ORIGINAL ARTICLE

Partitioning the forms of genotype-by-environment interaction in the reaction norm analysis of stability

Dominic L. Waters[1] · Julius H. J. van der Werf[1] · Hannah Robinson[2,3] · Lee T. Hickey[3] · Sam A. Clark[1]

Abstract

Key message **The reaction norm analysis of stability can be enhanced by partitioning the contribution of different types of G × E to the variation in slope.**

Abstract The slope of regression in a reaction norm model, where the performance of a genotype is regressed over an environmental covariable, is often used as a measure of stability of genotype performance. This method could be developed further by partitioning variation in the slope of regression into the two sources of genotype-by-environment interaction (G × E) which cause it: scale-type G × E (heterogeneity of variance) and rank-type G × E (heterogeneity of correlation). Because the two types of G × E have very different properties, separating their effect would enable a clearer understanding of stability. The aim of this paper was to demonstrate two methods which seek to achieve this in reaction norm models. Reaction norm models were fit to yield data from a multi-environment trial in Barley (*Hordeum vulgare*), with the adjusted mean yield from each environment used as the environmental covariable. Stability estimated from factor-analytic models, which can disentangle the two types of G × E and estimate stability based on rank-type G × E, was used for comparison. Adjusting the reaction norm slope to account for scale-type G × E using a genetic regression more than tripled the correlation with factor-analytic estimates of stability (0.24–0.26 to 0.80–0.85), indicating that it removed variation in the reaction norm slope that originated from scale-type G × E. A standardisation procedure had a more modest increase (055–0.59) but could be useful when curvilinear reaction norms are required. Analyses which use reaction norms to explore the stability of genotypes could gain additional insight into the mechanisms of stability by applying the methods outlined in this study.

Introduction

Genotype-by-environment interactions (G × E) occur when the performance of a genotype is dependent on the environment it exists in. The performance of some genotypes can be more sensitive to environmental factors, whilst others show greater stability across a range of environments. Genotypes that combine high overall performance with stability are appealing for their greater marketing potential and utility in mitigating the effects of climate change on crop production

✉ Dominic L. Waters
   dwaters9@myune.edu.au

[1] School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia

[2] InterGrain Pty Ltd, Perth, WA, Australia

[3] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia

(Powell et al. 2012). Therefore, there is a need to understand the models available to estimate genetic variation and select parent genotypes for stability in plant breeding programs.

Reaction norm (RN) models have been a popular choice for modelling G × E and the stability of genotypes across environments (Kraakman et al. 2004; Kusmec et al. 2017; Lacaze et al. 2009; Li et al. 2019; Ly et al. 2018; Sjoberg et al. 2020). In RN models, the performance of a genotype is regressed across an environmental covariable (EC), which describes the quality of the environment as a continuous value (Woltereck 1913). If the EC is centred with a mean of zero, the intercept represents performance in the average EC, whilst the slope of the regression captures the stability of the genotype. The regression can be either fixed or random; random regression is more common, as additive relationships can then be accounted for via pedigree or genomic data, which enables the estimation of additive genetic breeding values and genomic predictions of stability. The current paper focuses on modelling additive genetic random

regression RN models, although the concepts are relevant to all types of RN models.

There are two types of G×E which can contribute to variation in the slope of a RN model, and both have very different properties. Rank-type G×E occurs when the genetic correlation between environments is less than unity. A correlation significantly less than unity means that between environments, genotypes re-rank and the relative effect of the underlying genes changes (Falconer 1952). In contrast, scale-type G×E occurs when the genetic variance changes between environments. This "scales" the breeding values such that the differences between genotypes get larger or smaller, but it does not affect the genetic correlation (James 2009). Separating the two sources of variation in the slope of a reaction norm would enable a clearer picture of re-ranking between genotypes, and how the genetic architecture of a trait varies across the EC. If the effect of scale-type G×E was removed from the reaction norm slope, the resulting slope would capture how much the genotype re-ranks across the EC, or its "rank-stability". This could assist breeders to select genotypes which combine high overall performance with a stable ranking across a given EC, or alternatively, genotypes with adaptation to a specific EC level.

A method suggested by Falconer (1990) proposes that for a linear RN model with the intercept placed at the average EC, the square of the correlation between the intercept and slope (r^2) represents the proportion of variance in the slope that is due to scale-type G×E. As such, the proportion of variance in slope due to rank-type G×E can be estimated as ($1–r^2$). A genetic regression can also be used to derive breeding values for the reaction norm slope that is independent of the intercept variance (thus independent of scale-type G×E), as recently applied in the reaction norm analysis of body weight in sheep (Waters, Clark, et al. 2022).

