Genomic prediction for broad and specific adaptation in sorghum accommodating differential variances of SNP effects

Julio G. Velazco1,2 | David R. Jordan3 | Colleen H. Hunt3,4 | Emma S. Mace3,4 | Fred A. van Eeuwijk2

1 National Institute of Agricultural Technology (INTA), EEA Pergamino, Pergamino B2700WAA, Argentina
2 Biometris, Wageningen University and Research, Wageningen 6700AA, The Netherlands
3 Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Hermitage Research Facility, Warwick, QLD 4370, Australia
4 Department of Agriculture and Fisheries, Hermitage Research Facility, Warwick, QLD 4370, Australia

Abstract
This paper reports a first study exploring genomic prediction for adaptation of sorghum [Sorghum bicolor (L.) Moench] to drought-stress (D-ET) and nonstress (W-ET) environment types. The objective was to evaluate the impact of both modeling genotype × environment interaction (GxE) and accounting for heterogeneous variances of marker effects on genomic prediction of parental breeding values for grain yield within and across environment types (ETs). For this aim, different genetic covariance structures and different weights for individual markers were investigated in best linear unbiased prediction (BLUP)-based prediction models. The BLUP models used a kinship matrix combining pedigree and genomic information, termed K-BLUP. The dataset comprised testcross yield performances under D-ET and W-ET as well as pedigree and genomic data. In general, modeling GxE increased predictive ability and reduced empirical bias of genomic predictions for broad adaptation across both ETs vs. models that ignored GxE by fitting a main genetic effect only. Genomic predictions for specific adaptation to D-ET or W-ET were also improved by K-BLUP models that explicitly accommodated GxE and used data from both ETs relative to prediction models that used data from the targeted ET exclusively or models that used all the data but assumed no GxE. Allowing for heterogeneous marker variances...
through weighted K-BLUP produced clear increments (43–72%) in predictive ability of genomic prediction for grain yield in all adaptation scenarios. We conclude that G×E as well as locus-specific genetic variances should be accommodated in genomic prediction models to improve adaptability of sorghum to variable environmental conditions.

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 (Lesk, Rowhani, & Ramankutty, 2016; Wiltshire, Kay, Gornall, & Betts, 2013). More frequent drought events and erratic rainfall are expected to become major factors affecting crop yields in dryland farming systems (Daryanto, Wang, & Jacinthe, 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 (TPEs) where new varieties will be grown. Two main breeding 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 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., Atlin, Kleinjach, Singh, & Piepho, 2011; Cooper & Hammer, 1996; Piepho & Möhring, 2005). The same challenges are renewed from the genomic selection perspective (Heslot, Jannink, & Sorrells, 2015; Malosetti, Bustos-Korts, Boer, & van Eeuwijk, 2016). In this case, important questions are (a) how to stratify the TPE, (b) how to use this stratification to design the training sets, and (c) how to model of G×E for improved genomic predictions of 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, Cooper, Hammer, & Butler, 2000; Hammer et al., 2014). This fact has traditionally favored selection strategies focused on improving for broadly adapted genotypes (Borrell, Jordan, Mullet, Henzell, & Hammer, 2006; Cooper & Hammer, 1996). 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., Heslot, Akdemir, Sorrells, & Jannink, 2014; van Eeuwijk, Denis, & Kang, 1996). An alternative strategy is to rely on environmental information to define specific ETs and 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.

