allowing the similarity matrix to vary across traits in order to optimize prediction of the target. Given the complexity of traits and the highly structured breeding population used for this study, genomic predictions are likely to rely more on familial relationships and less on information from specific SNPs in LD with QTL (Habier et al. 2007). In this context, the BLUP-based multi-trait models applied here are expected to perform well compared to more refined alternatives (Jia and Jannink 2012; Haile et al. 2018).

4.4.4. Implications of multi-trait genomic prediction for grain sorghum breeding

While genotyping costs are being constantly reduced, the efficient use of phenotypic data becomes more relevant for plant breeding programs, since field phenotyping is still costly and labor-intensive. This work has shown, for the first time, how phenotypes of several traits routinely measured in grain sorghum breeding can be efficiently utilized through multi-trait analysis to assist genome-based selection for a target trait. We demonstrated empirically that genomic prediction of parental breeding values for GY benefits mainly from using PH as auxiliary trait. This trait is particularly promising in practice, since PH phenotype can be potentially measured in all trials, as opposed to SG, which can only be phenotyped under specific environmental conditions. In addition, high throughput field phenotyping technologies are expected to increase accuracy of PH measurement in sorghum (Wang et al. 2018), which could favor GY selection indirectly through multi-trait prediction.

The potential of exploiting genetic association between GY and PH predictively brings a new perspective regarding selection strategies in advanced sorghum testing. Our results imply that, despite strong selection for appropriate height is generally imposed in early stages of breeding, multi-trait evaluation could be implemented for increasing the capacity to use remaining genetic variation in PH when selecting for GY. That is, selection based on multi-trait models is able to directly use the variation in GY that is not associated with PH, and can simultaneously exploit GY variability indirectly through the available variation in PH (within the acceptable range).

Based on findings from this study, when SG is the target of prediction, most advantage from multi-trait genomic analysis can be obtained only if predicted lines have been phenotyped for GY. This may be beneficial when the interest is to predict the genetic aptitude for drought-adaptation in lines that have been field-tested for GY performance but could not yet experience post-flowering water stress. Such scenario is compatible with selection for

broad adaptation across the TPE, which is generally the best approach to deal with the largely unpredictable occurrence of drought in the Australian sorghum region (Chapman et al. 2000).

Expression of functional SG can also be a consequence of reduced sink demand relative to source, due to low grain production at plant level (Henzell and Gillieron 1973; Duncan et al. 1981). For this reason, Borrell et al. (2014) pointed out that simultaneous selection for SG and GY should be applied in sorghum breeding programs to correct for functional SG that is actually driven by low sink demand. This correction can be automatically performed by joint analysis of both traits using multi-trait genomic models. For instance, in our case, selection of parental lines for broad adaptation would be based on a predicted genetic score that is optimally derived by combining three sources of information: direct information from own GY breeding value in well-watered environments, and indirect information from SG and GY breeding value of relatives in water-limited environments.

4.5. Conclusions

This study demonstrates, based on extensive breeding data, that there is potential to improve genome-based predictions of grain yield and stay-green traits in grain sorghum by using multi-trait genomic analysis. Results suggest that better predictive abilities and accuracies for GY prediction are obtained when PH information on the training lines is included in multi-trait genomic models. When SG is the target, the quality of predictions is likely to improve only if GY performance data is available for both the training and the predicted lines *per se* since, in this case, direct information from SG-GY genetic correlations is exploited predictively through multi-trait analysis.

This article also shows, for the first time in plant breeding, how a similarity matrix using trait-specific combinations of pedigree and marker-based relatedness can further enhance multi-trait genomic prediction. Collectively, our results imply useful properties of multi-trait BLUP to evaluate alternative prediction schemes for genetic improvement of crop productivity and drought adaptability in grain sorghum. Given that the traits considered in this study are commonly measured in cereal breeding programs, findings presented here can be also relevant for practical implementation in other major crops.

Author Contributions

JV, DJ, MM and FvE designed the research. JV performed statistical analyses and wrote the article. DJ and FvE edited the article. DJ coordinated the experiments and data collection. CH and EM processed and prepared the dataset. All authors read and approved the final version of the manuscript.

Funding

The data used in this research was generated with funding from the Grains Research and Development Corporation (GRDC) of Australia. JV received financial support from the National Institute of Agricultural Technology (INTA) of Argentina, Res. DN 1126/13.

Chapter 5

Genomic prediction for broad and specific adaptation in sorghum accommodating information on individual marker effects

Velazco J.G., Jordan D.R., Mace E.S., Hunt C.H., van Eeuwijk F.A.
Submitted to *Crop Sci*.

Abstract:
This paper reports the first study exploring genomic prediction for adaptation of sorghum to drought-stress (D-ET) and non-stress (W-ET) environment types. The objective was to evaluate the impact of both modelling genotype-by-environment interaction (G×E) and accounting for heterogeneous variances of marker effects on genomic prediction of parental breeding values for grain yield within and across environment types (ET). The impact of modelling G×E was investigated by considering different structures for the genetic covariance matrix between environments in BLUP-based prediction models. To study the benefits of accounting for heterogeneous variances of single nucleotide polymorphism (SNP) effects, different weights for individual SNPs were used when constructing the genomic relationship matrix. The BLUP models used a kinship matrix combining pedigree and genomic information, termed K-BLUP. The dataset comprised testcross yield performances of sorghum parental lines under D-ET and W-ET conditions, pedigree data and genotypes of lines on 4.8K SNPs. In general, modelling G×E increased predictive ability and reduced empirical bias of genomic predictions for broad adaptation across both ETs compared to models that ignored G×E by fitting a main genetic effect only. Genomic predictions for specific adaptation to D-ET or to W-ET were also improved by K-BLUP models that used data from both ETs and explicitly accommodated G×E, relative to prediction models that used data from the targeted ET exclusively or models that used all the data but assumed no G×E. Allowing for SNP-specific variances through weighted K-BLUP produced marked increments (between 43 % and 72%) in predictive ability of genomic prediction for grain yield in all adaptation scenarios. We conclude that G×E as well as locus-specific variances should be accommodated in genomic prediction models to improve adaptability of sorghum to variable environmental conditions.

5.1. Introduction

Enhanced productivity and adaptability of crops to variable environmental conditions are increasing demands for sustainable food supply in light of projected global population growth and climate change (Wiltshire et al. 2013; Lesk et al. 2016). More frequent drought events and erratic rainfall are expected to become major factors affecting crop yields in dryland farming systems (Daryanto et al. 2017). For that reason, genetic improvement of crops for adaptation to climatic fluctuations may play a central role in achieving stable food production in the future.

Plant breeding programs of major field crops rely on extensive multi-environment trials (MET) conducted under dryland conditions with the aim of adequately sampling environmental variation in a *target population of environments* (TPE) where new varieties will be grown. Two main strategies have been traditionally considered by crop breeders when dealing with selection under variable environments: to select genotypes based on broad adaptation across the entire TPE, or to select for specific adaptation of genotypes to a targeted subset of environments within a subdivided TPE. The relative benefit of both strategies is influenced by the magnitude and repeatability of genotype by environment interactions (G×E), which will define the effectiveness of selecting for differential performance of genotypes in the TPE. The definition of the breeding strategy for adaptation as well as the implicit possibility of subdividing the TPE are key analytical challenges that have been thoroughly discussed in the conventional plant breeding context (e.g., Cooper and Hammer 1996; Piepho and Möhring 2005; Atlin et al. 2011). The same challenges are renewed from the genomic selection perspective (Heslot et al. 2015; Malosetti et al. 2016). In this case, important questions are how the stratification of the TPE, the use of it to design the training sets and the modelling of G×E affect the quality of genomic predictions for broad and specific adaptation. These topics are addressed in the present article through investigating prediction for adaptation of sorghum to drought and well-watered environments in Australia.

Sorghum is a major global crop particularly well adapted to risky production environments where marginal and unstable rainfall patterns are likely to occur. In the Australian cropping region, seasonal variation in water supply is the main cause of yield variability, typically associated with inconsistent patterns of G×E (Chapman et al 2000; Hammer et al. 2014). This fact has traditionally favored selection strategies focused on improving for broadly adapted genotypes (Cooper and Hammer 1996; Borrell et. al. 2006). Attempts to increase predictability of G×E patterns can be based on the explicit use of environmental data as covariates to model genotypic responses (e.g., van Eeuwijk et al. 1996; Heslot et al. 2014). An alternative strategy is to rely on environmental information to define specific environment types (ET) and the use these to classify trials sampling a TPE. Chapman et al. (2000) and Hammer et al. (2014) applied this approach in sorghum to characterize the intensity and

timing of water limitation within the TPE by using crop simulation models. The authors suggested that selection strategies to improve adaptation should be based on ETs rather than geographic locations.

Post-flowering water deficit is the most common and critical type of drought affecting sorghum yield in Australia and worldwide (Jordan et al. 2012; Hammer et al. 2014). This limiting condition triggers the phenotypic expression of a drought-adaptation mechanism known as stay-green, which emerges as the ability of some genotypes to retain active green leaves under drought stress during grain filling (Borrell et al. 2014a). Accordingly, the manifestation of stay-green phenotype in sorghum can be seen as an integral indicator of terminal water deficit conditions occurring in a specific environment. In our study, we used the presence or absence of stay-green expression to classify trials as two distinct types of environment within the TPE: post-flowering drought environment (D-ET) or water-sufficient environment (W-ET), respectively. This stratification of the TPE was then applied to evaluate genomic predictions according to three hypothetical breeding targets: specific adaptation to terminal water-stress environments, specific adaptation to non-stress environments and broad adaptation across both types of environments.

An efficient analysis of MET data is essential for informed assessment of G×E and for accurate estimation of broad and specific adaptation in plant breeding. Many modeling approaches have been proposed during the last decades for the analysis of MET in crop breeding (as reviewed by e.g. DeLacy et al. 1996; Smith et al. 2005). Statistical methods accounting for G×E have been updated to incorporate the increasing availability of genetic marker information. Mixed models accounting for marker by environment interaction were initially used in crops for identification of QTLs with across- and within-environment effects (Boer et al. 2007; van Eeuwijk et al. 2010). In the framework of genomic selection, a best linear unbiased prediction method using marker-based relationships, termed G-BLUP (VanRaden 2008), has been applied to model G×E in wheat breeding (Burgueño et al. 2012; Lopez-Cruz et al. 2015; Oakey et al. 2016). Methods for multi-environment genomic prediction have been also developed within semi-parametric and Bayesian approaches (Cuevas et al. 2016; 2017).

G-BLUP method is commonly used in plant breeding because it is simple to implement and produce competitive results for prediction of complex quantitative traits (Heslot et al. 2012; Gianola et al. 2014). However, its predictive performance is expected to decline when genetic control of the target trait departs from an infinitesimal model, since G-BLUP assumes equal variance for all marker effects. Bayesian methods can relax this assumption by allowing unequal variances across the genome, but at the expenses of important increases in computational burden and predictions that are sensitive to prior specification (Lehermeier et al. 2012). An alternative to combine benefits of both methodologies is to apply a BLUP-based prediction model that is able to account for heterogeneous variances of SNP effects by

incorporating weights in the genomic relationship matrix (Zhang Z. et al. 2010; Wang et al. 2012; Su et al. 2014; Zhang X. et al. 2016). Although this approach has been implemented for genomic prediction in animals, studies evaluating its potential in plant breeding are still lacking. Here, we use the SNP weighting approach to account for locus-specific variances of marker effects for grain yield across and within ETs.

The objective of this research was to evaluate the effectiveness of modelling genotype by ET interaction and of accounting for heterogeneous variances of SNP effects to improve genomic prediction of parental breeding values for broad and specific adaptation in sorghum.

5.2. Materials and Methods

5.2.1. Experimental data

The phenotypic data used in this research comprised testcross evaluations of 603 female parental lines. These were evaluated in hybrid combination with male testers across 26 field trials covering 12 locations during a period of seven years. The female lines are developed by the sorghum germplasm enhancement program of the University of Queensland and the Queensland's Department of Agriculture and Fisheries in Australia, which licenses these lines to commercial breeding programs to be used as breeding material or as hybrid parents. The MET dataset was considered to be representative of environmental conditions experienced by the main sorghum production region in Australia and defined here as the TPE.

In all trials where phenotypic variation for the stay-green trait was observed, it was visually rated on a 1-9 scale at maturity. The scores assessed the percentage of the canopy death with a score of 1 representing plots with 90-100% of the plant canopy being alive and a score of 9 representing plots with 90-100% of the plant being dead. Previous studies have shown leaf senescence to be highly correlated with the degree of post-flowering drought stress experienced by the plant. This information was used in our study as an environmental descriptor to subdivide the MET dataset according to two distinct ETs within the TPE: post-flowering drought ET (D-ET), for trials where stay-green expressed; and water-sufficient ET (W-ET), for trials where no manifestation of stay-green was observed. In the present series of trials, 9 were classified as D-ET and 15 as W-ET. The trials under D-ET conditions occurred in five of the seven years at seven different locations. The mean grain yield in D-ET trials was 24% lower than in W-ET trials. Similar levels of yield reduction caused by drought have been reported for rice and wheat in a meta-analysis study (Zhang et al. 2018). All the lines were tested under both ETs with different levels of replication due to data imbalance. Further details on the general structure of the dataset can be found in Velazco et al. (2019b).

All female lines were genotyped using an integrated DArT and genotyping-by-sequencing (GBS) methodology involving complexity reduction of the genomic DNA to remove

repetitive sequences using methylation sensitive restriction enzymes prior to sequencing on Next Generation sequencing platforms (DArT, www.diversityarrays.com). The sequence data generated were then aligned to the most recent version (v3.1.1) of the sorghum reference genome sequence (Paterson et al. 2009; McCormick et al. 2018) to identify SNPs. After running standard quality filtering, genotypes on 4782 SNP markers remained for the analyses. In addition to genomic data, genealogical information tracing back 28 generations was available for all the lines from deep pedigree data.

5.2.2. Phenotype analysis

Prior to the implementation of genomic prediction models, field plot data from the testcross progeny trials were analyzed using a two-stage approach to estimate adjusted line means in each ET (D-ET and W-ET). Spatial analysis of each field experiment in the first stage was performed as described in Velazco et al. (2019b). In the second stage, spatially-adjusted best linear unbiased estimates (BLUE) of testcross hybrid means from individual trials were combined for analysis using the following model:

$$y_{ijkl} = \mu_j + LE_{ij} + ME_{kj} + LME_{ikj} + T(E)_{l(j)} + LT(E)_{il(j)} + MT(E)_{kl(j)} + LMT(E)_{ikl(j)}, \quad (1)$$

where y_{ijkl} is the spatially-adjusted BLUE for grain yield of the i-th line crossed with the k-th tester in the l-th trial belonging to the j-th ET, μ_j is the general mean for the j-th ET, LE_{ij} is the effect of the i-th line in the j-th ET, ME_{kj} is the effects of the k-th male tester in the j-th ET, LME_{ikj} is the ik-th line-by-tester cross effect in the j-th ET, $T(E)_{l(j)}$ is the effect of the l-th trial nested within the j-th ET, $LT(E)_{il(j)}$ is the il-th line-by-trial interaction effect within the j-th ET, $MT(E)_{kl(j)}$ is the kl-th tester-by-trial interaction effect within the j-th ET, and $LMT(E)_{ikl(j)}$ is the ikl-th line-by-tester-by-trial interaction effect within the j-th ET. All the effects were considered fixed except for $T(E)$ and all its interactions, which were assumed random and independent normally distributed with zero mean. Heterogeneous variances across ETs were allowed for the random effects. Even though lines are conceptually random, they were taken a fixed at this stage to avoid double shrinkage of line effects in the genomic prediction stage.