A drawback of Falconer's method is that it is limited to a linear function, which might not adequately describe the response of a genotype across the EC in some situations. This could be addressed by directly standardising the breeding value of genotypes across the EC to a constant genetic variance, thereby removing variation in stability due to scale-type G×E. For a given genotype, the change in standardised breeding value could then be used as an estimate of stability due to rank-type G×E.

Factor-analytic (FA) models can also separate scale-type and rank-type G×E when analysing the stability of genotypes across environments (Smith and Cullis 2018). Whilst RN models employ a single user-defined environmental covariable, FA models generate an unsupervised set of c common factors that represent unobserved environmental covariables which explain changes in genetic effects across environments (Meyer 2009). The slopes of regression on the common factors can be used to derive breeding values for stability whilst accounting for G×E (Smith and Cullis 2018). A downside of using FA models is that they reveal unobserved environmental factors which may not always be interpretable. Therefore, the model is less useful than RN models when investigating a specific environmental factor, such as in designed experiments. However, the ability of FA models to separate scale-type and rank-type G×E could be leveraged to investigate scale-correction methods for RN models.

The aim of this study was to apply the two methods for separating variation due to scale and rank-type G×E when estimating breeding values for stability from random regression RN models. The dataset involves a MET from an Australian barley breeding program with large amounts of scale-type and rank-type G×E. The resulting breeding values for stability from the RN models were compared with those derived from FA models to test how well they accounted for scale-type G×E. We use genomic data to model the additive genetic relationships amongst genotypes.

Methods

Phenotypic data

The data were provided by InterGrain Pty Ltd (www.intergrain.com) and consisted of plot yields from 52 partially replicated trials distributed across 15 locations throughout Western Australia, New South Wales, Victoria, and South Australia. The trials took place in 2019 (10 trials) and 2020 (42 trials).

There were 3460 unique genotypes with yield phenotypes, although 2734 of the genotypes were only represented in one environment and did not have single nucleotide polymorphism (SNP) data. The remaining 726 genotypes were almost fully replicated across the 15 environments and had SNP data.

SNP data

The SNP data were collected using the Illumina Infinium Wheat Barley 40 K SNP array Version 1.0 and was imputed to a set of 410 499 SNPs, as described in Keeble-Gagnère et al. (2021). After imputation, SNPs with a minor allele frequency less than 0.05 (103 489 SNPs) and with more than 12.5% heterozygosity (202 SNPs) were removed, leaving a total of 306 808 SNPs for the 726 genotypes.

Analysis

Because there was a large number of genotypes without replication across environments, a two-stage analysis was used (Smith et al. 2001) rather than a usually preferable single stage. This allowed all the genotypes to be used for

modelling spatial variation within the trials in the first stage, whilst avoiding the inclusion of 2734 uninformative genotypes to model G×E across environments in the second stage. All the models were fit using restricted maximum likelihood (REML) in ASReml-R (Butler et al. 2018).

In the first stage, trials were analysed independently to obtain best linear unbiased estimates (BLUEs) of yield for each genotype, along with the corresponding weights. For each trial, the initial model consisted of an independent random genotype effect and a first-order separable autoregressive term for the error variance structure using the row and column dimensions of the trial. Based on this model, plots with mean yields deviating more than four standard deviations from their expectation were removed. Additional random 'row' and 'column' terms were fitted based on a log-likelihood ratio test with a threshold of $p = 0.05$. Once the final model was determined, the BLUEs and their corresponding weights were obtained by fitting genotype as a fixed effect, following the methodology of Smith et al., (2001).

In the second stage, the performance of genotypes across the environments was analysed with RN and FA models utilising the BLUEs of yield and corresponding weights obtained in the first-stage analysis. Only the 726 genotypes with replication across environments and SNP data were considered in this stage. Of the 726 genotypes, 74 were removed from the analysis, as they were heterozygous for more than 12.5% of their SNPs. This left 652 genotypes, of which 606 were replicated across all 15 environments, with an additional 45 genotypes replicated in 14 environments. The remaining genotype was replicated in only six environments but was included in the analysis since it would contribute information through its genomic relationship to the population.