Postflowering water deficit is the most common and critical type of drought affecting sorghum yield in Australia and worldwide (Hammer et al., 2014; Jordan, Hunt, Cruickshank, Borrell, & Henzell, 2012). 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: postflowering D-ET or 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 nonstress 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, Basford, Cooper, Bull, & McLaren, 1996; Smith, Cullis, & Thompson, 2005). Statistical methods accounting for G×E have been updated to incorporate the increasing availability of genetic marker information. Mixed models accounting for marker × environment interaction were initially used in crops for identification of quantitative trait loci (QTL) with across-and within-environment effects (Boer et al., 2007; van Eeuwijk, Bink, Chenu, & Chapman, 2010). In the framework of genomic selection, a BLUP method using marker-based relationships, termed G-BLUP (VanRaden, 2008), has been applied to model G×E in wheat (Triticum aestivum L.) breeding (Burgueño, de los Campos, Weigel, & Crossa, 2012; Lopez-Cruz et al., 2015; Oakey et al., 2016). Methods for multi-environment genomic prediction have been also developed within semiparametric and Bayesian approaches (Cuevas et al., 2016; 2017).

The G-BLUP models are commonly used in plant breeding because they are simple to implement and produce competitive results for prediction of complex quantitative traits (Gianola, Weigel, Krämer, Stella, & Schön, 2014; Heslot, Yang, Sorrells, & Jannink, 2012). 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 allows for unequal variances of SNP effects. This can be attained by incorporating weights for single nucleotide polymorphism (SNP) effects in the genomic relationship matrix. The SNP weighting approach has been implemented for genomic prediction in the animal breeding context (Su, Christensen, Janss, & Lund, 2014; Wang, Misztal, Aguilar, Legarra, & Muir, 2012; Zhang et al., 2010; Zhang, Lourenco, Aguilar, Legarra, & Misztal, 2016). However, studies evaluating its potential for plant breeding applications are still lacking. Here, we use a SNP weighting method based on BLUP 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 modeling genotype × 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.

2 MATERIALS AND METHODS

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 7 yr. 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-to-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 postflowering 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: postflowering D-ET for trials where stay-green was expressed and W-ET for trials where no expression of stay-green was observed. In the present series of trials, nine were classified as D-ET and 15 as W-ET. The trials under D-ET conditions occurred in 5 of the 7 yr at seven different locations. The mean grain yield in D-ET trials was 24% lower than in W-ET trials, 4.5 vs 5.9 t ha⁻¹, respectively. Similar levels of yield reduction caused by drought have been reported for rice (Oryza sativa L.) and wheat in a meta-analysis study (Zhang et al., 2018). All the lines were tested under both ETs with different levels of replication because of 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 Diversity Arrays Technology (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 (McCormick et al., 2018; Paterson et al., 2009) to identify SNPs. After running standard quality filtering of SNP data for minor allele frequency (<2.5%), missing values (>20%), and imputation, genotypes on 4782 SNP evenly spaced 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.

### 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 (BLUEs) of testcross hybrid means from individual trials were combined for analysis using the following model:

\[
y_{ijkl} = \mu + L_{ij} + M_{kj} + LME_{ikj} + T(E)_{ikjl} \\
+ LT(E)_{iklj} + MT(E)_{iklj} + LMT(E)_{ikklj}
\]

(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, \( \mu \) is the general mean for the \( j \)th ET, \( L_{ij} \) is the effect of the \( i \)th line in the \( j \)th ET, \( M_{kj} \) is the effect 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, \( T(E)_{iklj} \) is the effect of the \( l \)th trial nested within the \( j \)th ET, \( LT(E)_{iklj} \) is the \( il \)th line \( \times \) trial interaction effect within the \( j \)th ET, \( MT(E)_{iklj} \) is the \( klj \)th tester \( \times \) trial interaction effect within the \( j \)th ET, and \( LMT(E)_{ikklj} \) is the \( ikkl \)th line \( \times \) tester \( \times \) trial interaction effect within the \( j \)th ET. All the effects were considered as 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 as fixed at this stage to avoid double shrinkage of line effects in the genomic prediction stage.