5.2.3. Genomic prediction models

Genomic predictions for broad adaptation across ETs and for specific adaptation within ETs were performed with BLUP models differing in the genetic covariance matrix between ETs and in the weights used to build the environment-dependent relationship matrix among lines. The general model formulation is:

$$\mathbf{y} = \mathbf{1}\boldsymbol{\mu} + \mathbf{Z}_g\mathbf{g} + \mathbf{e}, \quad (2)$$

where $\mathbf{y} = (\mathbf{y}'_D, \mathbf{y}'_W)'$ is a vector of line BLUEs for yield under D-ET and W-ET from the combined analysis with model (1); $\mathbf{1}$ is a vector of ones with associated ET-specific general means $\boldsymbol{\mu} = (\mu_D, \mu_W)'$; $\mathbf{g} = (\mathbf{g}'_D, \mathbf{g}'_W)'$ is a multi-environment vector of total additive genetic effects or genomic estimated breeding values (GEBV), with corresponding incidence matrix $\mathbf{Z}_g$ relating $\mathbf{y}$ to $\mathbf{g}$; and $\mathbf{e}$ is a vector of random residuals distributed as $\mathbf{e} \sim N(0, \mathbf{R})$, where $\mathbf{R}$ is a diagonal matrix with elements computed as in Smith et al. (2001a). This matrix accounts for differences in reliability of estimated line means in each ET resulting from within-ET variance heterogeneity and unequal number of testers and trials used in line evaluation. The distribution of total additive genetic effects for individual ETs was assumed to follow a separable form as $\mathbf{g} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_E \otimes \mathbf{K}_E)$, where $\boldsymbol{\Sigma}_E$ is the genetic variance-covariance matrix between ETs, $\mathbf{K}_E$ is the environment-dependent kinship matrix among lines, and $\otimes$ is the Kronecker product operator. The forms of $\boldsymbol{\Sigma}_E$ and the derivations of $\mathbf{K}_E$ are described below.

5.2.3.1. Structures for matrix $\boldsymbol{\Sigma}_E$

The matrix $\boldsymbol{\Sigma}_E$, of order 2 × 2, has diagonal elements representing the genetic variances within each ET and the off-diagonal element representing the genetic covariance between both ETs. In order to assess the effect of modeling genotype-by-ET interaction on genomic predictions, four models assuming different structures for $\boldsymbol{\Sigma}_E$ were considered:

- Model G: genetic effects were modeled using $\boldsymbol{\Sigma}_E \sigma^2_g$, where $\boldsymbol{\Sigma}_E$ is a matrix with all ones and σ^2_g represents the common genetic variance for both ETs, which is also the covariance between ETs. This model is equivalent to fitting a main line effect across all trials, ignoring line-by-ET interactions, i.e., assuming that the TPE is not subdivided into ETs.
- Model GE: using a uniform or compound symmetry structure for $\boldsymbol{\Sigma}_E$, which assumes the same genetic variance in both ETs and a distinct term for the genetic covariance between ETs. This model is equivalent to the standard variance component model with a main line effect plus a line-by-ET interaction effect.
- Model GEH: a more general formulation of model GE using an unstructured form for $\boldsymbol{\Sigma}_E$, which also assumes genetic covariance between ETs but, in this case, allowing for heterogeneous genetic variances across ETs. This parameterization enables the most flexible modelling of genotype-by-ET interaction.
- Model ID: using a diagonal structure with heterogeneous variances for $\boldsymbol{\Sigma}_E$, which allows for a distinct genetic variance in each ET but not for genetic covariance between ETs. This model assumes independence of line effects between ETs and is analogous to conducting a separate analysis for each ET. The ID model was considered only for genomic prediction of specific adaptation to a targeted ET.

5.2.3.2. Computation of environment-dependent kinship matrices $\mathbf{K}_E$

In this study, we used estimated variances of individual SNP effects for grain yield within D-ET, within W-ET and cross both ETs to build three different environment-dependent kinship matrices: $\mathbf{K}_D$, $\mathbf{K}_W$ and $\mathbf{K}_B$, respectively. The kinship matrix $\mathbf{K}_E$, with $E = D, W, B$, combines pedigree and genomic information in the following form:

$$\mathbf{K}_E = w\mathbf{A} + (1 - w)\mathbf{G}_E, \tag{3}$$

where $\mathbf{A}$ is the numerator relationship matrix among lines based on the full pedigree, $\mathbf{G}_E$ is an environment-dependent genomic relationship matrix computed from the SNP marker data, and the parameter w represents the proportion of total additive genetic variance that is not captured by SNPs and is accounted for by genealogical information contained in $\mathbf{A}$. Under the BLUP method using this merged pedigree-genomic matrix, denoted K-BLUP, the value of w is empirically determined based on cross-validation in order to optimize predictive ability (see Velazco et al. 2019b for details). Based on results from the latter study for grain yield prediction, we used $w = 0.6$ for the present research.

The genomic matrix $\mathbf{G}_E$ used to construct $\mathbf{K}_E$ was computed according to VanRaden (2008):

$$\mathbf{G}_E = \mathbf{Z}\mathbf{D}_E\mathbf{Z}'\lambda, \tag{4}$$

where $\lambda = 1/(2\sum p_i(1 - p_i))$ is a scaling factor based on the minor allele frequency p_i of SNP i summed over all loci, $\mathbf{Z}$ is a matrix of centered genotypes for each line, and $\mathbf{D}_E$ is a diagonal matrix of weights for the variances explained by each SNP in a particular environmental context. The environment-dependent SNP weights included in $\mathbf{D}_E$ were derived from (purely) genomic breeding values (GBVs) of lines within D-ET ($\tilde{\mathbf{m}}_D$), within W-ET ($\tilde{\mathbf{m}}_W$), or across both ETs ($\tilde{\mathbf{m}}_B$), where $\tilde{\mathbf{m}}_B = 0.5\tilde{\mathbf{m}}_D + 0.5\tilde{\mathbf{m}}_W$. For this derivation, we applied the procedure proposed by Wang et al. (2012), which can be generally describe, for $E = D, W, B,$ as follows:

a) Environment-dependent GBVs ($\tilde{\mathbf{m}}_E$) were predicted with conventional G-BLUP by fitting an unweighted matrix, termed $\mathbf{G}$, which was computed using an identity matrix $\mathbf{I}$ instead of $\mathbf{D}_E$ in (4);

b) Predicted SNP effects ($\tilde{\mathbf{u}}_E$) were derived from $\tilde{\mathbf{m}}_E$ as: $\tilde{\mathbf{u}}_E = \lambda\mathbf{I}\mathbf{Z}'\mathbf{G}^{-1}\tilde{\mathbf{m}}_E$;

c) Individual variances of SNP effects ($\tilde{\sigma}^2_{u_{i,E}}$) were then estimated as: $\tilde{\sigma}^2_{u_{i,E}} = \tilde{u}^2_{i_E} 2p_i(1 - p_i)$;

d) The SNP variances were standardized and used to build $\mathbf{D}_E$.

Accordingly, the obtained weighting matrices $\mathbf{D}_D$, $\mathbf{D}_W$ and $\mathbf{D}_B$ were used as in (4) for computation of corresponding genomic matrices $\mathbf{G}_D$, $\mathbf{G}_W$ and $\mathbf{G}_B$, and these were finally included in (3) to build $\mathbf{K}_D$, $\mathbf{K}_W$ and $\mathbf{K}_B$, respectively. According to Wang et al. (2012), weights can be re-computed in order to optimize predictions by iterating either on SNP effects alone (i.e., by looping to step b) or on GBVs and SNP effects (i.e., by looping to step a). We obtained optimal model performances using SNP weights from the first iteration and these are the predictive results reported in our study. For illustration purpose, individual SNP variances or weights used in $\mathbf{G}_B$ are represented graphically in the Appendix.

To evaluate the impact of accounting for heterogeneous SNP variances, we also considered predictions based on the unweighted kinship matrix, denoted simply as $\mathbf{K}$, which was formed using the unweighted $\mathbf{G}$ matrix—by setting $\mathbf{D}_E = \mathbf{I}$. Note that the use of this genomic matrix implies that all SNPs are assumed to explain the same amount of genetic variance, which is equivalent to the conventional G-BLUP (using the first method in VanRaden 2008).

Narrow-sense heritability of grain yield in each ET was estimated for the different models as the ratio of additive-genetic variance to total variance. Note that these estimates are approximations of the true unknown heritabilities and are presented in this study only as a measure of the ability of the models to capture genetic variation. Additionally, the goodness of fit of the different models was assessed using the Akaike Information Criterion (AIC).

5.2.4. Evaluation of genomic prediction for broad and specific adaptation

Genomic prediction models were evaluated through cross-validation technique, where the realized genotypic values of lines ($\mathbf{y}$) were used as predictands to validate GEBVs ($\tilde{\mathbf{g}}$). Validation of predictions was framed in terms of the different breeding objectives for adaptation, where the target genetic values were either $\mathbf{y}_D$ or $\mathbf{y}_W$ when selecting for specific adaptation and $\mathbf{y}_B$ when selecting for broad adaptation, with $\mathbf{y}_B = 0.5\mathbf{y}_D + 0.5\mathbf{y}_W$ being the overall mean across both ETs. Accordingly, the selection targets $\mathbf{y}_D$, $\mathbf{y}_W$ and $\mathbf{y}_B$ were predicted by the respective GEBVs $\tilde{\mathbf{g}}_D$, $\tilde{\mathbf{g}}_W$ and $\tilde{\mathbf{g}}_B$. The GEBVs in $\tilde{\mathbf{g}}_B$ are given directly by model G, whereas they were obtained as $\tilde{\mathbf{g}}_B = 0.5\tilde{\mathbf{g}}_D + 0.5\tilde{\mathbf{g}}_W$ for the other models. Note that $\mathbf{y}_B$ and $\tilde{\mathbf{g}}_B$ were computed in the form of a selection index for broad adaptation by combing line performances in each ET (Piepho and Möhring 2005; Kelly et al. 2007). In this study, we used the same relative index weight of 0.5 for both ETs.

We implemented a five-fold cross-validation scheme where data of multi-environment yield performance from the 80% of randomly sampled lines formed the training set (TS), while data from the remaining 20% of lines were retained as validation set (VS). This partitioning of the data corresponds to a model evaluation for genomic prediction of new lines, which have not been field-tested yet. Prediction models were compared on the base of

predictive ability and unbiasedness of predictions in the VS. Predictive ability was calculated as the correlation between GEBVs and realized genotypic values. Empirical bias of genomic predictions was determined as the regression coefficient of realized genotypic values on GEBVs. For each prediction scenario, we present average results over 20 random replications of the cross-validation scheme. The Hotelling-Williams t-test for dependent correlations (Steiger 1980) was applied to determine statistical differences in predictive ability for pairwise comparisons of among prediction models.

5.3. Results

5.3.1. Parameter estimates and model fits

Table 1 presents estimates of heritabilities for each ET, genetic correlations between ETs and goodness of fit values from the K-BLUP models considered in this study. Independent of the

Table 1 Grain yield heritability estimates (h^2) for drought (D-ET) and water-sufficient (W-ET) environmental types, additive genetic correlations between D-ET and W-ET (r_g), and Akaike Information Criterion (AIC) values from K-BLUP models using different kinship matrices and genetic covariance structures between ETs.

Kinship matrix	Model	h^2		r_g	AIC[a]
		D-ET	W-ET		
K	G	0.47	0.47	1.00	244
	GE	0.50	0.50	0.94	240
	GEH	0.63	0.41	0.97	221
	ID	0.53	0.23	0.00	302
$\mathbf{K}_B$	G	0.43	0.43	1.00	57
	GE	0.48	0.48	0.90	39
	GEH	0.60	0.36	0.95	8
	ID	0.57	0.31	0.00	106
$\mathbf{K}_D$	G	0.46	0.46	1.00	59
	GE	0.51	0.51	0.90	36
	GEH	0.63	0.38	0.96	0
	ID	0.60	0.32	0.00	111
$\mathbf{K}_W$	G	0.46	0.46	1.00	56
	GE	0.51	0.51	0.91	37
	GEH	0.63	0.38	0.96	1
	ID	0.61	0.32	0.00	108

[a] Value expressed as difference relative the best-fitting model (with AIC = 0).

kinship matrix used, the model allowing estimation of a separate genetic variance for each ET and genetic correlation between ETs (GEH) gave better fits compared to models imposing equal genetic variances across ETs (G and GE) or assuming genetic independence between

ETs (ID). According to the best-fitting model GEH, grain yield heritability in D-ET was higher than in W-ET and genetic performances of lines in both ETs were highly correlated (r_g = 0.95–0.97) for all kinship matrices. The use of environment-dependent SNP weights caused important improvements in goodness of fit for all structures of $\mathbf{\Sigma}_E$, as reflected by marked decreases in the AIC values. However, small changes in parameter estimates from the different models were produced by weighting the kinship matrix, with more pronounced variation in estimated heritabilities only in the case of model ID.

5.3.2. Genomic prediction for broad adaptation

Predictive ability and empirical bias of genomic predictions for broad adaptation across ETs are shown in Table 2. In general, models GE and GEH improved predictive ability compared to model G, which ignores the stratification of the TPE for prediction of broadly adapted genotypes. These improvements were always significant ($p < 0.05$) and slightly magnified when weighted kinship matrices were used, representing a mean relative increase of 10% in predictive ability. More important and highly significant gains in predictive ability ($p < 0.001$) were generally achieved by weighting the kinship matrix for individual SNP variances, with an average increment of 62% across models of $\mathbf{\Sigma}_E$. K-BLUP models incorporating marker effect information from a specific ET—through $\mathbf{K}_D$ or $\mathbf{K}_W$—tended to produce marginally better results than models using across-ETs marker information—through $\mathbf{K}_B$—, although these differences were not significant.

Table 2 Predictive ability and empirical bias (regression coefficient) of genomic predictions for broad adaptation across drought and water-sufficient environmental types (ETs), obtained from K-BLUP models using different kinship matrices and genetic covariance structures between ETs. Mean values over 20 replicates of a five-fold cross-validation scheme.

	Kinship matrix				
Model	$\mathbf{K}_B$	$\mathbf{K}_B$	$\mathbf{K}_D$	$\mathbf{K}_W$	Mean
		Predictive ability			
G	0.342	0.537	0.547	0.549	0.494
GE	0.364	0.591	0.600	0.602	0.539
GEH	0.374	0.592	0.605	0.607	0.545
Mean	0.359	0.573	0.584	0.586	
		Empirical bias			
G	0.993	1.177	1.189	1.189	1.137
GE	0.735	0.931	0.928	0.931	0.881
GEH	0.816	1.052	1.043	1.045	0.989
Mean	0.848	1.054	1.053	1.055	

Almost unbiased genomic predictions for broad adaptation were obtained by model G using the unweighted **K**, while models GE and GEH with the same kinship matrix increased bias and led to overestimation or inflation of GEBVs, as indicated by regression coefficients lower than 1. Accounting for heterogeneous SNP variances with matrices $\mathbf{K}_B$, $\mathbf{K}_D$ or $\mathbf{K}_W$ reduced bias of GEBVs for both GE and GEH models and removed inflation of predictions for the latter. Finally, differences in regression coefficients for a specific model of $\mathbf{\Sigma}_E$ were very small among weighted kinship matrices.

5.3.2. Genomic prediction for specific adaptation

Table 3 shows the results on model performance for specific genomic predictions within D-ET. Overall GEBVs across ETs obtained from model G always gave the lowest predictive abilities for specific adaptation to D-ET, independently of the applied kinship matrix. Models GE and GEH, which explicitly account for genotype-by-ET interaction and produce ET-specific GEBVs, resulted in significant increments of predictive ability in D-ET ($p < 0.01$), representing an average improvement of 17% relative to model G. Even though model ID made use only of D-ET data for training and prediction, its predictive abilities were significantly higher than those of model G ($p < 0.01$) and lower but statistically comparable to models GE and GEH when using the same kinship matrix. A general increment of 72% was obtained in predictive ability for D-ET by applying SNP-specific weights in K-BLUP, compared to the mean performance of models based on the original matrix **K**. The use of different sets of weights did not cause significant differences in predictive ability between models, although weighting specifically for SNP effects in D-ET gave generally better results.

Table 3 Predictive ability and empirical bias (regression coefficient) of genomic predictions for specific adaptation to D-ET, obtained from different K-BLUP models using different kinship matrices and genetic covariance structures between ETs. Mean values over 20 replicates of a five-fold cross-validation scheme.