Reaction norm models

The RN models require an environmental covariable (EC) that captures the environmental quality. Whilst it is increasingly common for ECs to consist of real environmental covariables, this was not available in our dataset. Instead, a joint-regression analysis was used (Eberhart and Russell 1966; Finlay and Wilkinson 1963; Perkins and Jinks 1968), where the adjusted mean yield of each environment was used as the EC. RN models based on JR analysis have been used widely to investigate the genetic basis of stability for yield and quality traits in cereal crops (Calderini and Slafer 1999; Kraakman et al. 2004; Li et al. 2019; Sandhu et al. 2021; Tan et al. 2020). Because the distribution of genotypes across environments was slightly unbalanced in our dataset, the direct mean of yield in each environment would be a slightly biased estimator of the EC. To account for the unequal use of genotypes across environments, a simple model which fit genotypes as an independent random effect and

environment as a fixed effect was used to obtain a BLUE of yield for each environment, which formed the EC. The EC values ranged from 1.45 to 4.84 tonnes per hectare (t/ha) and were standardised to values with a variance of one and a mean of zero for the RN analysis. The distribution of genotypes along the EC is given in Fig. S1.

The first RN model was a standard linear reaction norm (RN-L), which modelled the genetic effect of each genotype as a linear function of the EC. The second RN model used cubic polynomials (i.e. a nonlinear function) to model the genetic effect across the EC and was denoted RN-NL.

The linear RN-L in simple matrix form was:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a}_0 + \mathbf{Z}_2\mathbf{a}_1 + \mathbf{e} \qquad (1)$$

where $\mathbf{y}$ was a vector of BLUEs for each genotype, $\mathbf{X}$ was an incidence matrix linking records to the fixed effects $\mathbf{b}$, and $\mathbf{Z}_1$ and $\mathbf{Z}_2$ were design matrices linking records to the genomic estimated breeding values (GEBVs) for the intercept ($\mathbf{a}_0$) and slope ($\mathbf{a}_1$), with $\mathbf{Z}_1$ containing 1's on the diagonal and $\mathbf{Z}_2$ containing the EC value corresponding to each BLUE in $\mathbf{y}$ on the diagonal, and $\mathbf{e}$ was the residual variance.

The environment was fitted as the only fixed effect to account for differences in the mean yield between environments, and residual variance was estimated independently for each environment. The genetic variance of $\mathbf{a}_0$ and $\mathbf{a}_1$ was modelled according to:

$$\begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \end{bmatrix} \sim \mathrm{N}(0, \mathbf{G} \otimes \mathbf{K}), \text{ where } \mathbf{K} = \begin{bmatrix} \sigma^2_{a_0} & \sigma_{a_1 a_0} \\ \sigma_{a_0 a_1} & \sigma^2_{a_1} \end{bmatrix} \text{ and } \mathbf{G}$$

was a relationship matrix based on SNP data, constructed using the first van Raden method (VanRaden 2008) and $\otimes$ refers to a Kronecker product of matrices.

For RN-NL, a cubic polynomial was fitted to describe the genetic effects across the EC. The model was a simple extension of RN-L:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a}_0 + \mathbf{Z}_2\mathbf{a}_1 + \mathbf{Z}_3\mathbf{a}_2 + \mathbf{Z}_4\mathbf{a}_3 + \mathbf{e} \qquad (2)$$

where $\mathbf{Z}_3$ and $\mathbf{Z}_4$ were design matrices linking records to the GEBVs for quadratic ($\mathbf{a}_2$) and cubic ($\mathbf{a}_3$) components, with the diagonal of $\mathbf{Z}_3$ and $\mathbf{Z}_4$ containing the square and cube of the EC value corresponding to the BLUE in $\mathbf{y}$, respectively. The remaining components $\mathbf{y}$, $\mathbf{X}$, $\mathbf{b}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{a}_0$, $\mathbf{a}_1$ and $\mathbf{e}$ were the same as in (1). The genetic variance of $\mathbf{a}_0$, $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ were also modelled in the same way as in RN-L, where:

$$\begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} \sim \mathrm{N}(0, \mathbf{G} \otimes \mathbf{K}), \text{ where}$$

$$\mathbf{K} = \begin{bmatrix} \sigma^2_{a_0} & \sigma_{a_1 a_0} & \sigma_{a_2 a_0} & \sigma_{a_3 a_0} \\ \sigma_{a_0 a_1} & \sigma^2_{a_1} & \sigma_{a_2 a_1} & \sigma_{a_3 a_1} \\ \sigma_{a_0 a_2} & \sigma_{a_1 a_2} & \sigma^2_{a_2} & \sigma_{a_3 a_2} \\ \sigma_{a_0 a_3} & \sigma_{a_1 a_3} & \sigma_{a_2 a_3} & \sigma^2_{a_3} \end{bmatrix}.$$

r both RN models, the genetic (co)variance matrix (**A**) across the EC was obtaining using:

$$\mathbf{A} = \phi\mathbf{K}\phi' \qquad (3)$$

In RN-L, ϕ was a 15×2 matrix containing a vector of 1's in the first column and a vector of the EC values for each environment. In RN-NL, ϕ was a had an additional two columns (15×4) consisting of the EC values squared and cubed, respectively.