### 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 as follows:

\[
y = X\mu + Z_{g}g + e
\]

(2)

where \( y = (y_{D}', y_{W}')' \) is a vector of line BLUEs for yield under D-ET and W-ET from the combined analysis with model (1); \( X \) is an incidence matrix for ET-specific general means \( \mu = (\mu_{D}, \mu_{W})' \); \( g = (g_{D}', g_{W}')' \) is a multi-environment vector of total additive genetic effects or genomic estimated breeding values (GEBVs), with corresponding incidence matrix \( Z_{g} \) relating \( y \) to \( g \); and \( e \) is a vector of random residuals distributed as \( e \sim N(0, R) \), where \( R \) is a diagonal matrix with elements computed as in Smith, Cullis, and Gilmour (2001). 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 \( g \sim N(0, \Sigma_{E} \otimes K_{E}) \), where \( \Sigma_{E} \) is the genetic variance–covariance matrix between ETs, \( K_{E} \) is the environment-dependent kinship matrix among lines, and \( \otimes \) is the Kronecker product operator. The forms of \( \Sigma_{E} \) and the derivations of \( K_{E} \) are described below.

### 2.3.1 Structures for matrix \( \Sigma_{E} \)

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

- **Model G**: genetic effects were modeled using \( \Sigma_{E} \sigma_{g}^{2} \), where \( \Sigma_{E} \) is a matrix with all ones and \( \sigma_{g}^{2} \) 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 \( \times \) ET interactions, that is, assuming that the TPE is not subdivided into ETs.

- **Model GE**: using a uniform or compound symmetry structure for \( \Sigma_{E} \), 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 \( \times \) ET interaction effect.

- **Model GEH**: a more general formulation of model GE using an unstructured form for \( \Sigma_{E} \), which also assumes genetic covariance between ETs but, in
this case, allowing for heterogeneous genetic variances across ETs. This parameterization enables the most flexible modeling of genotype \times ET interaction.

Model ID: using a diagonal structure with heterogeneous variances for $\Sigma_E$, 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.

### 2.3.2 Computation of environment-dependent kinship matrices $K_E$

In this study, we used estimated variances of individual SNP effects for grain yield within D-ET, within W-ET, and across both ETs to build three different kinship matrices: $K_D$, $K_W$, and $K_B$, respectively. These relationship matrices are environment dependent since each one contains information derived from a particular adaptation environment. The kinship matrix $K_E$ (with $E = D, W, B$) combines pedigree and genomic information in the following form:

$$K_E = wA + (1 - w)G_E$$

where $A$ is the numerator relationship matrix among lines based on the full pedigree, $G_E$ 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 $A$. 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 $G_E$ used to construct $K_E$ was computed according to VanRaden (2008):

$$G_E = Z\Sigma_E Z\lambda$$

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, $Z$ is a matrix of centered genotypes for each line, and $D_E$ 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 $D_E$ were derived from (purely) genomic breeding values (GBVs) of all lines within D-ET ($\tilde{m}_D$), within W-ET ($\tilde{m}_W$), or across both ETs ($\tilde{m}_B$), where $\tilde{m}_B = 0.5\tilde{m}_D + 0.5\tilde{m}_W$. For this derivation, we applied the procedure proposed by Wang et al. (2012), which can be generally described (for $E = D, W, B$) as follows:

(a) Environment-dependent GBVs ($\tilde{m}_E$) were predicted with conventional G-BLUP by fitting an unweighted matrix, termed $G$, which was computed using an identity matrix $I$ instead of $D_E$ in (4);
(b) Predicted SNP effects ($\tilde{u}_E$) were derived from $\tilde{m}_E$ as: $\tilde{u}_E = \lambda IZG^{-1}\tilde{m}_E$;
(c) Individual variances of SNP effects ($\tilde{\sigma}_{u,i,E}^2$) were then estimated as: $\tilde{\sigma}_{u,i,E}^2 = \tilde{u}^2_{i,E}2p_i(1 - p_i)$;
(d) SNP variances were standardized and used to build $D_E$.