	Kinship matrix				
Model	$\mathbf{K}_B$	$\mathbf{K}_B$	$\mathbf{K}_D$	$\mathbf{K}_W$	Mean
	Predictive ability				
G	0.265	0.468	0.485	0.479	0.424
GE	0.325	0.548	0.563	0.559	0.498
GEH	0.324	0.544	0.562	0.558	0.497
ID	0.317	0.538	0.555	0.551	0.490
Mean	0.308	0.524	0.541	0.537	
	Empirical bias				
G	1.040	1.386	1.425	1.403	1.314
GE	0.828	1.039	1.046	1.047	0.990
GEH	0.783	1.043	1.035	1.032	0.973
ID	0.800	1.081	1.071	1.065	1.004
Mean	0.863	1.137	1.144	1.137	

Model G gave less biased GEBVs relative other structures of $\boldsymbol{\Sigma}_E$ when unweighted K-BLUP was used to predict yield in D-ET (Table 3). However, GEBVs from this model showed the largest biases under weighted K-BLUP and resulted in marked under-prediction of line performances. In contrast, the use of weighted kinship matrices in models GE, GEH and ID reduced bias and eliminated inflation of genomic predictions for D-ET.

Table 4 Predictive ability and empirical bias (regression coefficient) of genomic predictions for specific adaptation to W-ET, obtained from different K-BLUP models using different kinship matrices and genetic covariance structures between ETs. Mean values over 20 replicates of a five-fold cross-validation scheme.

	Kinship matrix				
Model	$\mathbf{K}_B$	$\mathbf{K}_B$	$\mathbf{K}_D$	$\mathbf{K}_W$	Mean
		Predictive ability			
G	0.309	0.422	0.419	0.430	0.395
GE	0.295	0.434	0.430	0.438	0.399
GEH	0.296	0.438	0.441	0.450	0.406
ID	0.284	0.400	0.390	0.403	0.369
Mean	0.296	0.424	0.420	0.430	
		Empirical bias			
G	0.930	0.960	0.945	0.966	0.950
GE	0.646	0.774	0.752	0.760	0.733
GEH	0.874	1.068	1.055	1.066	1.016
ID	1.148	1.164	1.141	1.151	1.151
Mean	0.900	0.991	0.973	0.986	

When specific adaptation to W-ET was the target of prediction, the performances of different models for $\boldsymbol{\Sigma}_E$ were statistically similar in terms of predictive ability if the unweighted $\mathbf{K}$ was fitted (Table 4). Differences were significant only between models ID and GEH ($p < 0.05$) under weighted K-BLUP. In contrast to results for D-ET, predictive abilities of K-BLUP using model G were not significantly different from those of the models producing ET-specific GEBVs, while model ID gave the lowest predictive abilities for W-ET with all kinship matrices. In line with the findings on prediction for broad adaptation and for specific adaptation to D-ET, differences in predictive ability between weighted and unweighted K-BLUP models were highly significant for yield prediction in W-ET. The average improvement rate was 43% in this case; lower than for the other prediction scenarios. Predictive abilities were generally higher when the marker information used to scale the kinship matrix was derived from the same W-ET. For this targeted ET, predictive ability of genomic predictions was optimal with K-BLUP fitting $\mathbf{K}_W$ and model GEH.

As shown in Table 4, regression coefficients from model G were always the closest to 1, although this model caused inflation of GEBVs for W-ET in all cases. Accounting for genotype-by-ET interaction with model GE resulted in even stronger inflation. However, this

problem was diminished by accommodating heterogeneous genetic variances for individual ETs with model GEH. Empirical bias was generally lower for most K-BLUP models accommodating locus-specific variances, with the combined use of matrix $\mathbf{K}_D$ and model GEH giving the most unbiased and not inflated genomic predictions in W-ET.

5.4. Discussion

This study investigated the ability of different genomic models to predict multi-environment GEBVs of sorghum parental lines for grain yield by evaluating testcross hybrid performance under drought-stress and non-stress environments. Models considered here for multi-environment genomic evaluation accommodated different genetic covariance structures to model G×E and accounted for homogeneous or heterogeneous variances of individual SNP effects while incorporating additional information from pedigree. These features were fully implemented within a unified BLUP-based framework.

Previous genomic selection studies in sorghum have focused on within-trial (Hunt et al. 2018) or across-trial predictions (Velazco et al. 2019a; 2019b). This has been the common approach in other crops as well (e.g., Burgueño et al. 2012; Albrecht et al. 2014; Oakey et al. 2016). For the present research, we adopted an alternative approach where the MET series sampling the TPE was stratified according to two types of environments—or mega-environments—, D-ET and W-ET, and each one of these was assumed to be represented by a random sample of trials. Accordingly, our study focused on multi-environment GEBVs that are conditioned on the ETs. In the terms discussed by McLean et al. (1991), these predictors are then applicable to an intermediate inference space, which is broader than the narrow inference space of specific predictions within trials. At the same time, our intermediate space GEBVs apply to a narrower inference space relative to global across-trial predictions. Nevertheless, the obtained ET-specific genomic predictions result more informative since they make direct reference to the targets of adaptation studied. Many studies have investigated the effects of modeling G×E in the context of GP, but none had yet reportedly compared broad vs specific adaptation strategies.

5.4.1. Impact of modeling genotype-by-ET interaction:

Genomic analysis indicated that the most realistic descriptions of the genetic covariance between ETs were obtained using the unstructured matrix of model GEH (Table 1). This covariance model revealed that line performances under D-ET were genetically more informative than under W-ET, as reflected by the differences in within-ET heritabilities. A possible explanation for differences in heritability is that, in terminal-stress environments, the additive-genetic variation of grain yield in the testcross hybrids is magnified as a consequence

of the genetic variability in stay-green expression among hybrids. This would not be the case under adequate water conditions, where genetic differences in stay-green are expected to have no impact on yield. In addition, the best-fitting model of $\boldsymbol{\Sigma}_E$ indicated that additive-genetic effects for yield in both ETs were highly and positively correlated. These results may indicate that most of the genetic variation for yield in wet environments is also contributing to yield under drought environments, but with additional genetic variability in the latter case resulting from varying drought adaptation capacity of genotypes. Previous studies by Jordan et al. (2012) and Borrell et al. (2014b) suggested that stay-green trait had limited or no penalty on yield under water-sufficient conditions, which would be in line with the consistency of yield performance across ETs found in our study. It should be noted, however, that the high genetic correlation between D-ET and W-ET may be partly a consequence of the environmental classification used. Given that for this research trials were grouped considering only the presence or absence of stay-green, D-ET may include trials where the stay-green phenotype was scored but the levels of drought stress were insufficient to cause changes in genotype rankings for yield, with respect to non-stress environments. The genetic correlation between ETs is expected to be lower if D-ET is defined only by trials where stay-green expression is a clear determinant of genetic differences in yield.

From a practical perspective, the presence of strong genetic correlation between ETs also implies the possibility of exploiting across-ET information for genomic prediction. When considering prediction for broad adaptation, our result showed that accommodating genotype-by-ET interaction through models GE or GEH improved predictive ability compared to modeling only a main genetic effect across ETs, as with model G (Table 2). These improvements for across-environment prediction of untested lines are similar in magnitude to those previously reported in barley using pure genomic models (Malosetti et al. 2016) and slightly better than those found in wheat using equivalent models that combined pedigree and marker information (Burgueño et al. 2012; Sukumaran et al. 2017). Besides predictive ability, we also considered the empirical bias of predictions since this measure may indicate the systematic tendency of a model to under- or over-predict GEBVs in a specific environment, with over-prediction or inflation of GEBVs being particularly detrimental for genetic gain in the long term (Aguilar et al. 2010; Velazco et al. 2019b). Our study showed that the best combinations of predictive ability and unbiasedness were generally achieved when genotype-by-ET interaction and heterogeneous genetic variances across ETs were accommodated in K-BLUP models. Collectively, these results suggest that even when the genetic correlation between ETs was high, accounting for a subdivided TPE in genomic analysis might be a better strategy than ignoring this subdivision when targeting prediction of broadly adapted sorghum lines. In agreement with our findings, Piepho and Möhring (2005) also demonstrated that considering a stratified TPE in BLUP models can maximize the expected response to selection for broad adaptation, even when the G×E component is not large.

For prediction of specific adaptation, the relative performance of K-BLUP models varied according to the targeted ET. The main contrasts in predictive performance were observed between model G and model ID, with the latter outperforming the former in D-ET and the opposite occurring in W-ET (Tables 2 and 3). The reason for this may be that model ID uses only data from the target ET as training set, while model G uses data from both ETs, assuming that all the data are equally informative for prediction in the targeted ET. Therefore, in the case of prediction for D-ET, GEBVs from model ID were exclusively based on training data from the genetically more informative ET. In contrast, GEBVs from model G were also based on the low-heritability data from W-ET, which seemed to deteriorate predictions for D-ET. For the same reason, predictions in W-ET benefited by borrowing genetic signal from the high-heritability D-ET data with model G, as opposed to predictions that were exclusively derived from the less informative W-ET data using model ID.

Besides these two alternative models, we found that models GE or GEH were generally better options to improve specific predictions for both ETs. The superiority in performance may be explained by how these models combine information from both ETs to assist prediction in the targeted ET by applying an optimal weighting of the training data. The weighting procedure, which is inherent in BLUP-based estimation, has been cleverly described by Piepho and Möhring (2005) for the context of prediction within a subdivided TPE. In our case, the weights depended on the amount of genetic information provided by each ET, which was determined by the heritabilities in individual ETs and the genetic correlation between ETs. Piepho and Möhring (2005) showed that applying this automatic weighting, by contemplating a subdivided TPE in BLUP analysis, increased the expected response to selection for specific adaptation when compared to models that use all the data but ignore the subdivision (as with our model G) or models that use only data from the targeted environment (as with our model ID). Their expectations were corroborated empirically in the present study when inspecting predictive ability and bias of multi-environment genomic predictions. In addition, our results demonstrate that the benefits of using multi-environment models over genetic main effect models also depend on the difference in genetic information content between the target and the non-target environments, and not only on the genetic correlation between environments, as previously suggested by Atlin et al. (2000) and Dawson et al. (2013).

In our study, prediction for broad adaptation was based on a selection index that gave the same relative weight of 0.5 to both ETs, implying that line's testcross performances under post-flowering drought and non-drought conditions are equally important for the breeder. However, these index weights could differ between ETs according to informed selection decisions. For instance, the weights could be defined considering information on long term frequency of drought events within a particular TPE (Atlin et al. 2011) or on economic risk assessment of drought-related yield loss (Zhang et al. 2019). Jordan et al. (2012) suggested

that selecting for yield under terminal drought may be broadly beneficial for improving sorghum yield in Australia. This implies that more relative weight should be given to D-ET when targeting broad adaptation. Even though the definition of user-supplied weights is beyond the scope of the present research, it should be noted that, as previously mentioned, a higher weight is automatically given to D-ET data in our case by the K-BLUP models accommodating genotype-by-ET interaction.

Finally, the delineation of ETs in our study was based on a direct indicator of water stress experienced by the crop in extensive unmanaged field trials. It is important to mention, however, that the complementation with indirect environmental indicators of drought patterns (Chenu et al. 2011; Bustos-Korts et al. 2019) as well as with managed drought screening (Cooper et al. 2014) could be helpful to better define ETs and optimize selection. Moreover, when environmental variables are available, this information can be exploited directly by genomic models to predict not only for new lines but also for new environments (Malosetti et al. 2016; van Eeuwijk et al. 2019).

5.4.2. Impact of accounting for heterogeneous SNP variances:

This is the first study in plant breeding that explores how accounting for the genetic variance captured by each SNP in BLUP models can affect the quality of genomic predictions. We derived weights for locus-specific variances from GBVs of lines using the procedure of Wang et al. (2012). These weights were then used to obtain GEBVs incorporating additional pedigree information through a combined pedigree-genomic matrix. The integration of pedigree and marker-based information with K-BLUP has been shown to increase predictive ability and reduce bias of single- and multi-trait genomic predictions relative to conventional G-BLUP when the additive genetic variance is not fully captured by SNPs (Velazco et al. 2019a; 2019b). The method we used here for estimation of SNP weights is based on a fully-fledged mixed model procedure, as opposed to alternative weighting approaches that require Bayesian analysis (Zhang et al. 2010; Su et al. 2014). The SNP weights can be optimized predictively through an iterative process, as proposed by Wang et al. (2012). In the present research, only one iteration was required for optimal predictive results, while more iterations were actually detrimental (not shown). A similar finding was reported by Zhang et al. (2016) in a simulation study where prediction accuracy decreased after the first iteration for the most "complex" trait determined by 500 minor-effect QTLs. The authors attributed this decline to the fact that SNPs with large effects are emphasized while small-effect SNPs are excessively shrunk with successive iterations and re-computations of weights. The same could be argued in our case for grain yield considering that this complex polygenic trait is controlled by many genes with small effects (Melchinger et al. 1998; Schön et al. 2004).

Our results showed that accommodating heterogeneous variances of SNP effects had a big impact on improving genomic predictions for yield across and within ETs in sorghum. The marked gains in predictive performance imply that assuming SNP-specific variances gave better descriptions of the additive genetic variation among lines than assuming a common distribution for all markers. Previous simulation studies using Bayesian and weighting methods have suggested that allowing for SNP variance heterogeneity is likely to be more beneficial when traits are influenced by a few large-effect QTLs (Zhang et al. 2010; Daetwyler et al. 2010; Zhang et al. 2014). This seems to not be the case in our study, where weighting for SNPs produced important increases in predictive ability, even when individual SNPs explained less than 0.6 % of the genetic variation in yield (see Appendix). Accordingly, the benefit is apparently a consequence of considering the varying contribution of loci to the genetic variance, irrespective of the presence of major genes. In addition, it should be considered that the levels of gains from using SNP weighting might be partly determined by the relatively small size of the training population and the low-density panel used (less than 5K SNPs). The improvements are expected to be diluted with increased number of reference lines genotyped and high-density marker information.

For this research, we derived three sets of environment-dependent SNP weights by considering yield in D-ET, in W-ET and across ETs as different traits. We found that weights derived from ET-specific marker effects did optimize predictive ability and unbiasedness in all adaptation scenarios. However, the small differences obtained from changing sets of SNP weights indicate that the strategy to obtain weights based on environmental types was not crucial in our research. This was somewhat expected considering that the additive-genetic correlation between D-ET and W-ET was very high. Environment-dependent SNP weights are likely to have more impact on genomic prediction for specific adaptation in situations where the expression of yield QTLs changes more markedly across environments.

So far, the approach used in plant breeding to incorporate marker-specific information into genomic prediction models has been based on including additional effects for the markers identified as significant in a previous genome-wide association study (GWAS). However, the existing studies applying this approach on real data have reported no improvements or even worse predictive results for grain yield relative to conventional G-BLUP in maize and rice (Bustos-Korts et al. 2016; Spindel et al. 2016). Using the same approach, Rice and Lipka (2019) found that prediction accuracies were frequently lower for complex traits simulated from maize and sorghum diversity panels. One possible explanation for these results is that, by including a select group of highly significant marker as separate effects in BLUP models, the selected SNPs with larger effects receive less or no regularization (if they are considered fixed) compared to the non-significant SNPs with smaller effects. The implicit emphasis given by these prediction models to the SNPs with larger effects is expected to be beneficial for prediction of traits affected by major QTLs, but not for traits that are mainly controlled by

many small-effect QTLs. This has been demonstrated by Bernardo (2014) using simulations and corroborated empirically by the previously mentioned studies. Furthermore, it should be considered that the effect of each SNP is estimated individually in GWAS. As a result, the sum of SNP effects may overestimate QTL effects if several linked markers are actually capturing the effect of the same QTL (Su et al. 2014). Finally, the definition of the selected group of significant markers—and consequently the definition of the prediction model—will change according to the significance threshold chosen for GWAS, which is usually arbitrary (Bush and Moore 2012). As shown in the studies by Bustos-Korts et al. (2016) and Sarinelli et al. (2019) using wheat data, GWAS can fail to identify significant SNPs for grain yield when stringent thresholds are used, precluding the use of locus-specific information under this strategy.