RN overall breeding value and stability estimates

In both RN-L and RN-NL, the GEBV for the intercept (a_0) was treated as the overall GEBV for each genotype. The slope GEBV ($\mathbf{a}_1$) in RN-L captured how much the overall GEBV of each genotype changed across the EC and was therefore used as a GEBV for stability. This is the standard model for deriving stability from linear RN models (e.g. de Souza et al. 2020; Kusmec et al. 2017; Lacaze et al. 2009), so it was treated as the control measure of stability. None of the GEBVs in RN-NL could be directly interpreted as stability. Stability could only be estimated in RN-NL using the second scale-correction method (outlined later).

The slope GEBV ($\mathbf{a}_1$) could include variation due to both scale-type and rank-type $G \times E$. The following two methods aim to remove the variation due to scale-type $G \times E$, so that the resulting GEBVs describe variation due to rank-type $G \times E$ across the EC.

The first method was derived from Falconer (1990). Here, the correlation between the intercept and slope is attributed entirely to scale-type $G \times E$, whilst the remaining variation that is independent of the correlation is attributed to rank-type $G \times E$. We can obtain GEBVs for the slope that are independent of the genetic correlation between the intercept and slope by using a genetic regression:

$$\mathbf{a}_1^* = \mathbf{a}_1 - \frac{\sigma_{a_1 a_0}}{\sigma_{a_0}^2}\mathbf{a}_0 \qquad (4)$$

where $\mathbf{a}_0$ and $\mathbf{a}_1$ are the GEBVs for the intercept and slope, respectively, $\sigma_{a_0}^2$ is the variance in intercept, and $\sigma_{a_1 a_0}$ is the covariance between the intercept and slope. The resulting GEBVs ($\mathbf{a}_1^*$) are the essentially the residuals of a genetic regression between the intercept and slope components. This method, referred to hereafter as the genetic regression, was only applied to RN-L, as it cannot be used for reaction norms with higher-order polynomials (i.e. RN-NL).

The second method also attempts to remove variation in the reaction norm slope arising from scale-type $G \times E$ whilst allowing the use of RN models with higher-order polynomials. Briefly, the GEBVs of genotypes along the EC are standardised to a constant genetic variance to remove scale-type $G \times E$. After standardisation, the change in GEBV

for each genotype across the EC is used to estimate stability. This method, hereafter referred to as the standardisation method, was applied independently to both RN-L and RN-NL.

In more detail, the GEBV for genotype i at EC value t in a RN model is a function of the GEBVs for each regression coefficient ($a_0 \dots a_p$) and was calculated as follows:

$$\mathrm{GEBV}_{it} = a_{0_i} + a_{1_i} \times t + \dots + a_{p_i} \times t^p$$

Likewise, the genetic variance at the different values of t can be calculated using $\mathbf{A} = \phi\mathbf{K}\phi'$, as described in Eq. 3. The GEBV for each genotype at each value of t can then be standardised (SGEBV) based on the genetic variance:

$$\mathrm{SGEBV}_{i,t} = \frac{\mathrm{GEBV}_{i,t}}{\sqrt{\mathrm{Va}_t}}$$

where Va_t is the genetic variance at EC value t. By performing this across the range of EC values in the data, the SGE-BVs for each genotype can be plotted across the EC. The resulting standardised reaction norm describes the breeding value of each genotype across the EC whilst being adjusted for changes in the scale of the genetic variance. The SGE-BVs only represent relative performance and do not carry units. Since the SGEBVs are not linear with respect to t, there is not a simple measure for stability. A solution is to use the change in SGEBV over a specified interval across the EC to estimate the stability (**S**):

$$S_i = \frac{\mathrm{SGEBV}_{i,t_x} - \mathrm{SGEBV}_{i,t_y}}{t_x - t_y} \qquad (5)$$

where t_x and t_y represent the maximum and minimum EC values in the specified interval. It is therefore possible to estimate the stability of each genotype for different segments of the EC. This could be used to explore whether different genotypes are more sensitive to different parts of the EC. The current study used an interval containing the entire range of the EC. Therefore, **S** was interpreted for each genotype as the change in GEBV relative to the population across the range of the EC and was calculated using both RN-L and RN-NL.