Accordingly, the obtained weighting matrices $D_D$, $D_W$, and $D_B$ were used as in (4) for computation of corresponding genomic matrices $G_D$, $G_W$, and $G_B$, and these were finally included in (3) to build $K_D$, $K_W$, and $K_B$, respectively.

According to Wang et al. (2012), weights can be recomputed 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, the individual SNP variances used as weights in the computation of $G_B$ are represented graphically in Figure 1.

To evaluate the impact of accounting for heterogeneous SNP variances, we also considered predictions based on the unweighted kinship matrix, denoted simply as $K$, which was formed using the unweighted $G$ 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 based on the first method of 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. 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. Additive-genetic correlations between ETs were computed for each model as the Pearson’s correlation coefficient using the estimated genetic variances and covariances. Additionally, the goodness of fit of the models was assessed using the Akaike information criterion (AIC).

### 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 (\( y \)) were used as predictands to validate GEBVs (\( \hat{g} \)). Validation of predictions was framed in terms of the different breeding objectives for adaptation, where the target genetic values were either \( y_D \) or \( y_W \) when selecting for specific adaptation and \( y_B \) when selecting for broad adaptation, with \( y_B = 0.5y_D + 0.5y_W \) being the overall mean across both ETs. Accordingly, the selection targets \( y_D, y_W, \) and \( y_B \) were predicted by the respective GEBVs \( \hat{g}_D, \hat{g}_W \) and \( \hat{g}_B \). The GEBVs in \( \hat{g}_B \) are given directly by model G, whereas they were obtained as \( \hat{g}_B = 0.5\hat{g}_D + 0.5\hat{g}_W \) for the other models. Note that \( y_D \) and \( y_B \) were computed in the form of a selection index for broad adaptation by combing line performances in each ET (Kelly, Smith, Eccleston, & Cullis, 2007; Piepho & Möhring, 2005). 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, while data from the remaining 20% of lines were retained as validation set. 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 basis of predictive ability and unbiasedness of predictions in the validation set. 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 among prediction models.

### 2.5 Data analysis software

All analyses were implemented within the R environment (R Core Team, 2019). Spatial analysis of individual trials was performed under the REML-based mixed model framework using the freely available R-package SpATS (Rodriguez-Alvarez, Boer, Eilers, & van Eeuwijk, 2018). Phenotypic and genomic multi-environment analyses were implemented with the mixed model package ASReml-R (Butler, Cullis, Gilmour, Gogel, & Thompson, 2017).

### 3 RESULTS

#### 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 kinship matrix used, the model allowing estimation of a separate genetic variance for each ET and genetic correlation between ETs (model GEH) gave better fits than models imposing equal genetic variances across ETs (models G and GE) or assuming genetic independence between ETs (model 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 \( \Sigma_E \), as reflected by marked decreases in 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.

#### 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 with model G, which ignores the stratification of the TPE for prediction of broadly adapted genotypes. These improvements were always significant (\( p < .05 \)) and slightly magnified when weighted kinship matrices were used, representing a mean relative increase of 10% in predictive ability. More important and highly
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