In contrast to the GWAS-based approach, the SNP weighting method used in the present article estimates all marker effects simultaneously, and it does not require pre-selection of SNPs based on significance tests from extra GWAS analysis. Instead, a continuous gradient of weights is applied to all SNPs across the genome reflecting the relative contributions of each locus to the genetic variation of the trait. Our results have demonstrated that the weighted BLUP models were effective to improve genomic prediction for yield in sorghum. These models could be extended to include information on large QTLs affecting other traits that have been associated with improved drought adaptation, such as nodal root angle and stay-green (Mace et al. 2012; Borrell et al. 2014b). Further research is required to explore the possibilities of combining these sources of information in a multi-trait modelling framework.

An alternative weighting approach was proposed by Zhang et al. (2014) in which only the "most important" markers are weighted according to external results from publicly available GWAS. A potential drawback of this method results from assuming that significant QTL regions identified by previous independent studies are universally applicable, ignoring specificities of target breeding population such as the extent of LD or the interactions of QTLs with the genetic background (Vadez et al. 2011; Sandhu et al. 2018). This may partly explain the marginal increases in predictive ability obtained by Zhang et al. (2014) in rice when applying external weights. Although the weights used in our research are certainly conditional on dataset used, they are expected to be more suitable for prediction in the specific population of genotypes and environments that is relevant for the breeding program. Moreover, the reliability of these weights can be constantly improved by including updated phenotypic information, with the additional possibility of utilizing phenotypes of ungenotyped lines connected through pedigree (Zhang et al. 2016).

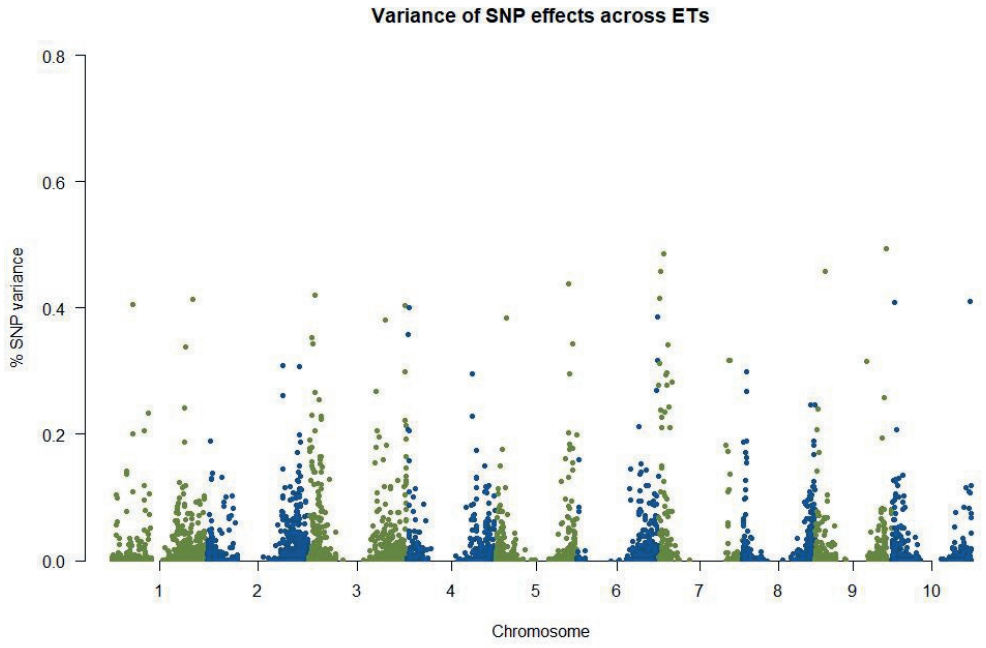
5.5. Conclusion

Results indicate that genomic prediction for broad adaptation across post-flowering stress (D-ET) and non-stress (W-ET) environments in sorghum can be enhanced by considering a subdivided TPE and addressing genotype-by-ET interaction explicitly with multi-environment models (GE or GEH). Specific adaptation to a targeted ET may be better predicted if genetic information from the non-targeted ET is borrowed through multi-environment genomic models, when compared to using only information from the targeted ET or to borrowing across-ET information through a main genetic effect model.

Our study also showed that accounting for heterogeneous variances of SNP effects with weighted K-BLUP models can be very effective in maximizing predictive ability and reducing bias of genomic predictions for grain yield. The levels of improvements may be only slightly dependent on the environmental context in which SNP weights were derived, mainly because of the strong correlation of line performances under post-flowering drought and non-drought conditions in sorghum.

Appendix

Genetic variance captured by each SNP, expressed as proportion of the total variance in grain yield.



Acknowledgments

We are grateful to Daniela Bustos-Korts for useful discussions and comments on the manuscript. The data used in this research was generated with funding from the Grains Research and Development Corporation (GRDC) of Australia. JV acknowledges financial support from the National Institute of Agricultural Technology (INTA) of Argentina, Res. DN 1126/13.

Chapter 6

General Discussion

6.1. Introduction

Further improvement in productivity and adaptability of crops requires the adoption of innovative methodologies to improve the efficiency of genetic evaluation in modern breeding programs. Genomic selection is a new method for improving quantitative traits that promises to overcome the limitations of traditional phenotypic selection and conventional MAS by exploiting genome-wide marker information to predict genetic merit. GS has been thoroughly studied in major cereal crops such as wheat and maize and its routine use in breeding programs is now a reality. This is not the case for grain sorghum where the implementation of GS has not been extensively explored yet.

The present thesis provides a first comprehensive insight into some challenges and opportunities for the implementation of GS in grain sorghum. The research questions were focused on investigating different statistical modelling strategies for improving predictive performance of genomic models. Specifically, this thesis investigated the benefits of: combining pedigree and genomic information (Chapter 3; Velazco et al. 2019b); exploiting information from multiple traits (Chapter 4; Velazco et al. 2019a); modelling G×E (Chapter 5); and accounting for heterogeneous variances of SNP effects (Chapter 5). Because genetic-marker effects are ultimately estimated from the phenotype, the appropriate analysis of phenotypic data is a key prerequisite for the implementation of GS. For this reason, Chapter 2 (Velazco et al. 2017) focused initially on the spatial adjustment of phenotypic variation in field trials.

The entire research of this thesis was carried out using real data sets from an ongoing sorghum breeding program. Consequently, an important motivation was to develop analytical strategies that considered the structure of typical sorghum breeding data while making use of the information that is routinely generated by the breeding program.

6.2. Simplified spatial analysis of plant breeding trials

The presence of spatial variation in field trials has long been recognized as a critical issue that can affect evaluation and selection of genotypes if not properly taken into account. This has been demonstrated in the contexts of conventional and genomic-based plant breeding (e.g., Qiao et al. 2004; Bernal-Vasquez et al. 2014). The importance of accounting for field trends is reflected by the large numbers of spatial analysis methodologies that have been developed, as reviewed in Chapter 2. The most commonly used methods, including the one proposed by Gilmour et al. (1997), are based on modelling field trends as spatial correlations among plots by fitting a specific covariance structure for the residuals in a mixed model. Within this approach, several modeling choices have to be made since different options for the spatial covariance structure exist, which can be fitted either in one or two dimensions, and the

inclusion of additional model terms may be also considered to account for the remaining trends not captured by spatial correlations. In order to define the spatial model for analysis, a multi-step model-building strategy is used, where several candidate models are tried in search for the best-fitting spatial model. As a result of this strategy, the standard spatial method demands intensive modelling and in-depth analysis of each trial. This may not be the most efficient practice when multiple trials have to be analyzed one at a time.

Chapter 2 of this thesis compared the performance of the standard spatial method of Gilmour el al. (1997) to that of a new simplified approach based on the SpATS mixed model (Rodríguez-Álvarez et al. 2018a). Analyzing a series of early-generation sorghum breeding trials, it was shown that the standard and the novel methods produce equivalent results regarding improvements in heritability and predictions of genetic effects. However, the main advantage of the spatial analysis with SpATS is that it allowed a simplified modeling strategy. This is a consequence of two distinctive features of SpATS model. First, large- and small-scale trends are simultaneously modeled by the smoothed spatial surface in a single modeling step. Conversely, under the standard method, both types of trends are fitted independently by including different model terms in a sequential multi-step modeling procedure. Second, the same SpATS model can be used to analyze different trials since the flexible spatial surface is automatically tuned by the smoothing parameters in order to optimize the balance of model fitting and parsimony in each case (Eilers et al. 2015). This contrasts with the conventional method, where the optimal spatial model has to be explicitly identified by the user in each trial after scrutinizing a number of model choices.

The gain in efficiency of analysis from using the new spatial method can be illustrated with our case study: for spatial analysis of 21 trials and two traits, more than 300 candidate models were tried in total under the standard method (not shown), while virtually the same results were obtained by fitting one SpATS model to analyze whole dataset. This reduction in required computational steps makes the spatial method based on SpATS a more suitable alternative for efficient routine analysis of large series trials in plant breeding programs. In addition, SpATS has been proved to be computationally more robust than models based on an autoregressive covariance structure (Rodríguez-Álvarez et al. 2018a). Finally, if a diagnostic assessment is required, SpATS provides explicit estimate of the spatial surface that can be graphically represented and easily interpreted, as shown in Chapter 2. This diagnostic may be also complemented with outlier detection analysis using the extracted spatially independent residuals of the model. Given the good performance of SpATS in Chapter 2, this method was adopted to correct for spatial trends in the subsequent studies of Chapters 3, 4 and 5.

More recently, Verbyla et al. (2018) showed that interaction surfaces based on tensor cubic smoothing splines can be fitted using a commercial mixed model software. However, they propose to use the smooth surface to model treatment and temporal effects, while residuals are still modeled using the standard approach with a separable variance–covariance structure.

In addition, Selle et al. (2019) demonstrated how the standard spatial model and a flexible model similar to SpATS can be fitted using full Bayesian analysis with the open-source R package INLA (www.r-inla.org). This implementation enables modelling several trials simultaneously. The authors cautioned, however, about the potential limitations of INLA method due to excessive computational demand when the number of trial-specific parameters increases.

6.3. Combining different sources of information for genomic prediction

Chapters 3, 4 and 5 investigated the effect of exploiting different sources of information when implementing genomic prediction for productivity and adaptability traits in sorghum. The Chapters considered data from early breeding stages aiming at the development of elite inbred lines, which will eventually serve as parents of commercial hybrids. At this stage, selection of parental lines is primarily based on their additive genetic values—or general combining abilities—estimated from testcross hybrid performance trials. Accordingly, the breeding values of parental lines were the genetic targets of prediction examined in these studies.

6.3.1. Information from pedigree and markers

In Genomic selection, complex quantitative traits are predicted by statistical models that utilize information from dense genome-wide markers. These markers are expected to be associated with the trait of interest through SNPs that are either located at causal loci or in linkage disequilibrium (LD) with causal genes, or through markers tracing familial relationships (Dekkers 2004; Habier et al. 2007). A derived implicit assumption when using purely genomic models is that markers explain all the additive-genetic variation of the target trait. However, because finite numbers of markers are used in real situations and markers are typically in imperfect LD with QTLs, genomic models inevitably fail to fully capture the genetic variation (de los Campos et al. 2013; de los Campos et al. 2015). This phenomenon was termed “missing heritability” when first realized in GWAS of human height using identified associated markers (Visscher 2008; Maher 2008). Yang et al. (2010) showed that even when all available SNPs were fitted simultaneously still a proportion of the genetic variance for this trait remained unexplained. This observation has been also reported for several complex traits in GS studies of animal breeding (e.g., Haile-Mariam et al. 2013). The consequence in practice is that this “missing genetic variance” cannot be used by genomic models, affecting the potential predictive performance that could be achieved by GS (Dekkers 2007). In Chapter 3 (Velazco et al. 2019b), information from traditional pedigree relationships was included in genomic prediction models to account for the residual polygenic variance not captured by SNP markers. The impact of using additional pedigree information

in genomic models was investigated by evaluating the quality of predictions for different traits in sorghum.

Results showed that markers alone did not explain the full additive-genetic variance and a combination of genomic and pedigree information was required to optimize predictive ability and reduce prediction bias for all traits and prediction scenarios studied. In this research, genomic predictions were optimized by BLUP models fitting a combined kinship matrix $\mathbf{K} = w\mathbf{A} + (1 - w)\mathbf{G}$, termed here K-BLUP. Previous studies have suggested theoretical calculations to derive a unique optimal value of w based on general assumptions (VanRaden et al. 2008; Goddard et al. 2011). In contrast, Chapter 3 proposed a more straightforward approach where specific weights are identified empirically for each trait and prediction scenario in order to maximize predictive ability using cross-validation. This predictive strategy using K-BLUP also differs from the common approach used so far in plant breeding, which is based on models that fit both **A** and **G** separately (e.g. Albrecht et al. 2011; Sukumaran et al. 2017). Under the latter approach (referred as AG-BLUP), the weight w is implicitly set to optimize model-fitting in the training data, but it is not purposely defined to maximize predictive performance in the validation set, as in K-BLUP method. Differences favoring K-BLUP over AG-BLUP were marginal. However, the main advantage of K-BLUP is that less variance parameters have to be estimated since a single relationship matrix is fitted. This may be attractive in practice since less computational time is required and more stable results are expected, mainly when working with small training samples.

Results from Chapter 3 demonstrated that integrating pedigree and genomic relationships can be an effective strategy to improve genomic prediction in sorghum compared to conventional G-BLUP. This is in line with previous studies in animal breeding (Liu et al. 2011; Rodríguez-Ramilo et al. 2014). Interestingly, the inclusion of pedigree information in our study was more beneficial for grain yield, which was the trait with the lowest heritability and is also the main target trait for sorghum breeding. It should be considered that only 40%-50% of the additive-genetic variance was captured by the ~4.8K SNPs available. In this case, the pedigree information seemed to have been particularly valuable to account for familial additive relationships, which is an important factor affecting prediction of largely polygenic traits (Habier et al. 2007; Habier et al. 2013). Obviously, the results are conditioned by the low marker density available in this study. The contribution of pedigree data to improve the performance of genomic models is likely to decrease if more markers are used. However, the method presented in this Chapter may be a suitable option for breeding programs of many crops where dense genotyping is still costly or even prohibitive. Although improvements in the quality of prediction can be moderate, including pedigree information to assist genomic models should always be considered since any benefit obtained would come without extra cost.

6.3.2. Information from correlated traits

In a following step exploring strategies for efficient implementation GS in sorghum, the integration of information from multiple traits was considered in Chapter 4 (Velazco et al. 2019a). The simultaneous analysis of different traits has long been recognized as a method to increase accuracy of genetic evaluation (Henderson and Quaas, 1976; Thompson and Meyer 1986). Multi-trait analysis is receiving renewed attention in the context of genomic prediction since more phenotypic records on the training population can be potentially used to predict breeding values. In addition, by exploiting relationships among relatives, individuals that are not phenotyped for all the traits of interest can still be included in multi-trait genomic analysis, further augmenting the volume of data that can be available for training prediction models. The amount of multi-trait information that can be actually used to assist prediction will depend on the correlations between traits and the heritabilities of the traits. It has been shown in simulation studies that benefits from multi-trait genomic analysis relative to single-trait analysis are expected for traits with low heritability when genetically correlated with traits of higher heritability (Calus and Veerkamp 2011; Jia and Jannink 2012; Guo et al. 2014). The potential advantage of multi-trait genomic prediction has been investigated empirically in cereal crops such as maize, rice and wheat with dissimilar results (e.g., dos Santos et al., 2016; Wang et al., 2017; Sun et al. 2017). In Chapter 4, a first study on this topic was presented for grain sorghum.