Factor-analytic models

Unlike RN models, FA models 'find' environmental variables, called common factors, that explain patterns of $G \times E$ across environments. The regression on these common factors can be used to estimate the stability of genotypes across environments. If the EC used in the RN models is a major driver of $G \times E$, we can expect estimates of stability and overall performance to be similar between these methods. Since FA models have special properties which allow the user to measure the amount of scale-type $G \times E$ in the regression coefficients for the environmental variables,

we used stability estimated from FA models as a reference to compare the effectiveness of the two RN correction methods to account for scale-type G×E.

The first FA model (FA-2) used two common factors to model genetic effects that were common across the environments, along with specific genetic and residual variances for each of the 15 environments. FA-2 was described in matrix notation as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_g}\left(\mathbf{\Lambda_a} \otimes \mathbf{I_n}\right)\mathbf{f_a} + \mathbf{Z_g}\mathbf{\delta_a} + \mathbf{e} \tag{6}$$

where $\mathbf{y}$, X and b were the same as in Eq. 1, $\mathbf{Z_g}$ was a design matrix linking records to the random effects, $\mathbf{\Lambda_a} = [\mathbf{\lambda_{a_1}}, \mathbf{\lambda_{a_2}}]$ was a $b \times 2$ matrix of estimated loadings for the b environments and first ($\mathbf{\lambda_{a_1}}$) and second ($\mathbf{\lambda_{a_2}}$) common factors, and $\mathbf{I_n}$ was an $n \times n$ identity matrix, where n was the number of genotypes, $\mathbf{f_a} = (\mathbf{f'_{a_1}}, \mathbf{f'_{a_2}})'$ was a $n \times 2$ vector of genotype regression coefficients for the first ($\mathbf{f_{a_1}}$) and second ($\mathbf{f_{a_2}}$) common factor, $\mathbf{\delta_a} = (\mathbf{\delta'_{a1}}, \mathbf{\delta'_{a2}}, \dots, \mathbf{\delta'_{ab}})$ was a $n \times b$ vector containing the specific genetic effects ($\mathbf{\delta_{aj}}$) for the b environments, and $\mathbf{e}$ was the residual variance. The residual variance was estimated independently for each environment.

The variance of the common factors and specific effects was modelled according to:

$$\begin{bmatrix} \mathbf{f_a} \\ \mathbf{\delta_a} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{I} \otimes \mathbf{G} & 0 \\ 0 & \mathbf{\psi_a} \otimes \mathbf{G} \end{bmatrix} \right)$$

where $\mathbf{I_{c_a}}$ was an identity matrix of order two, $\mathbf{G}$ was the genomic relationship matrix as calculated for the RN models, and $\mathbf{\psi_a}$ was a $n \times n$ diagonal matrix with elements ψ_{aj}, which were the specific genetic variances for individual environments.

The matrix $\mathbf{A}$, which describes the additive genetic variance and covariance between the b environments, was calculated following:

$$\mathbf{A} = \mathbf{\Lambda_a}\mathbf{\Lambda'_a} + \mathbf{\psi_a}$$

The second FA model (RR-2) was a reduced-rank implementation of FA-2. RR-2 was identical to FA-2, except that the specific genetic variances for each of the 15 environments (ψ_{aj}) were assumed to be zero. RR-2 was fitted because RN models also implicitly assume that the specific genetic effects are zero. Therefore, the results from RR-2 should account for the effect of estimating specific genetic effects when comparing the RN and FA models.

FA models with more than two common factors were not considered. This was because RN models essentially capture the performance of a genotype in two components: intercept (overall breeding value) and slope (change in breeding value over a unit change in EC). Similarly, the first common factor in a FA model usually represents the overall breeding value (comparable to the intercept), with the second common factor capturing stability (comparable to the slope). Going beyond two common factors would make it difficult to compare stability between FA and RN models and was not required to achieve the objective of the study, which was to test how well the RN model scale corrections accounted for scale-type G×E. For reference, the two common factors accounted for on average 72.8% of the additive genetic variance (see results section), whilst fitting a third common factor explained an additional 9.14% (results not shown).

## Overall breeding value and stability in FA models

Breeding values for overall performance and stability for each genotype were derived from FA-2 and RR-2 based on the methods outlined in Smith and Cullis (2018). This involved first rotating the loadings and regression coefficients for each common factor using singular value decomposition to make them orthogonal. After rotation, the loadings were analysed to ascertain how much scale or rank-type G×E contributed to the regression coefficients for each common factor.

The sign of the loadings represents contrasts between the environments (Smith and Cullis 2018). Since all genotypes have the same intercept for each common factor (0,0), there is no re-ranking of genotypes (or no contrast) across a common factor when the loadings all have the same sign. In this case, the regression coefficients will exclusively capture scale-type G×E.