<table>
<thead>
<tr>
<th>Kinship matrix</th>
<th>Model</th>
<th>$h^2$ D-ET</th>
<th>$h^2$ W-ET</th>
<th>$r_g$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$</td>
<td>G</td>
<td>0.47</td>
<td>0.47</td>
<td>1.00</td>
<td>244</td>
</tr>
<tr>
<td></td>
<td>GE</td>
<td>0.50</td>
<td>0.50</td>
<td>0.94</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>GEH</td>
<td>0.63</td>
<td>0.41</td>
<td>0.97</td>
<td>221</td>
</tr>
<tr>
<td></td>
<td>ID</td>
<td>0.53</td>
<td>0.23</td>
<td>0.00</td>
<td>302</td>
</tr>
<tr>
<td>$K_B$</td>
<td>G</td>
<td>0.43</td>
<td>0.43</td>
<td>1.00</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>GE</td>
<td>0.48</td>
<td>0.48</td>
<td>0.90</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>GEH</td>
<td>0.60</td>
<td>0.36</td>
<td>0.95</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>ID</td>
<td>0.57</td>
<td>0.31</td>
<td>0.00</td>
<td>106</td>
</tr>
<tr>
<td>$K_D$</td>
<td>G</td>
<td>0.46</td>
<td>0.46</td>
<td>1.00</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>GE</td>
<td>0.51</td>
<td>0.51</td>
<td>0.90</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>GEH</td>
<td>0.63</td>
<td>0.38</td>
<td>0.96</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>ID</td>
<td>0.60</td>
<td>0.32</td>
<td>0.00</td>
<td>111</td>
</tr>
<tr>
<td>$K_W$</td>
<td>G</td>
<td>0.46</td>
<td>0.46</td>
<td>1.00</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>GE</td>
<td>0.51</td>
<td>0.51</td>
<td>0.91</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>GEH</td>
<td>0.63</td>
<td>0.38</td>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ID</td>
<td>0.61</td>
<td>0.32</td>
<td>0.00</td>
<td>108</td>
</tr>
</tbody>
</table>

* $K$, unweighted kinship matrix; $K_B$, $K_D$, $K_W$, kinship matrices weighted by estimated variances of individual SNP effects across both ETs, within drought ET, and within water-sufficient ET, respectively.

* $G$, fitting a main genetic effect across ETs; GE and GEH, fitting a genotype × ET effect with common and heterogeneous genetic variances across ETs, respectively; ID, fitting independent genetic effects across ETs.

* Expressed as differences relative to the best-fitting model (with AIC = 0).

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Kinship matrix</th>
<th>Predictive ability</th>
<th>Empirical bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
<td>$K_B$</td>
<td>$K_D$</td>
</tr>
<tr>
<td>G</td>
<td>0.342</td>
<td>0.537</td>
<td>0.547</td>
</tr>
<tr>
<td>GE</td>
<td>0.364</td>
<td>0.591</td>
<td>0.600</td>
</tr>
<tr>
<td>GEH</td>
<td>0.374</td>
<td>0.592</td>
<td>0.605</td>
</tr>
<tr>
<td>Mean</td>
<td>0.359</td>
<td>0.573</td>
<td>0.584</td>
</tr>
<tr>
<td>G</td>
<td>0.993</td>
<td>1.177</td>
<td>1.189</td>
</tr>
<tr>
<td>GE</td>
<td>0.735</td>
<td>0.931</td>
<td>0.928</td>
</tr>
<tr>
<td>GEH</td>
<td>0.816</td>
<td>1.052</td>
<td>1.043</td>
</tr>
<tr>
<td>Mean</td>
<td>0.848</td>
<td>1.054</td>
<td>1.053</td>
</tr>
</tbody>
</table>

* $G$, fitting a main genetic effect across ETs; GE and GEH, fitting a genotype × ET effect with common and heterogeneous genetic variances across ETs, respectively.

* $K$, unweighted kinship matrix; $K_B$, $K_D$, $K_W$, kinship matrices weighted by estimated variances of individual SNP effects across both ETs, within drought ET, and within water-sufficient ET, respectively.

Significant gains in predictive ability ($p < .001$) were generally achieved by weighting the kinship matrix for individual SNP effects, with an average increment of 62% across models of $\Sigma_E$. The K-BLUP models incorporating marker effect information from a specific ET—through $K_D$ or $K_W$—tended to produce marginally better results than models using across-ETs marker information—through $K_B$—although these differences were not significant.
Almost unbiased genomic predictions 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 (Table 2). Accounting for heterogeneous SNP variances with matrices $K_B$, $K_D$, or $K_W$ 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 $\Sigma_E$ were very small among weighted kinship matrices.