The Chapter focused on the possibility of using information from secondary or auxiliary traits to improve predictions for a primary target trait. A predictive assessment was adopted to identify the best multi-trait genomic model by trying all possible combinations of available traits in search for optimized prediction quality. Previous studies in crops have usually evaluated multi-trait models that included available auxiliary traits all together (e.g. Rutkoski et al. 2016; Sun et al. 2017; dos Santos et al. 2016). Moreover, Wang et al. (2017) suggested that larger gains in predictive ability might be obtained as more auxiliary traits are included in multi-trait genomic analysis. Results in Chapter 4 showed that including all auxiliary traits may not be the best option and actually combining specific subsets of traits can give better predictions. For instance, we found that using information from plant height alone was more effective for improving grain yield predictions than adding also information from stay-green and flowering time. The idea of using a selected subset of informative traits instead of all available traits has also been proposed for attaining optimal detection power in multi-trait QTL mapping (Cheng et al. 2013). Despite the definition of optimal multi-trait models in our study was purely based on a predictive approach, results confirm that the potential to exploiting information from correlated traits for genomic prediction can be inferred in principle by considering populational estimates of heritabilities and correlations among traits.

An important finding of this study was that the availability of plant height records for the training lines can be effectively exploited in sorghum to assist prediction of grain yield breeding values for new lines. This is a direct consequence of plant height being a trait with high heritability and presenting strong additive-genetic correlation with the lowly-heritable grain yield trait. Given that, in advanced line testing trials, plant height is routinely measured to control for agronomic type, any improvement in precision of plant height phenotyping could be beneficial for GS of yield in sorghum. In this respect, the implementation of new field phenotyping techniques may represent a promising tool to obtain more detailed measurements and integrate information over space and time that is potentially useful for prediction (van Eeuwijk et al.2019).

The improvements in genomic prediction for yield reported in Chapter 4 are a consequence of using information on auxiliary traits in the training population that is transmitted from relatives through genetic relationships. This can be referred to as "indirect" information from auxiliary traits. Such situation is compatible with a scenario where (new) predicted lines have not been field-tested for yield or any other trait (scenario Aux@TL; Figure 6.1). Previous studies on multi-trait genomic prediction of grain yield have considered the scenario where selection candidates are indeed observed for auxiliary traits, but not yield (Rutkoski et al. 2016; Sun et al. 2017). This situation could apply to very early stages of testing—where reliability of yield evaluation is actually quite doubtful—or when the target trait is too expensive to be measured directly at a large scale, such as baking quality traits in wheat (Lado et al. 2018). Our study did not consider such scenario for grain yield since, in advanced testing stages of cereal crops, direct measurement of grain yield is still the gold standard and indirect selection for this trait may be difficult to justify in practice.

When the target trait for multi-trait genomic prediction was stay-green, the study showed that large improvements in predictive ability, unbiasedness and accuracy could only be achieved when grain yield records were also available for the predicted lines (scenario Aux@TL+VL; Figure 6.1). In this case, the improvements are a consequence of exploiting "indirect" information plus "direct" information on auxiliary traits that is sourced directly from the selection candidates *per se*. The fact that yield was crucial to assist in the prediction of stay-green suggests that, when "direct" information on auxiliary traits is used, the improvement in prediction depends on the strength of genetic correlations between traits, regardless of differences in trait heritabilities. This speculation has been recently confirmed in a simulation study by Runcie and Cheng (2019), who demonstrated that, under Aux@TL+VL scenarios, prediction accuracy improved whenever genetic correlations among traits were large.

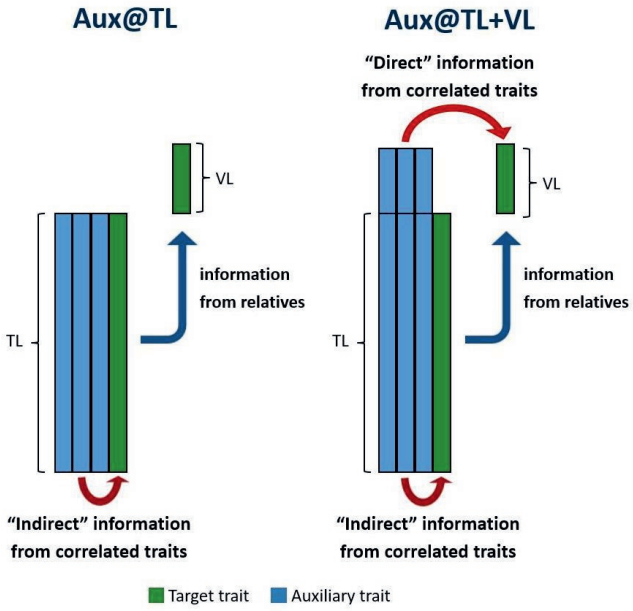


Figure 6.1 Schematic representation of multi-trait prediction scenarios Aux@TL, where data of auxiliary traits in the training lines set (TL) are included for prediction of the target trait in the validation lines set (VL), and Aux@TL+VL, where data of auxiliary traits in both TL and VL are included for prediction of the target trait in VL. The arrows represent the different sources of information used for prediction.

Chapter 4 also demonstrated how the K-BLUP method developed in Chapter 3 can be extended to improve the performance of multi-trait genomic prediction. Results showed that using a weighted combination of pedigree and SNP information was particularly beneficial for prediction of stay-green when using "direct" information from grain yield. We hypothesize that this is a consequence of obtaining better estimates of additive-genetic correlations between traits when both sources of information are used simultaneously, compared to using only markers. Gianola et al. (2015) showed that genetic correlations and purely genomic correlations are actually distinct parameters. The authors argue that part of the genetic correlation resultant from pleiotropy and LD among QTL is likely to be partially or totally "missed" due to imperfect LD between SNP and QTL. In addition, LD between SNPs can inflate the pleiotropy effects of QTL, distorting genetic correlations. In the case of our study, the inclusion of pedigree information in multi-trait genomic models seemed to have contributed to account for the portion of pleiotropy and LD between QTL that was not traced by SNP markers. Besides speculations on correlation parameter estimates, results for prediction of stay-green in sorghum showed that using a blended pedigree-genomic matrix was useful in practice to increase predictive ability of multi-trait genomic models.

6.3.3. Information from correlated environments

Based on classical theory of quantitative genetics, the same trait in different environments can be considered as different correlated traits (Falconer and Mackay 1996). Within this approach, genotype-by-environment interaction (G×E) may be addressed as a particular form of multi-trait analysis where genetic correlations between environments can be exploited to increase accuracy of genotypic evaluation. The analysis of G×E is central in plant breeding since it allows an appropriate modeling of differential responses of genotypes across a set of target environmental conditions (i.e., across the TPE). These differential responses provide the breeder with essential information to identify and select genotypes with superior performance in a subset of environments (specifically adapted) or across the entire TPE (broadly adapted). In Chapter 5, the impact of modelling G×E was investigated in the context of GS for broad and specific adaptation of sorghum lines. This research was conducted in the context of a subdivided TPE where the subdivision delineated two types of environments (ET): drought-stress (D-ET) and non-stress (W-ET) environments. The delineation of selection environments based on environmental types rather than geographic locations has been suggested to cope with highly unpredictable occurrence of drought, which is the main driver of yield variability for sorghum in Australia (Chapman et al. 2000; Borrell et al. 2006; Hammer et al. 2014).

Different models for multi-environment genomic prediction were explored in order to address relevant questions regarding GS for adaptability in sorghum. When aiming at prediction for broad adaptation, we investigated whether the subdivision of the TPE should be ignored, by using a main genetic effect model, or exploited, by using prediction models that explicitly accommodated line-by-ET interaction. The study showed that the latter option was generally a better strategy to predict for broadly adapted lines across the TPE. In particular, allowing for genetic correlation between ETs and unequal genetic variances for each ET optimized predictive ability and unbiasedness of genomic predictions. The benefits of accounting for G×E even when predicting for overall performance across environments have been also demonstrated by previous studies in conventional and genomic-based breeding (e.g., Kelly et al. 2007; Malosetti et al. 2016).

When investigating specific adaptation, an important question was if genomic models trained with data from the target ET exclusively can give better predictions than models that also use data from the non-targeted ET for training. We found that only for prediction in drought environments using data from the targeted ET alone could be a suitable option. This may be related to the fact that the heritability of grain yield under drought was higher than grain yield under non-stress conditions. Then including data from the W-ET may not be too informative to improve prediction in D-ET. However, the best predictive results for specific adaptation to both ETs were always obtained when line-by-ET interaction was accounted for

by fitting a compound symmetry or an unstructured matrix to model the genetic covariance between ETs. Based on the study by Piepho and Möhring 2005, it could be argued that, in our case, the superiority of these BLUP-based genomic models is likely to be a consequence of their ability to optimally combine information from the training data of both ETs to improve predictions for the target ET.

From a practical point of view, results of Chapter 5 generally suggest that the stratification of the TPE in drought stress and non-stress environments should be considered and line-by ET interaction should always be modelled to predict either for specific or broad adaptation of sorghum. It is worth noting that selecting lines for specific adaptation to drought or to non-stress conditions exclusively may not be a realistic option for sorghum breeding in Australia, where improving for broad adaptability has been traditionally the preferred strategy (Borrell et al. 2006; Jordan et al. 2012). However, the specific adaptation scenarios were considered in this Chapter to attain an empirical assessment of how informative yield performance under drought-stress environments is for prediction of yield under non-tress conditions, and vice versa. This information could be useful to assign different weights to each source of data when selecting sorghum lines for improved adaptation, for instance using a selection index.

6.3.4. Information from SNP-specific variances

In addition to studying the impact of modeling G×E, Chapter 5 also explored how accounting for heterogeneous variances of SNP effects in BLUP models may affect genomic prediction for grain yield in sorghum. Conventional G-BLUP method assumes that, given the same allele frequency, all marker effects across the genome have equal variance (Goddard 2009). This is consistent with the "infinitesimal model" of quantitative genetics, where a very large number of genes with small effects is assumed to contribute equally to the trait of interest. Several empirical studies have demonstrated that the implicit assumption of G-BLUP can work well for many complex traits of economic interest in plant breeding, which are usually lowly heritable and largely polygenic in nature, such as grain yield (e.g. Heslot et al. 2012; Wimmer et al. 2013). Nevertheless, it has been shown that quantitative traits are actually controlled by a finite number of genes (Hayes & Goddard 2001), and the true genetic architecture of any trait is expected to deviate from the assumptions of the infinitesimal model to some extent.

In Chapter 5, potential departures from the infinitesimal model were accommodated by allowing for unequal variances of SNP effects within BLUP-based method. For this, weights accounting for estimated SNP-specific variances were used in the computation of the genomic relationship matrices. Results showed that using SNP weighting in K-BLUP had a strong impact on increasing predictive ability and reducing empirical bias of genomic prediction for yield in all the adaptation scenarios studied. Predictive results as well as the marked improvements in goodness-of-fit of models using weighted **K** suggest that SNPs contributed

differently to the additive-genetic variation of grain yield among lines. The observed benefits are somewhat surprising for this trait since assuming heterogeneous SNP variances is expected to be particularly advantageous in presence of major QTLs affecting a trait (Habier et al. 2007; Wimmer et al. 2013). However, in the study of Chapter 5 each SNP explained less than 0.6% of the genetic variance, which is far below the threshold of 10% previously suggested as required for declaring a major QTL in plants (Bernardo 2014). Then, one possible explanation for the predictive benefits of weighting SNPs may be related to the fact that a low marker density was used in our study. With high marker density, the shrinkage that is imposed to SNP effects in conventional BLUP models is compensated because more SNPs are likely to support the same QTL (Habier et al. 2013). When working with low marker density, as in our case, the weights seem to play an important role in compensating for shrinkage by explicitly accounting for the differential contribution of SNPs to the genetic variation of the trait. Previous studies in animal breeding have reported that a weighted **G** matrix can produce better genomic predictions than the original unweighted **G** using real and simulated data (e.g. Su et al. 2014; Zhang et al. 2016). Here, we presented a first study confirming the benefits of using the SNP weighting approach for genomic prediction in plant breeding.

Several methods have been proposed to derive the weighting factors used to compute **G**. Most methods are based on using external weights, either from significant SNPs identified in a previous GWAS (de los Campos et al. 2013; Zhang et al. 2014) or from the posterior variances of marker effects estimated by Bayesian analysis (Zang et al 2010; Su et al. 2014). Our study used an alternative approach based on the iterative procedure developed by Wang et al. (2012). In this case, SNP weights are directly derived within the linear mixed model framework by converting GEBV into SNP effects and these weights are optimized iteratively to achieve higher predictive ability. Previous comparisons of weighting methods have shown minor differences in predictive performance (e.g., Wang et al. 2012; Su et al. 2014). The main advantage of the weighting method applied to K-BLUP in Chapter 5 is that weights are obtained without requiring additional GWAS or Bayesian analysis. In this Chapter, locus-specific variances were assumed since weights were applied to individual SNPs. An alternative is to assign common weights for groups of SNPs, where SNPs can be grouped based on a fixed length of genomic region, or based on a fixed number of adjacent SNPs. Some studies have shown that the group-marker weighting can reduce bias of GEBV and slightly increase prediction accuracy compared to single-marker weighting, apparently due to a reduction in the uncertainty of estimated weights (Su et al. 2014; Zhang et al. 2016). These studies used high-density SNP chips (> 45K SNPs); the differences between both weighting approaches still need to be investigated under lower marker densities.

Finally, a suitable feature of the method used in Chapter 5 is that the weighted **K** matrix can be directly used in different linear mixed models, providing a flexible framework for

model formulation. Furthermore, the SNP weights are based on information that is relevant for genomic prediction not only in the target breeding population but also in the TPE of interest. For instance, we computed environment-dependent weights derived from estimated SNP effects for yield under drought and under non-stress environments, which were the specific environmental condition of interest for improving adaptation. Although computing environment-dependent SNP weights had negligible effect on predictions in our study, mainly due to the high additive-genetic correlation between D-ET and W-ET, more benefits could be expected from this approach for genomic prediction in a TPE with stronger G×E effects.

In summary, results from Chapter 5 have proved that combining pedigree and genomic information while accounting for locus-specific variances with weighted K-BLUP can be very effective in increasing prediction quality of parental breeding values and should be implemented for improving grain yield productivity of sorghum through GS.

6.4. Perspective and future research

This thesis explored statistical modelling strategies to cope with different challenges encountered in modern plant breeding. Although the research presented in this thesis was performed using data from sorghum and the discussion of results was mainly concentrated on sorghum breeding, the methodologies and outcomes of this thesis are generally relevant for other crops as well. In addition, the results of the analyses presented in this study provide a framework for future research; some directions are suggested in the following paragraphs.

An important initial step in the plant breeding pipeline is the appropriate analysis of genotypes' performance in individual field trials. Chapter 2 showed how the new method for spatial adjustment of field trends, SpATS, can be applied to simplify the analysis of series of plant breeding trials while obtaining competitive results when compared to the more elaborate conventional method. This study focused in on modelling spatial correlations for traits that were measured at a single time. However, due to recent developments in high throughput phenotyping techniques, the breeder is increasingly having access to series of repeated measurements of the same trait over time. Data is also collected at multiple times in breeding trials of perennial crops, such as sugarcane or forage grasses, where genotypes are harvested repeatedly during the experiment (Pastina et al. 2012; Piepho and Eckl 2014). For these situations, an interesting area of research would be the extension of the SpATS method to account for spatial correlation as well as the temporal correlation induced by the repeated measures nature of the data. Smoothing spatio-temporal data using a three-dimensional P-spline function has been previously proposed in the literature (e.g., Lee and Durbán 2011). However, more research is still needed to adapt this methodology to the specificities of plant breeding trials.