When there is a mix of positive and negative loadings, the regression coefficients capture at least some variation due to rank-type G×E, because the genotypes pass through (0,0) and re-ranking occurs. The degree to which the regression coefficients represent rank-type G×E can be assessed by comparing the absolute mean of the positive and negative loadings (or average contrasts), respectively, for the common factor. When the average contrasts are very different (Fig. 1a), most of the variation in the regression coefficients will be due to scale-type G×E, because the variance of GEBVs for the common factor will be very different between the average contrasts. When the absolute mean of the positive and negative loadings is equal, the variance of GEBVs for the common factor at the average contrasts will also be equal (Fig. 1b), so the variation in regression coefficients will be mostly due to rank-type G×E.

According to Smith and Cullis (2018), the loadings for the first common factor are usually the same sign, whilst the remaining common factors tend to be a mix of positive and negative loadings. This experience was reflected in our analysis (see results). This meant that the breeding value for overall performance across environments (**OP**) could be calculated as a function of the first common factor, since it represented a genetic effect common to all environments with no re-ranking. Therefore, **OP** was

Fig. 1 Schematic example of the regression of four genotypes (coloured lines) across a common factor, 'x' in a FA model. The vertical black lines represent the variation in GEBVs for the common factor 'x' at the average of the positive and negative loadings (or average contrasts). In **a**, the average negative and positive loading are -0.15 and 0.7 units, respectively, and in **b**, -0.7 and 0.7 units respectively

(a) Mean positive and negative loadings of different magnitude

(b) Mean positive and negative loadings of same magnitude
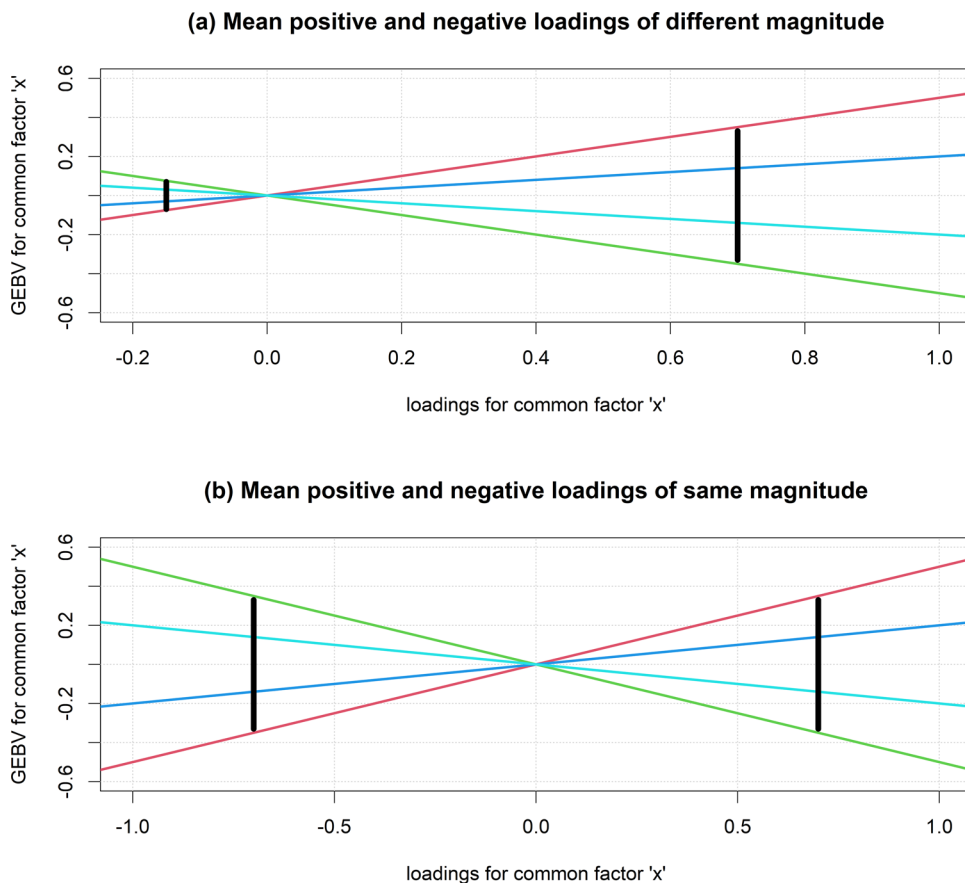
calculated as the mean additive genetic effect across the loadings for the first common factor:

$$\mathrm{OP}_i = \frac{1}{b} \sum_{j=1}^{b} \widetilde{\lambda}^*_{a_{1j}} \widetilde{f}^*_{a_{1i}}$$

where $\widetilde{\lambda}^*_{a_{1j}}$ was the rotated loading for the first common factor for environment j, and $\widetilde{f}^*_{a_{1i}}$ was the rotated regression coefficient for genotype i on the first common factor.