### 3.3 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 × ET interaction and produce ET-specific GEBVs, resulted in significant increments of predictive ability in D-ET ($p < .01$), 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 ($p < .01$) 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 vs. 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.

Model G gave less-biased GEBVs relative to other structures of $\Sigma_E$ 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.

When specific adaptation to W-ET was the target of prediction, the performances of different models for $\Sigma_E$ were statistically similar in terms of predictive ability if the unweighted $K$ was fitted (Table 4). Differences were significant only between models ID and GEH ($p < .05$) 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
predicted ability of genomic predictions was optimal with K-BLUP fitting $K_W$ 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 $\times$ 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 $K_D$ and model GEH giving the most unbiased and not inflated genomic predictions in W-ET.

## 4 | DISCUSSION

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

Previous genomic selection studies in sorghum have focused either on within-trial (Hunt, van Eeuwijk, Mace, Hayes, & Jordan, 2018) or on across-trial predictions (Velazco et al., 2019a; 2019b). This has been the common approach in other crops as well (e.g., Albrecht et al., 2014; Burgueño et al., 2012; 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 megaenvironments), 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 multienvironment GEBVs that are conditioned on the ETs. In the terms discussed by McLean, Sanders, and Stroup (1991), these predictors are then applicable to an intermediate inference space, which is broader than the narrow inference space of within-trial predictions. At the same time, our intermediate space GEBVs apply to a narrower inference space relative to the more 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 $G \times E$ in the context of GP, but none had yet reportedly compared broad vs specific adaptation strategies.

### 4.1 Impact of modeling genotype $\times$ environmental type 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

---

**Table 4** Predictive ability and empirical bias (regression coefficient) of genomic predictions for specific adaptation to water-sufficient environmental type (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

<table>
<thead>
<tr>
<th>Model(^a)</th>
<th>Kinship matrix(^b)</th>
<th>$K$</th>
<th>$K_B$</th>
<th>$K_D$</th>
<th>$K_W$</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0.309</td>
<td>0.422</td>
<td>0.419</td>
<td>0.430</td>
<td>0.395</td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>0.295</td>
<td>0.434</td>
<td>0.430</td>
<td>0.438</td>
<td>0.399</td>
<td></td>
</tr>
<tr>
<td>GEH</td>
<td>0.296</td>
<td>0.438</td>
<td>0.441</td>
<td>0.450</td>
<td>0.406</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>0.284</td>
<td>0.400</td>
<td>0.390</td>
<td>0.403</td>
<td>0.369</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.296</td>
<td>0.424</td>
<td>0.420</td>
<td>0.430</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Empirical bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0.930</td>
<td>0.960</td>
<td>0.945</td>
<td>0.966</td>
<td>0.950</td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>0.646</td>
<td>0.774</td>
<td>0.752</td>
<td>0.760</td>
<td>0.733</td>
<td></td>
</tr>
<tr>
<td>GEH</td>
<td>0.874</td>
<td>1.068</td>
<td>1.055</td>
<td>1.066</td>
<td>1.016</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>1.148</td>
<td>1.164</td>
<td>1.141</td>
<td>1.151</td>
<td>1.151</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.900</td>
<td>0.991</td>
<td>0.973</td>
<td>0.986</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

\(a\) G, fitting a main genetic effect across ETs; GE and GEH, fitting a genotype $\times$ ET effect with common and heterogeneous genetic variances across ETs, respectively.

\(b\) $K$, unweighted kinship matrix; $K_B$, $K_D$, $K_W$, kinship matrices weighted by estimated variances of individual SNP effects across both ETs, within drought ET, and within water-sufficient ET, respectively.
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 $\Sigma_E$ 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 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 variation 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 nonstress 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 relative 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 (Hordeum vulgare L.) 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; Sukumarman, Crossa, Jarquin, Lopes, & Reynolds, 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 overpredict GEBVs in a specific environment, with overprediction or inflation of GEBVs being particularly detrimental for genetic gain in the long term (Aguilar et al., 2010; Velazco et al., 2019a).