The rest of the research presented in this thesis focused on evaluating different modelling strategies to improve genomic predictions of parental breeding values in sorghum (Chapters 3, 4 and 5). This topic is particularly relevant for sorghum and other hybrid crops since the implementation of GS opens the possibility to perform early selection of parental lines on the basis of GEBV instead of using testcross progeny testing, reducing time and cost of field evaluation. Chapter 3 started by exploring the relative value of using genomic information instead of traditional pedigree-based information for prediction of breeding values. It was demonstrated that replacing **A** by **G** in prediction models was always beneficial to improve predictive ability for different traits. However, the best predictive results were achieved when both relationship matrices were combined optimally. It should be noted that the results are dependent on the quality of information contained in **A** and **G**. We used a deep and well-curated pedigree, and thus little improvement in the information provided by **A** can be attained. A remaining question is, however, if the complete pedigree or a reduced number of generations should be used to optimize predictive ability of combined models. Additionally, for computation of **G** all available SNPs were used, following a pure GS approach, i.e., assuming that all markers are in LD with at least one QTL controlling the trait of interest. However, an alternative opportunity for research may be to evaluate the effect of computing **G** using SNPs that were previously selected based on their effects on the trait of interest, as assessed from GWAS. While this option may not be attractive for grain yield, the use of pre-selected SNPs could be more suitable for other traits, such as stay-green and flowering time, for which significant QTL regions have been identified in sorghum (Mace et al. 2013; Borrell et al. 2014b).

Chapter 4 investigated if genomic predictions for yield and drought adaptability in sorghum can be further improved by exploiting information from several traits simultaneously through joint multi-trait analysis. We concluded that this methodology is effective to increase the quality of predictions for both traits when plant height information is also incorporated in multi-trait genomic models. The research considered traits that are routinely measured in sorghum—and other crops—breeding programs using traditional phenotyping procedures. In recent years, remote sensing and hyperspectral imaging have facilitated low-cost and efficient collection of data on a large number of genotypes for many "new" traits that are associated with yield, such a spectral vegetation indices and canopy temperature (Montesinos-López et al. 2017). Future research should investigate the application of the multi-trait approach developed in Chapter 4 to make use of these new phenotypes as auxiliary traits to predict yield and stay-green in sorghum.

While Chapters 3 and 4 considered multi-environment genomic prediction across the whole TPE, Chapter 5 also addressed predictions for specific types of environments within a subdivided TPE. The latter Chapter provided a first insight on how information from yield performance in drought stress and non-stress environments can be used to improve GS for

adaptability in sorghum. The main outcome of the research is that G×E should be modelled in genomic models not only for predicting specific adaptation, but also for broad adaptation across the TPE. A potential limitation of this study lays on the classification of the environmental types since we used only the presence of stay-green phenotype as a general indicator of drought conditions. This classification may be masking stronger G×E effects because the variations in the degree of drought stress within D-ET are ignored. More refined conclusions on sorghum adaptation may be obtained in future studies by using better definitions of environmental types, for instance, based on direct characterization of water availability patterns (Bustos-Korts et al. 2019; Rincent et al. 2019).

Interestingly, important improvements in prediction of grain yield were achieved when heterogeneous variances of SNP effects were accounted for in genomic analysis (Chapter 5). This finding suggests that using the SNP weighting method can be an effective strategy to optimize predictive performance when working with low marker density, even for traits that are largely polygenic in nature. Our study on weighted K-BLUP was restricted to the use of grain yield data alone. An interesting extension of this method would be to include information from other traits associated with yield in a multi-trait analysis. For this, multi-trait models should accommodate not only heterogeneous variances of SNP effects for each trait, but also heterogeneous covariances of SNP effects between traits. The multi-trait weighting approach has been recently implemented in G-BLUP for the animal breeding context using simulated data (Karaman et al. 2018). However, the performance of this approach still remains unexplored for real applications in plant breeding.

In summary, results from this thesis show promising prospects for the use of GS in sorghum breeding programs. One of the major obstacles for the deployment of GS in existing breeding programs is still the high investment costs required. In this context, the adoption of the modeling methodologies implemented in this thesis may be valuable to make efficient use of all sources of available information, contributing to close the gap between prospects and real application of GS.

List of References

Acquaah G (2012) Breeding Sorghum. *Principles of Plant Genetics and Breeding*, Second Edition, John Wiley & Sons, Ltd, Chichester, UK.

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci* 93: 743–752

Albrecht T, Auinger HJ, Wimmer V, Ogutu JO, Knaak C, Ouzunova M, Piepho HP, Schön CC (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. *Theor Appl Genet* 127:1375–1386

Albrecht T, Wimmer V, Auinger HJ, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön CC (2011) Genome-based prediction of testcross values in maize. *Theor Appl Genet* 123:339–350

Ashraf B, Edriss V, Akdemir D, Autrique E, Bonnett D, Crossa J, Janss L, R Singh et al. (2016) Genomic prediction using phenotypes from pedigree lines with no markers. *Crop Sci* 56:957–964

Atlin GN, Baker RJ, McRae KB, Lu X (2000) Selection response in subdivided target regions. *Crop Sci* 40:7–13

Atlin GN, Kleinknecht K, Singh KP, Piepho HP (2011) Managing genotype × environment interaction in plant breeding programs: A selection theory approach. *J Indian Soc Agric Statistics* 65:237–247

Auinger HJ, Schönleben M, Lehermeier C, Schmidt M, Korzun V, Geiger HH, Piepho HP, Gordillo A et al. (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). *Theor Appl Genet* 129:2043–2053

Bartlett MS (1978) Nearest neighbour models in the analysis of field experiments. *J R Stat Soc* Ser B 40:147–174

Basford KE, Williams ER, Cullis BR, Gilmour A (1996) Experimental design and analysis for variety trials. p. 125–138. *In*: *Plant adaptation and crop improvement*, Cooper M and Hammer GL (ed), CAB Int., Wallingford, UK.

Bernal-Vasquez AM, Möhring J, Schmidt M, Schönleben M, Schön CC, Piepho HP (2014) The importance of phenotypic data analysis for genomic prediction—a case study comparing different spatial models in rye. *BMC Genom* 15:646

Bernal-Vasquez AM, Möhring J, Schmidt M, Schönleben M, Schön CC, Piepho HP (2014) The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. *BMC Genom* 15(1):646. doi: 10.1186/1471-2164-15-646

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci* 48: 1649–1664

Bernardo R (2010) *Breeding for Quantitative Traits in Plants*, Stemma Press, Woodbury, Minnesota

Bernardo R (2014) Genomewide selection when major genes are known. *Crop Sci* 54:68–75

Bink MCAM, Uimari P, Sillanpaa MJ, Janss LLG, Jansen RC (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *Theor Appl Genet* 104:751–762

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A Mixed-Model Quantitative Trait Loci (QTL) Analysis for Multiple-Environment Trial Data Using Environmental Covariables for QTL-by-Environment Interactions, with an Example in Maize. *Genetics* 177: 1801–1813

Borrell AK, Hammer GL, Henzell RG (2000) Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. *Crop Sci* 40(4):1037–1048

Borrell AK, Jordan DR, Mullet JE, Henzell RG, Hammer G (2006) Drought adaptation in sorghum. *In*: J-M Ribaut (Eds). *Drought adaptation in cereals*, Haworth Press Inc., New York, pp 335–339

Borrell AK, Mullet JE, George-Jaeggli B, van Oosterom EJ, Hammer GL, Klein PE, Jordan DR (2014a) Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. *J Exp Bot* 65(21): 6251–6263

Borrell AK, van Oosterom EJ, Mullet JE, George-Jaeggli B, Jordan DR, Klein PE, Hammer GL (2014b) Stay-green alleles individually enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns. *New Phytol* 203(3): 817–830

Brownie C, Bowman DT, Burton JW (1993) Estimating spatial variation in analysis of data from yield trials: a comparison of methods. *Agron J* 85:1244–1253

Brownie C, Gumpertz ML (1997) Validity of spatial analyses for large field trials. *JABES* 2(1):1–23

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci* 52(2):707–719

Bush WS, Moore JH (2012) Chapter 11: Genome-wide association studies. *PLoS Comput Biol* 8(12):e1002822

Bustos-Korts D, Malosetti M, Chapman S, Biddulph B, van Eeuwijk FA (2016) Improvement of predictive ability by uniform coverage of the target genetic space. *G3* 6:3733–3747

Bustos-Korts D, Malosetti M, Chapman S, Chenu K, Boer M, Zheng B, van Eeuwijk (2019) From QTLs to adaptation landscapes: using genotype to phenotype models to characterize G×E over time. *Front Plant Sci* 10:1540

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (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

Butler DG., Cullis BR, Gilmour AR, Gogel BG, Thompson R (2017) ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Calus MP, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol* 43:26

Cappa EP, Cantet RJC (2007) Bayesian estimation of a surface to account for a spatial trend using penalized splines in an individual-tree mixed model. *Can J For Res* 37:2677–2688

Cappa EP, Lstiburek M, Yanchuk AD, El-Kassaby YA (2011) Two-dimensional penalized splines via Gibbs sampling to account for spatial variability in forest genetic trials with small amount of information available. *Silvae Genet* 60:25–35

Cappa EP, Muñoz F, Sanchez L, Cantet RJC (2015) A novel individual-tree mixed model to account for competition and environmental heterogeneity: a Bayesian approach. *Tree Genet Genomes* 11: 120.

Chapman SC, Cooper M, Hammer GL, Butler DG (2000) Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Aust J Agric Res* 51:209–221

Cheng R, Borevitz J, Doerge R (2013) Selecting informative traits for multivariate quantitative trait locus mapping helps to gain optimal power. *Genetics* 195:683–691

Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer MF, Chapman SC (2011) Environment characterization as an aid to wheat improvement. Interpreting genotype-environment interactions by modelling water-deficit patterns in North-Eastern Australia. *J Exp Bot* 62(6):1743–1755

Christensen OF (2012) Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet Sel Evol* 44:37

Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. *Genet Sel Evol* 42:2

Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G (2012) Single-step methods for genomic evaluation in pigs. *Animal* 6:1565–1571

Christopher JT, Manschadi AM, Hammer GL, Borrell AK (2008) Developmental and physiological traits associated with high yield and stay-green phenotype in wheat. *Aust J Agric Res* 59:354-364

Clark S, Hickey J, Van Der Werf J (2011) Different models of genetic variation and their effect on genomic evaluation. *Genet Sel Evol* 43:18.

Comstock RE (1977) Quantitative genetics and the design of breeding programmes. *In*: Pollack E, Kempthorne O, Bailey TBJ (eds) *Proceedings of the International Conference of Quantitative Genetics*. Iowa State University Press, Ames, Iowa, pp 705–718

Cooper M, Gho C, Leafgren R, Tang T, Messina C (2014) Breeding drought-tolerant maize hybrids for the US corn-belt. Discovery to product. *J Exp Bot* 65(21):6191–6204

Cooper M., Hammer GL (1996) Synthesis of strategies for crop improvement. *In*: M. Cooper & G.L. Hammer (Eds.). *Plant Adaptation and Crop Improvement*. CAB International, Wallingford, England, pp 591–623

Corrada Bravo H, Wahba G, Lee KE, Klein BEK, Klein R, Iyengar SK (2009) Examining the relative influence of familial, genetic and environmental covariate information in flexible risk models. *Proc Natl Acad Sci* 106: 8128–8133

Crossa J, de los Campos G, Pérez-Rodríguez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquin D, de los Campos G, Burgueño J, González-Camacho JM et al. (2017) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci* 22(11):961–975

Cuevas J, Crossa J, Montesinos-López OA, Burgueño J, Pérez-Rodríguez P et al. (2017) Bayesian genomic prediction with genotype × environment interaction kernel models. *G3* 7(1):41–53

Cuevas J, Crossa J, Soberanis V, Pérez-Elizalde S, Pérez-Rodríguez P et al. (2016) Genomic prediction of genotype × environment interaction kernel regression models. *Plant Genome* 9: 1–20

Cullis BR, Gleeson, AC (1991) Spatial analysis of field experiments – an extension to two dimensions. *Biometrics* 47:1449–1460.

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

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. *JABES* 11:381–393

Cullis BR, Thomson FM, Fisher JA, Gilmour AR, Thompson R (1996) The analysis of the NSW wheat variety database. II. Variance component estimation. *Theorl Appl Genet* 92:28–39

Currie ID, Durbán M (2002) Flexible smoothing with P-splines: a unified approach. *Stat Model* 2:333–349

Currie ID, Durbán M, Eilers PHC (2006) Generalized linear array models with applications to multidimensional smoothing. *J R Statist Soc* Ser B 68:259–280

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185: 1021–1031

Dahlberg J, Berenji J, Sikora V and Latković D (2011) Assessing sorghum [*Sorghum bicolor* (L.) Moench] germplasm for new traits: food, fuels & unique uses. *Maydica* 56:85–92

Daryanto S, Wang L, Jacinthe PA (2017) Global synthesis of drought effects on cereal, legume, tuber and root crops production: A review. *Agric Water Manag* 179:18–33

Dawson JC, Endelman JB, Heslot N, Crossa J, Poland J, Dreisigacker S, Manès Y, Sorrells ME, Jannink JL (2013) The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crop Res* 154:12–22

de la Mata R, Zas R (2010) Transferring Atlantic maritime pine improved material to a region with marked Mediterranean influence in inland NW Spain: a likelihood-based approach on spatially adjusted field data. *Eur J Forest Res* 129:645–658

de los Campos G, Gianola D, Rosa GJ, Weigel KA, Crossa J (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet Res (Camb)* 92: 295–308.

de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375–385

de los Campos G, Sorensen D, Gianola D (2015) Genomic heritability: what is it? *PLoS Genet* 11:e1005048

de los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D (2013) Prediction of complex human traits using the Genomic Best Linear Unbiased Predictor. *PLoS Genet* 9:e1003608

Dekkers JCM (2007) Marker-assisted selection for commercial crossbred performance. *J Anim Sci* 85:2104-2114

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet* 3:22–32

DeLacy IH, Basford KE, Cooper M, Bull JK, McLaren CG (1996) Analysis of multi environmental trials: an historical perspective. *In*: Cooper M, Hammer GL (eds) *Plant adaptation and crop improvement*. Wallingford, UK, pp 39–124

Denis JB, Piepho HP, van Eeuwijk HP (1997) Modelling expectation and variance for genotype by environment data. *Heredity (Edimb)* 79:162–171

dos Santos JPR, Vasconcellos RCC, Pires LPM, Balestre M, Von Pinho RG (2016) Inclusion of dominance effects in the multivariate GBLUP model. *PLoS ONE* 11(4):e0152045

Duncan RR, Bockholt AJ, Miller FR (1981) Descriptive comparison of senescent and nonsenescent Sorghum genotypes. *Agron J* 73(5):849–853

Durbán M, Currie ID, Kempton R (2001) Adjusting for fertility and competition in variety trials. *J Agric Sci (Camb.)* 136:129–140

Durbán M, Hackett CA, McNicol JW, Newton AC, Thomas WTB, Currie ID (2003). The practical use of semiparametric models in field trials. *JABES* 8:48–66

Dutkowski GW, Costa e Silva J, Gilmour AR, Lopez GA (2002) Spatial analysis methods for forest genetic trials. *Can J For Res* 32:2201–2214

Dutkowski GW, Costa e Silva J, Gilmour AR, Wallendorf H, Aguiar A (2006) Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Can J For Res* 36:1851–1870

Edmondson RN (1993) Systematic row-and-column designs balanced for low order polynomial interactions between rows and columns. *J R Statist Soc* Ser B 55:707–723

Eilers PHC (1999) Discussion on: the analysis of designed experiments and longitudinal data by using smoothing splines (by Verbyla et al.). *J R Statist Soc* Ser C 48:307–308

Eilers PHC, Marx BD (1996) Flexible smoothing with B-splines and penalties. *Stat Sci* 11:89–102

Eilers PHC, Marx BD (2003) Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometr Intell Lab Syst* 66:159–174

Eilers PHC, Marx BD, Durbán M (2015) Twenty years of P-splines. *SORT-Stat Oper Res T* 39(2):149-186

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*, Pearson Prentice Hall, Harlow, Essex

Federer WT (1998) Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle design experiments. *Biometrics* 54:471–481

Fernandes SB, Dias KOG, Ferreira DF, Brown PJ (2018) Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theor Appl Genet* 131:747–755

Fernando RL, Cheng H, Golden BL, Garrick DJ (2016) Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. *Genet Sel Evol* 48(1):96

Fernando RL, Gianola D (2018) Bayesian inference of genomic similarity among individuals from markers and phenotypes. *Proceedings of the World Congress on Genetics Applied to Livestock Production*, Volume Theory to Application 2, 942

Fisher RA (1918) The correlation between relatives under the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh* 52:399–433

Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol* 43:1

Frensham A, Cullis BR, Verbyla A (1997) Genotype by environment variance heterogeneity in a two-stage analysis. *Biometrics* 53:1373–1383

Fu YB, Yanchuk AD, Namkoong G (1999) Spatial patterns of tree height variations in a series of Douglas-fir progeny trials: implications for genetic testing. *Can J For Res* 29:714–723

Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, Su G (2012) Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet Sel Evol* 44:8

Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 41:55

Gianola D (2013) Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 194: 573–596

Gianola D, de los Campos G (2008) Inferring genetic values for quantitative traits non-parametrically. *Genet Res* 90: 525–540

Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando RL (2009) Additive genetic variability and the Bayesian alphabet. *Genetics* 183(1):347–63

Gianola D, de los Campos G, Toro MA, Naya H, Schön CC, Sorensen D (2015) Do molecular markers inform about pleiotropy? *Genetics* 201:23–9.