Because the second common factors in FA-2 and RR-2 contained a mix of positive and negative loadings that were similar in magnitude, they captured variation in stability due to rank-type G×E. Therefore, the responsiveness of genotypes to the second common factor (**R**2) was used as the GEBV for stability and was calculated as follows:

$$\mathrm{R}2_i = (\overline{\lambda}_{2+} - \overline{\lambda}_{a2-})\widetilde{f}^*_{a_{2i}}$$

where $\overline{\lambda}_{2+}$ and $\overline{\lambda}_{2-}$ represent the mean of the rotated positive and negative loadings for the second common factor, respectively. A smaller value of **R**2 indicated a genotype with greater stability in rank. The GEBVs for **R**2 were used as the reference estimate of stability to which the various

GEBVs for stability estimated from the RN models were compared.

The percentage of genetic variance in FA-2 that was explained by common factor d (either 1 or 2) in environment j in was calculated using:

$$v_{a_{dj}} = 100(\widetilde{\lambda}^*_{a_{dj}})^2 \Big/ \left(\sum_{s=1}^{2} (\widetilde{\lambda}^*_{a_{sj}})^2 + \psi_{a_j} \right)$$

where $\widetilde{\lambda}^*_{a_{dj}}$ was the rotated loading for common factor d in environment j, and ψ_{a_j} was the specific genetic effect in environment j. The average value of v_{a_d} across the j environments was used to calculate the mean genetic variance explained by the common factors.

Results

Model summary

The FA models provided a better fit to the data than the RN models based on the Akaike information criterion (AIC), Bayesian information criterion (BIC) and log likelihood (Table 1). The linear RN model (RN-L) was

Table 1 Summary of model fit. The number of parameters excludes those which fixed to zero during the model fitting process

[a]Model	Number of parameters	[b]AIC	[c]BIC	Log likelihood
RN-L	18	− 8076.89	− 7947.08	4056.445
RN-NL	25	− 8215.64	− 8035.35	4132.818
RR-2	44	− 8565.06	− 8247.76	4326.532
FA-2	58	− 9601.79	− 9183.52	4858.896

[a]*RN-L* linear reaction norm; *RN-NL* Nonlinear reaction norm; *RR-2* Reduced-rank factor-analytic model with two common factors and zero specific genetic effects; *FA-2* Factor-analytic model with two common factors and specific genetic effects

[b]Akaike information criterion

[c]Bayesian information criterion

Table 2 Pearson correlations between GEBVs for overall performance in the RN models ($\mathbf{a}_0$) and FA models (**OP**)

[a]Model	RN-L	RN-NL	FA-2	RR-2
RN-L	1	0.90	0.93	0.95
RN-NL		1	0.94	0.95
FA-2			1	0.99

[a]*RN-L* Linear reaction norm; *RN-NL* Nonlinear reaction norm; *FA-2* Factor-analytic model with two common factors and specific genetic effects; *RR-2* Reduced-rank factor-analytic model with two common factors and zero specific genetic effects

a significantly poorer fit that the cubic RN model (RN-NL) (ΔAIC = 139), whilst the reduced-rank factor-analytic model (RR-2) provided a poorer fit than the factor-analytic model (FA-2) (ΔAIC = 1037).

Overall, the first and second common factors in FA-2 explained 50.6 and 22.4% of the genetic variance across the environments, respectively (72.8% in total, Table S1). The percent of genetic variance explained by each common factor within the environments is also reported in Table S1. The genetic (co)variance of the random regression coefficients for RN-L and RN-NL are given in Table S2, along with correlations. The genetic correlation between the intercept and slope (*r*) in RN-L was high (0.76), indicating that genotypes with larger intercepts also tended to have larger slopes. Based on the r^2, this also implied that 57.8% of the variation in slope was due to scale-type G×E, whilst 42.2% was due to rank-type G×E.