Our study showed that the best combinations of predictive ability and unbiasedness were generally achieved when genotype x 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 GxE 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 with 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 multienvironment genomic predictions. In addition, our results demonstrate that the benefits of using multienvironment models over genetic main effect models also depend on the difference in genetic information content between the target and the nontarget environments and not only on the genetic correlation between environments, as previously suggested by Atlin, Baker, McRae, and Lu (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 the line’s testcross performances under postflowering drought and nondrought 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 × 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 (Bustos-Korts et al., 2019; Chenu et al., 2011) as well as with managed drought screening (Cooper, Gho, Leafgren, Tang, & Messina, 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).

### 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 pedigreed 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 multitrait 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 (Su et al., 2014; Zhang et al., 2010). 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 predictive accuracy decreased after the first iteration for the most complex trait determined by 500 minor-effect QTL. 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 recomputations 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, Utz, & Schön, 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 QTL (Daetwyler, Pong-Wong, Villanueva, & Woolliams, 2010; Zhang 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 <0.6% of the genetic variation in yield (see Figure 1). 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 (<5,000 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 QTL 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 (Zea mays L.) and rice (Bustos-Korts, Malosetti, Chapman, Biddulph, & van Eeuwijk, 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 with the nonsignificant 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 QTL, but not for traits mainly controlled by many small-effect QTL. 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 & 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 research estimates all marker effects simultaneously, and it does not require preselection 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 QTL affecting other traits that have been associated with improved drought adaptation such as nodal root angle and stay-green (Borre et al., 2014b; Mace et al., 2012). Further research is required to explore the possibilities of combining these sources of information in a multitrait modeling 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 linkage disequilibrium or the interactions of QTL with the genetic background (Sandhu et al., 2018; Vadez et al., 2011). 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 the 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 using phenotypes of ungenotyped lines connected through pedigree (Zhang et al., 2016).

5 | CONCLUSION

Results indicate that genomic prediction for broad adaptation across postflowering D-ET and W-ET environments in sorghum can be enhanced by considering a subdivided TPE and addressing genotype × ET interaction explicitly with multienvironment models (GE or GEH). Specific adaptation to a targeted ET may be better predicted if genetic information from the nontargeted ET is borrowed through multienvironment genomic models when compared with 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 have been only slightly dependent on the environmental context in which SNP weights were derived because of the strong correlation of line performances under postflowering drought and nondrought conditions in sorghum.

**AUTHOR CONTRIBUTIONS**

J.G. Velazco, D.R. Jordan, and F.A. van Eeuwijk conceived and designed the research. C.H. Hunt and E.S. Mace curated the data. J.G. Velazco performed the statistical analysis and wrote the original draft of the paper. D.R. Jordan and F.A. van Eeuwijk reviewed and edited the manuscript. All authors read and approved the final version of the manuscript.

**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. J.G. Velazco acknowledges financial support from the National Institute of Agricultural Technology (INTA) of Argentina, Res. DN 1126/13.

**CONFLICT OF INTEREST**

The authors declare that there is no conflict of interest.

**ORCID**

Julio G. Velazco https://orcid.org/0000-0001-6061-9997

**REFERENCES**


mental covariables for QTL-by-environment interactions, with an example in maize. *Genetics*, 177, 1801–1813. https://doi.org/10.1534/genetics.107.071068


Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J. L., ... de los Campos, G. (2015). Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3: Genes, Genomes, Genetics, 5, 569–582. https://doi.org/10.1534/g3.114.016097


How to cite this article: Velazco JG, Jordan DR, Hunt CH, Mace ES, van Eeuwijk FA. Genomic prediction for broad and specific adaptation in sorghum accommodating differential variances of single nucleotide polymorphism effects. *Crop Science*. 2020;1–15. https://doi.org/10.1002/csc2.20221