Gianola D, de los Campos G, Toro MA, Naya H, Schön CC, Sorensen D (2015) Do molecular markers inform about pleiotropy? *Genetics* 201(1):23–9

Gianola D, Schön CC (2016) Cross-validation without doing across-validation in genome-enabled prediction. *G3* 6: 3107–3128

Gianola D, Weigel KA, Krämer N, Stella A, Schön C-C (2014) enhancing genome-enabled prediction by bagging Genomic BLUP. *PLoS ONE* 9(4):e91693

Gilmour AR (2000) Post blocking gone too far! Recovery of information and spatial analysis in field experiments. *Biometrics* 56(3): 944–945

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments. *JABES* 2:269–293

Goddard ME (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257

Goddard ME, Hayes BJ, Meuwissen THE (2011) Using the genomic relationship matrix to predict the accuracy of genomic selection. *J Anim Breed Genet* 128:409–21

Goldberger AS (1962) Best linear unbiased prediction in the generalized linear regression model. *J Amer Statist Assoc* 57:369–375

González-Recio O, Rosa GJM, Gianola D (2014) Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest Sci* 166:217–231

Gregersen PL, Culetic A, Boschian L, Krupinska K (2013) Plant senescence and crop productivity. *Plant Mol Biol* 82(6):603–22

Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G (2014) Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet* 15:30

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397

Habier D, Fernando RL, Dekkers JCM (2009) Genomic selection using low density marker panels. *Genetics* 182:343–353

Habier D, Fernando RL, Garrick DJ (2013) Genomic-BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194:597–607

Haile JK, N'Diaye A, Clarke F, Clarke J, Knox R, Rutkoski, J et al. (2018) Genomic selection for grain yield and quality traits in durum wheat. *Mol Breed* 38:75

Haile-Mariam M, Nieuwhof GJ, Beard KT, Konstatinov KV, Hayes BJ (2013) Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations. *J Anim Breed Genet* 130: 20–31

Hammer GL, McLean G, Chapman S, Zheng B, Doherty A, Harrison MT, van Oosterom E, Jordan DR (2014) Crop design for specific adaptation in variable dryland production environments. *Crop Pasture Sci* 65(7):614–626

Hayes BJ, Goddard M (2001) The distribution of the effects of genes affecting quantitative traits in livestock. *Genet Sel Evol* 33:209–229

Hayes BJ, Vissher PM, Goddard M E (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res* 91:47–60

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. *Crop Sci* 49:1–12

Henderson CR (1973) Sire Evaluation and Genetic Trends. *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush*, 10–41, Champaign: American Society of Animal Science and the American Dairy Science Association.

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447

Henderson CR, Quass RL (1976) Multiple trait evaluation using relatives' records. *J Anim Sci* 43:1188–1197

Henzell RG, Gillieron W (1973) Effects of partial and complete panicle removal on the rate of death of some Sorghum bicolor genotypes under moisture stress. *Qld J Agric Anim Sci* 30:291–299

Heslot N, Akdemir D, Sorrells ME, Jannink JL (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet* 127:463–480

Heslot N, Jannink JL, Sorrells ME (2015) Perspectives for genomic selection applications and research in plants. *Crop Sci* 55(1): 1–55

Heslot N, Yang HP, Sorrells ME, Jannink JL (2012) Genomic selection in plant breeding: a comparison of models. *Crop Sci* 52:146–160

Holland JB (2006) Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS proc MIXED. *Crop Sci* 46:642–654

Hunt CH, van Eeuwijk FA, Mace ES, Hayes BJ, Jordan DR (2018) Development of genomic prediction in sorghum. *Crop Sci* 58:690–700

Ilska JJ, Meuwissen THE, Kranis A, Woolliams JA (2017) Use and optimization of different sources of information for genomic prediction. *Genet Sel Evol* 49:90

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177

Jensen J, Su G, Madsen P (2012) Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. *BMC Genet* 13:44

Jia Y, Jannink JL (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192(4):1513–22

John JA, Williams ER (1995) *Cyclic and computer generated designs*. 2nd ed. Chapman & Hall

Jordan DR, Hunt CH, Cruickshank AW, Borrell AK, Henzell RG (2012) The relationship between the stay-green trait and grain yield in elite sorghum hybrids grown in a range of environments. *Crop Sci* 52:1153–1161

Jordan DR, Mace ES, Cruickshank AW, Hunt CH, Henzell RG (2011) Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Sci* 51:1444–1457

Jordan DR, Tao Y, Godwin ID, Henzell RG, Cooper M, McIntyre CL (2003) Prediction of hybrid performance in grain sorghum using RFLP markers. *Theor Appl Genet* 106:559–567

Jordan DR, Tao YZ, Godwin ID, Henzell1 RG, Cooper M, McIntyre CL (2004) Comparison of identity by descent and identity by state for detecting genetic regions under selection in a sorghum pedigree breeding program. *Mol Breed* 14:441–454

Karaman E, Lund MS, Anche MT, Janss L, Su G (2018) Genomic prediction using multi-trait weighted GBLUP accounting for heterogeneous variances and covariances across the genome. *G3* 8(11):3549–3558

Kelly AM, Smith AB, Eccleston JA, Cullis BR (2007) The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Sci* 47:1063–1070

Kulwal PL (2016) Association Mapping and Genomic Selection—Where Does Sorghum Stand? *In*: Rakshit S, Wang YH (eds) *The Sorghum Genome. Compendium of Plant Genomes*. Springer, Cham

Lado B, Matus I, Rodriguez A, Inostroza L, Poland J et al. (2013) Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3* 3:2105–2114

Lado B, Vázquez D, Quincke M, Silva P, Aguilar I, Gutiérrez L (2018) Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor Appl Genet* 131(12):2719–2731

Lee DJ, Durbán M (2011) P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Stat modelling* 11(1):49–69.

Lee DJ, Durbán M, Eilers PHC (2013) Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. *Comput Stat Data Anal* 61:22–37

Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. *J Dairy Sci* 92:4656–4663

Legarra A, Christensen OF, Aguilar I, Misztal I (2014) Single Step, a general approach for genomic selection. *Livest Sci* 166:54–65

Legarra A, Robert-Granié C, Manfredi E, Elsen JM (2008) Performance of genomic selection in mice. *Genetics* 180(1):611–618

Lehermeier C, Wimmer V, Albrecht T, Auinger H, Gianola D, Schmid VJ, Schön CC (2012) Sensitivity to prior specification in Bayesian genome-based prediction models. *Stat Appl Genet Mol Biol* 12(3):375–391

Leiser WL, Rattunde HF, Piepho HP, Parzies HK (2012) Getting the most out of sorghum low-input field trials in West Africa using spatial adjustment. *J Agron Crop Sci* 198:349–359

Lesk C, Rowhani P, Ramankutty N (2016) Influence of extreme weather disasters on global crop production. *Nature* 529:84–87.

Liu SM, Constable GA, Cullis BR, Stiller WN, Reid PE (2015) Benefit of spatial analysis for furrow irrigated cotton breeding trials. *Euphytica* 201:253–264

Liu Z, Goddard ME, Reinhardt F, Reents R (2014) A single-step genomic model with direct estimation of marker effects. *J Dairy Sci* 97:5833–5850

Liu Z, Seefried FR, Reinhardt F, Rensing S, Thaller G, Reents R (2011) Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet Sel Evol* 43:19

Loo-Dinkins JA, Tauer CG, Lambeth CC (1990) Selection system efficiencies for computer simulated progeny test field designs in loblolly pine. *Theor Appl Genet* 79:89–96

Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink JL, Singh RP, Autrique E, de los Campos G (2015) Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3* 5(4):569–582

Luan T, Woolliams JA, Ødegård J, Dolezal M, Román-Ponce SI, Bagnato A, Meuwissen THE (2012) The importance of identity-by-state information for the accuracy of genomic selection. *Genet Sel Evol* 44:28

Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Lyra DH, Mendonça LF, Galli G, Alves FC, Granato ISC, Fritsche-Neto R (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. *Mol Breed* 37:80

Mace ES, Hunt CH, Jordan DR (2013) Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theor Appl Genet* 126:1377–1395

Mace ES, Singh V, van Oosterom E, Hammer GL, Hunt CH, Jordan DR (2012) QTL for nodal root angle in sorghum (*Sorghum bicolor* L. Moench) co-locate with QTL for traits associated with drought adaptation. *Theor Appl Genet* 124:97–109

Magnussen S (1993) Bias in genetic variance estimates due to spatial autocorrelation. *Theor Appl Genet* 86:349–355

Magnussen S (1994) A method to adjust simultaneously for spatial microsite and competition effects. *Can J For Res* 24:985–995

Maher B (2008) Personal genomes: The Case of the Missing Heritability. *Nature News* 456(7218):18–21.

Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA (2016) Predicting responses in multiple environments: issues in relation to genotype × environment interactions. *Crop Sci* 56(5):2210–2222

Malosetti M, Linden CG, Vosman B, Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* 175:879–889

McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D et al. (2018) The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J* 93(2):338–354

McGuire S (2012) WHO,World Food Programme, and International Fund for Agricultural Development. The State of Food Insecurity in the World 2012. Economic growth is necessary but not sufficient to accelerate reduction of hunger and malnutrition. Rome, *FAO Adv Nutr* 2012 4:126–127

McLean RA, Sanders WL, Stroup WW (1991) A unified approach to mixed linear models. *Am Stat* 45:54–64

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149:383–403

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829

Meuwissen THE, Hayes BJ, Goddard ME (2016) Genomic selection: A paradigm shift in animal breeding. *Anim Front* 6(1):6–14

Meyer K (1989) Approximate accuracy of genetic evaluation under an individual animal model. *Livest Prod Sci* 21:87–100

Möhring J, Piepho HP (2009) Comparison of weighting in two-stage analyses of series of experiments. *Crop Sci* 49:1977–1988

Momen M, Mehrgardi AA, Sheikhy A, Esmailizadeh A, Fozi MA, Kranis A, Valente BD et al. (2017) A predictive assessment of genetic correlations between traits in chickens using markers. *Genet Sel Evol* 49:16

Montesinos-López OA, Montesinos-López A, Crossa J, de los Campos G, Alvarado G, Suchismita M, Rutkoski J, González-Pérez L, Burgueño J (2017) Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* 13:4

Mrode RA (2005) *Linear Models for the Prediction of Animal Breeding Values*. 2nd ed. CABI Publishing, Oxfordshire, UK.

Müller BU, Kleinknecht K, Möhring J, Piepho HP (2010) Comparison of spatial models for sugar beet and barley trials. *Crop Sci* 50:794–802

Oakey H, Cullis BR, Thompson R, Comadran J, Halpin C, Waugh R (2016) Genomic selection in multi-environment crop trials. *G3* 6:1313–1326

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006) Joint modelling of additive and non-additive genetic line effects in single field trials. *Theor Appl Genet* 113:809–819

Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. *Theor Appl Genet* 109:508–514

Pastina MM, Malosetti M, Gazaffi R, Mollinari M, Margarido GRA, Oliveira KM, Pinto LR, Souza AP, van Eeuwijk FA, Garcia AAF (2012) A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theor Appl Genet* 124:835–849

Paterson AH (2008) Genomics of sorghum. *Int J Plant Genomics* 2008:362–451

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556

Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 31:100–109

Patterson HD, Williams ER, Hunter EA (1978) Block designs for variety trials. *J Agric Sci (Camb)* 90:395–400

Payne RW (2006) New and traditional methods for the analysis of unreplicated experiments. *Crop Sci* 46:2476–2481

Piepho HP (1997) Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics* 53:761–767

Piepho HP, Eckl T (2014) Analysis of series of variety trials with perennial grasses. *Grass Forage Sci* 69:431–440

Piepho HP, Möhring J (2005) Best linear unbiased prediction of cultivar effects for subdivided target regions. *Crop Sci* 45:1151–1159

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008a) BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161:209–228

Piepho HP, Möhring J, Pflugfelder M, Hermann W, Williams ER (2015) Problems in parameter estimation for power and AR(1) models of spatial correlation in designed field experiments. *Communications in Biometry and Crop Science* 10(1):3–16

Piepho HP, Richter C, Williams ER (2008b) Nearest neighbour adjustment and linear variance models in plant breeding trials. *Biom J* 50(2):164–189

Piepho HP, Williams ER (2010) Linear variance models for plant breeding trials. *Plant Breed* 129:1–8

Qiao CG, Basford KE, Delacy IH, Cooper M (2000) Evaluation of experimental designs and spatial analyses in wheat breeding trials. *Theor Appl Genet* 100:9–16

Qiao CG, Basford KE, DeLacy IH, Cooper M (2004) Advantage of single-trial models for response to selection in wheat breeding multi-environment trials. *Theor Appl Genet* 108:1256–1264

R Development Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

R Development Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* 8:e66428

Rice B, Lipka AE (2019) Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *Plant Genome* 12:180052

Rincent R, Malosetti M, Ababaei B, Touzy G, Mini A, Bogard M, Martre P, Le Gouis J, van Eeuwijk FA (2019) Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. *Theor Appl Genet* 132:3399

Robinson GK (1991) That BLUP is a good thing: The estimation of random effects. *Stat Sci* 6:15–32

Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC (2016a) Spatial Models for Field Trials. arXiv:1607.08255v1 [stat. ME]

Rodríguez-Álvarez MX, Boer MP, Eilers PHC, van Eeuwijk FA (2016b) SpATS: spatial analysis of field trials with splines. R package version 1.0-4. https://cran.r-project.org/package=SpATS

Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC (2018a) Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spat Stat* 23:52–71

Rodríguez-Álvarez MX, Boer MP, Eilers PHC, van Eeuwijk FA (2018b) SpATS: spatial analysis of field trials with splines. R package version 1.0–8. https://CRAN.R-project.org/package=SpATS

Rodríguez-Ramilo ST, García-Cortés LA, González-Recio O (2014) Combining genomic and genealogical information in a reproducing kernel Hilbert spaces regression model for genome-enabled predictions in dairy cattle. *PLoS ONE* 9:e93424

Runcie D, Cheng H (2019) Pitfalls and remedies for cross validation with multi-trait genomic prediction methods. *G3* 9(11):3727–3741

Ruppert D, Wand M, Carroll R (2003) *Semiparametric Regression*. Cambridge University Press.

Rutkoski JE, Poland J, Mondal S, Autrique E, González Pérez L, Crossa, J et al. (2016) Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3* 6(9):2799–808

Sandhu N, Dixit S, Swamy BM, Vikram P, Venkateshwarlu C, Catolos M, Kumar A (2018) Positive interactions of major-effect QTLs with genetic background that enhances rice yield under drought. *Sci Rep* 8(1):1626

Sarinelli JM, Murphy JP, Tyagi P, Holland JB, Johnson JW, Mergoum M, Mason RE et al. (2019) Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theor Appl Genet* 132:1247–1261

Sarker A, Singh M (2015) Improving breeding efficiency through application of appropriate experimental designs and analysis models: A case of lentil (*Lens culinaris* Medikus subsp. *culinaris*) yield trials. *Field Crops Res* 179:26–34

Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167:485–498

Schopp P, Müller D, Wientjes YCJ, Melchinger AE (2017) Genomic prediction within and across biparental families: means and variances of prediction accuracy and usefulness of deterministic equations. *G3* 193(2):621–631

Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y (2016) Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor Appl Genet* 129(2):273–287

Schulthess, A. W., Zhao, Y., Longin, C. F. H., Reif, J. C. (2018). Advantages and limitations of multiple-trait genomic prediction for Fusarium head blight severity in hybrid wheat (*Triticum aestivum* L.). *Theor Appl Genet* 131(3):685–701

Selle ML, Steinsland I, Hickey JM, Gorjanc G (2019) Flexible modelling of spatial variation in agricultural field trials with the R package INLA. *Theor Appl Genet* 132:3277

Smith AB, Cullis BR, Gilmour A (2001a) The analysis of crop variety evaluation data in Australia. *Aust N Z J Stat* 43(2):129–145

Smith AB, Cullis BR, Luckett D, Hollamby G, Thompson R (2002) Exploring variety–environment data using random effects AMMI models with adjustments for spatial field trend. Part 2: Applications. *In*: Kang MS (ed) *Quantitative genetics, genomics and plant breeding*. Wallingford, UK: CABI Publishing, 337–351

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

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *J Agric Sci (Camb)* 143:449–462

Spindel JE, Begum H, Akdemir D, Collard B, Redona E, Jannink JL et al. (2016) Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity (Edimb)* 116:395–408

Stefanova KT, Smith AB, Cullis BR (2009) Enhanced diagnostics for the spatial analysis of field trials. *JABES* 14:392–410

Steiger JH (1980) Tests for comparing elements of a correlation matrix. *Psychol Bull* 87(2):245–251

Stroup WW, Baenziger PS, Mulitze DK (1994) Removing spatial variation from wheat yield trials: a comparison of methods. *Crop Sci* 34:62–66

Su G, Christensen OF, Janss L, Lund MS (2014) Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *J Dairy Sci* 97:6547–6559

Sukumaran S, Crossa J, Jarquin D, Lopes M, Reynolds MP (2017) Genomic prediction with pedigree and genotype × environment interaction in spring wheat grown in South and West Asia, North Africa, and Mexico. *G3* 7(2):481–495

Sun J, Rutkoski JE, Poland, J, Crossa J, Jannink, JL, Sorrells ME (2017) Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome* 10(2)

Talbot M (1984) Yield variability of crop varieties in the UK. *J Agric Sci (Camb.)* 102:315–321

Talwar HS, Elangovan M, Patil JV (2013) Sorghum - A potential crop to adapt to future climate change scenario. *In*: Elangovan M and JV Patil (Eds.). *Managing Intellectual Property under PVP and PGR*. Directorate of Sorghum Research (DSR), Rajendranagar, Hyderabad 500030, Andhra Pradesh, India, pp 254–256

Taye G, Njuho PM (2008) Smoothing fertility trends in agricultural field experiments. *Statistics* 42(3):275–289

Thompson R, Meyer K (1986) A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest Prod Sci* 15(4):299–313

UNCTAD (2011) Water for Food. Innovative water management technologies for food security and poverty alleviation. United Nations, New York and Geneva, 1–3

Vadez V, Deshpande SP, Kholova J et al. (2011) Stay-green quantitative trait loci's effects on water extraction, transpiration efficiency and seed yield depend on recipient parent background. *Funct Plant Biol* 38(7):553–566

van Eeuwijk FA, Bink MCAM, Chenu K, Chapman S (2010) Detection and use of QTL for complex traits in multiple environments. *Curr Opin Plant Biol* 13:193–205

van Eeuwijk FA, Bustos-Korts D, Millet EJ, Boer MP, Kruijer W, Thompson A, Malosetti M, Iwata H, Quiroz R, Kuppe C, Muller O, Blazakis KN, Yu K, Tardieu F, Chapman SC (2019) Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci* 282:23–39

van Eeuwijk FA, Denis JB, Kang MS (1996) Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. *In*: MS Kang and HG Gauch (eds.) *Genotype-by-environment interaction*. CRC Press Inc., Boca Raton, Florida, 15–50

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Velazco JG, Jordan DR, Mace ES, Hunt CH, Malosetti M, van Eeuwijk FA (2019a) Genomic prediction of grain yield and drought-adaptation capacity in sorghum is enhanced by multi-trait analysis. *Front Plant Sci* 10:997

Velazco JG, Malosetti M, Hunt CH, Mace ES, Jordan DR, van Eeuwijk FA (2019b) Combining pedigree and genomic information to improve prediction quality: an example in sorghum. *Theor Appl Genet* 132(7):2055–2067

Velazco JG, Rodríguez-Álvarez MX, Boer MP, Jordan DR, Eilers PHC, Malosetti M, van Eeuwijk FA (2017) Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. *Theor Appl Genet* 130:1375–1392

Verbyla AP, Cullis BR, Kenward MG, Welham SJ (1999) The analysis of designed experiments and longitudinal data using smoothing splines (with discussion). *J R Stat Soc* Ser C 48:269–312

Verbyla AP, De Faveri J, Wilkie JD, Lewis T (2018) Tensor cubic smoothing splines in designed experiments requiring residual modelling. *JABES* 23:478

Visscher PM (2008) Sizing up human height variation. *Nat Genet* 40:489–490

Vitezica Z, Aguilar I, Misztal I, Legarra A (2011) Bias in genomic predictions for populations under selection. *Genet Res (Camb)* 93:357–366

Wang H, Misztal I, Aguilar I, Legarra A, Muir WM (2012) Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb)* 94:73–83

Wang X, Li L, Yang Z, Zheng X, Yu S, Xu C et al. (2017) Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity (Edinb)* 118(3):302–310

Wang X, Singh D, Marla S, Morris G, Poland J (2018) Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. *Plant Methods* 14:53

Wang X, Yang Z, Xu C (2015) A comparison of genomic selection methods for breeding value prediction. *Sci Bull* 60:925–935

Welham SJ, Gogel BJ, Smith AB, Thompson R, Cullis BR (2010) A comparison of analysis methods for late-stage variety evaluation trials. *Aust N Z J Stat* 52:125–149

Wientjes YCJ, Veerkamp RF, Calus MPL (2013) The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* 193:621–631

Wilkinson GN, Eckert SR, Hancock TW, Mayo O (1983) Nearest neighbour (NN) analysis of field experiments. *J R Stat Soc* Ser B 45:151–211

Williams ER, John JA, Whitaker D (2006) Construction of resolvable row-column designs. *Biometrics* 62:103–108

Williams ER, John JA, Whitaker D (2014) Construction of more flexible and efficient p-rep designs. *Aust N Z J Stat* 56:89–96

Wiltshire AJ, Kay G Gornall JL, Betts RA (2013) The impact of climate, CO2 and population on regional food and water resources in the 2050s. *Sustainability* 5:2129–2151

Wimmer V, Albrecht T, Auinger HJ, Schön CC (2012) synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28:2086–2087

Wimmer V, Lehermeier C, Albrecht T, Auinger HJ, Wang Y, Schön CC (2013) Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics* 195:573–587

Wood SN (2006) Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* 62(4):1025–1036

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci* 48:391–407

Yang J, Benyamin, B, Mcevoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME,Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42:565–569.

Yang RC, Ye TZ, Blade SF, Bandara M (2004) Efficiency of spatial analysis of field pea variety trials. *Crop Sci* 44:49–55

Yates F (1940) The recovery of inter-block information in balanced incomplete block designs. *Annals of Eugenics* 10:317–325

Yu X, Li X, Guo T, Zhu C, Wu Y, Mitchell SE, Roozeboom KL, Wang D et al. (2016) Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nat Plants* 2:16150

Zas R (2006) Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials. *Tree Genet Genomes* 2:177–186

Zhang J, Zhang S, Cheng M, Jiang H, Zhang X, Peng C, Lu X, et al. (2018) Effect of Drought on Agronomic Traits of Rice and Wheat: A Meta-Analysis. *Int J Environ Res Public Health* 15:839

Zhang X, Lourenco D, Aguilar I, Legarra A, Misztal I (2016) Weighting Strategies for Single-Step Genomic BLUP: An Iterative Approach for Accurate Calculation of GEBV and GWAS. *Front Genet* 7:151

Zhang Z, Liu J, Ding X, Bijma P, de Koning DJ, Zhang Q (2010) Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS ONE* 5:e12648.

Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J et al. (2014) Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS ONE* 9:e93017

Zimmerman DL, Harville DA (1991) A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* 47:223–239

Summary

The increasing global food demand in the context of climate change is a major concern of the 21st century. Developing crop varieties with improved adaptability to variable environmental conditions might be crucial to ensure food supply. Sorghum [*Sorghum bicolor* (L.) Moench] is a staple cereal crop in semi-arid regions of the world and is also grown worldwide as feed and bioenergy crop. Its drought-tolerance ability makes it a strategic crop for sustainable grain production. Modern approaches to quantitative genetics and statistical models such as genomic-assisted breeding techniques offer new opportunities for further improvements in crop productivity and adaptability. Compared to other major cereal crops such as maize and wheat, the application of these approaches remains largely unexplored for grain sorghum. The motivation of this thesis is to implement statistical models that exploit information from phenotypic, pedigree and genomic data for improving genetic evaluation and selection in sorghum breeding.

Chapter 2 assessed the performance of a novel method for spatial analysis of plant breeding field experiments based on two-dimensional smoothing with P-splines. This method was evaluated in comparison with the conventional spatial models by considering the improvement in precision and predictions of genetic effects in early generation sorghum breeding trials. The Chapter shows that both spatial methods produced equivalent performance. Differences in model parameterization as well as the advantages of the new spatial approach for routine application are discussed.

In Chapter 3, the impact of using pedigree and genomic information on prediction quality was explored for different traits in sorghum. For this, the Chapter proposes to use BLUP models fitting weighted combinations of pedigree and genomic relationship matrices, where the best-predictive combination is identified empirically in each prediction scenario. Results showed that the use of a merged pedigree–genomic matrix always improved predictive ability and unbiasedness of prediction relative to conventional G-BLUP, mainly for the traits with lower heritabilities. Based on these outcomes, the inclusion of pedigree information in genomic models is recommended to optimize predictions when the additive variation is not fully explained by markers.

Chapter 4 presents an extension of the study in Chapter 3 to the context of multi-trait genomic prediction. Specifically, we assessed the capacity of multi-trait models to improve genomic prediction for grain yield and stay-green in sorghum by using information from

correlated auxiliary traits. In general, results showed that genomic prediction for both target traits can be enhanced by combining information from specific sets of traits. Predictions from conventional multi-trait G-BLUP were also optimized by combining pedigree and genomic information.

Chapter 5 investigated the effect of modelling genotype-by-environment interaction (G×E) on genomic prediction for grain yield in drought-stress and non-stress environments. Results indicated that accommodating G×E in genomic models was beneficial to improve the quality of prediction for specific adaptation as well as for broad adaptation to both types of environments. This Chapter also tested if better genomic predictions can be obtained by accounting for heterogeneous variances of marker effects. We found that weighting individual markers based on estimated locus-specific variances produced important improvements in predictive performance of genomic models, even for a largely polygenic trait such as grain yield.

To conclude, this thesis deals with different challenging aspects that may affect genetic evaluation in modern sorghum breeding. Specifically, several statistical modeling strategies making use of different sources of information have been proposed and assessed. The findings presented in this thesis are expected to contribute to increase the efficiency of selection schemes not only for sorghum but for crops in general.

Acknowledgements

First of all, I would like to thank my promotor Fred van Eeuwijk for giving me the opportunity to do my PhD with him and his research group. Thank you Fred for your constant support and encouragement, and for being open and committed to stimulate inspiring discussions throughout my research. I feel honored and proud of having published papers with my name listed next to yours.

I also want to thank my co-promotor Marcos Malosetti for his guidance and motivation. Marcos, you were very important for a smooth start of my PhD. Your positive and straightforward guidance was key to consolidate the direction of this thesis. In addition, I always enjoyed discussing our research topics in Uruguayan Spanish, including the “bo” and “tá”, which I could understand very well and made me feel closer to home in some way.

I acknowledge the contribution of David Jordan, who also deserves the credits of a supervisor. In spite of being both of us in opposite sides of the world, he was always there answering my emails, providing steady inputs and ideas. Thank you David for your generosity in sharing your knowledge and expertise on sorghum breeding.

I would like to thank all coauthors who contributed to the different chapters of this thesis. María Xosé Rodriguez-Álvarez and Martin Boer, who made important contributions to the second chapter, thank you for fruitful discussions that helped me to better understand the use of P-splines for spatial analysis. And thank you for developing together with Fred and Paul Eilers such a useful software package, SpATS. I also thank Emma Mace and Colleen Hunt for their contribution preparing the data and their patience when answering all the questions I had about the datasets.

I am especially grateful for the financial support received from the Argentine Government through INTA, which made possible this research. I am deeply indebted to my colleagues of INTA, Emilce, Catalina and Mariela for taking care of my duties during my PhD.

Many thanks to all my colleagues of Biometris for creating a great working environment. Especially to my former and current fellow PhDs for all the good and fun moments we shared also outside work. You certainly helped in making these PhD years a great experience. Special thanks go to Guus and Daniela not only for being my paranymphs but also for your friendship. Guus, I will miss our running sessions where we could elaborate on a variety of topics, from philosophy and sociology to just food—when getting hungry. Daniela, I am deeply thankful to you for you kindness and constant support, since the first day until the end of my stay in the Netherlands. I feel lucky to have gotten to know you and your husband Marcelo. And I cannot forget to mention your sister Mariana, who saved my life one dark, cold, snowy night.

I also had the chance to build personal relationships with wonderful people outside Biometris, thanks to all of you for being part of my life during this period.

Last but not least, I dearly thank my friends and my family back in Argentina for their unconditional support and for carrying the weight of my dreams without crushing under it. Immense gratitude to the mother of my children for all the love and care given to them, you are amazing. Finally, I dedicate this thesis to the memory of my parents, Rosa and Abel, and also to my beautiful kids, Fausti and Euge; you are true love and inspiration.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (6 ECTS)

- Statistical methods for genetic evaluation in plant breeding

Writing of project proposal (4.5 ECTS)

- Statistical modelling for the optimization of crop breeding schemes: a sorghum case study

Post-graduate courses (4.2 ECTS)

- Introduction to Bayesian statistic for quantitative geneticists; Technical University of Munich (2015)
- Introduction to R; Wageningen University & Research (2015)
- Genomic prediction and selection in plant breeding; Mediterranean Agronomic Institute of Zaragoza (2016)
- Construction of genetic maps/theory and practice; Wageningen University & Research (2016)

Deficiency, refresh, brush-up courses (0.9 ECTS)

- Modern statistics for the life sciences; WUR (2016)

Competence strengthening / skills courses (1.5 ECTS)

- Scientific writing; WUR (2016-2017)

PE&RC Annual meetings, seminars and the PE&RC weekend (1.5 ECTS)

- PE&RC First years weekend (2015)
  PE&RC Midterm weekend (2017)

Discussion groups / local seminars / other scientific meetings (7.5 ECTS)

- StatGen Colloquium (2015-2019)
- Multi-Parent Population (MPP) discussion group (2016)

International symposia, workshops and conferences (8.9 ECTS)

- XVI Eucarpia Meeting: section biometrics in plant breeding (2015)
- XXIXth International biometric conference (2018)
- Global conference sorghum in the 21st century (2018)
- XVIIth Meeting of the EUCARPIA section biometrics in plant breeding (2018)
- XXIVth EUCARPIA maize and sorghum conference (2019)

Lecturing / supervision of practicals / tutorials (3 ECTS)

- Modern statistics for the life sciences (2017)
- Advanced statistics (2018)

The research presented in this thesis is part of a PhD project funded by the National Institute of Agricultural Technology (INTA) of Argentina.

The data used in this thesis was generated with funding from the Grain Research and Development Corporation (GRDC) of Australia.

David Jordan and Alan Cruickshank are acknowledged for providing the photograph of a sorghum field in Australia used in the cover of this thesis.