Despite the differences in model fit, the correlations between GEBVs for overall performance in the RN models ($\mathbf{a}_0$) and FA models (**OP**) were high, ranging between 0.93 and 0.95 (Table 2). The lowest correlation between GEBVs for overall performance was between the RN models (0.90). Although not shown, the correlation between GEBVs for stability in FA-2 and RR-2 (**R2**) was also very high (0.96).
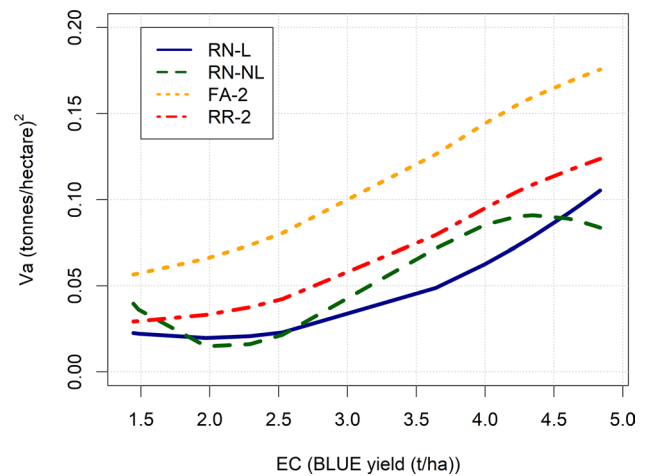

Fig. 2 Additive genetic variance (Va) across the environmental covariable (EC) for the four models. In the RN-L and RN-NL, the line connects the estimates of Va for each environment, whilst a smoothing spline was fitted to the estimates of Va across the EC in RR-2 and FA-2. RN-L: linear reaction norm; RN-NL: nonlinear reaction norm; RR-2: reduced-rank factor-analytic model with two common factors and zero specific genetic effects: FA-2: factor-analytic model with two common factors and specific genetic effects

Table 3 Minimum (min), maximum (max) and mean pairwise correlations between the 15 environments for the four models. The mean correlation was calculated using Fisher's Z-transformation

[a]Model	RN-L	RN-NL	FA-2	RR-2
Min	0.08	0.02	− 0.43	− 0.51
Max	1.00	1.00	0.97	1.00
Mean	0.96	0.89	0.50	0.86

[a]*RN-L* linear reaction norm; *RN-NL* Nonlinear reaction norm; *FA-2* Factor-analytic model with two common factors and specific genetic effects; *RR-2* Reduced-rank factor-analytic model with two common factors and zero specific genetic effects

Magnitude of G×E

Genetic variance increased along the EC in RN-L and RN-NL by 430% and 486%, respectively, indicating strong scale-type G×E (Fig. 2). A smoothing spline fitted to the genetic variance of FA-2 and RR-2 also revealed similar increases in genetic variance across the EC. Overall, FA-2 estimated approximately 30–45% more genetic variance across the EC compared to RR-2, which was due to modelling specific genetic effects for each environment. The RN models estimated slightly less genetic variance across the EC compared to RR-2.

Genetic correlations between low and high yielding environments in the RN models were close to zero, indicating substantial levels of rank-type G×E (Table 3). The FA models tended to estimate lower genetic correlations between the environments.

Loadings of factor-analytic models

The average of the negative loadings for the first common factor was 92.0 and 90.4% smaller than the average of the positive loadings in FA-2 and RR-2, respectively (Table 4). Additionally, only one loading in RR-2 and two loadings in FA-2 were negative for the first common factor (Table S3). Therefore, the first common factor captured variation that was almost entirely due to scale-type $G \times E$ in both models. The average of the negative loadings for the second common factor were only 34.1 and 1.9% smaller than the average of the positive loadings in FA-2 and RR-2, respectively. Because the magnitude of the average positive and negative loadings was similar (especially in RR-2), the second common factor captured variation primarily due to rank-type $G \times E$. Therefore, we can expect the regression on the second common factor, $\mathbf{R}2$, to capture variation due to rank-type $G \times E$.

Comparison of stability measures

The GEBVs for slope in RN-L ($\mathbf{a}_1$) were lowly correlated with $\mathbf{R}2$ estimated in both FA-2 (-0.24) and RR-2 (-0.26) (Fig. 3a and b). Therefore, the standard method for determining stability in RN models gave very different results to the FA models. However, applying a genetic regression to the GEBVs (Eq. 4) for slope ($\mathbf{a}_1^*$) more than tripled the correlation with $\mathbf{R}2$, to -0.80 in FA-2 and -0.85 in RR-2 (Fig. 3c and d). Estimating stability after standardising the genetic variance in RN-L ($\mathbf{S}$) also increased the correlation with $\mathbf{R}2$ relative to $\mathbf{a}_1$, although the increase was more modest than $\mathbf{a}_1^*$ (Fig. 3e and f). Using the same standardisation method with RN-NL ($\mathbf{S}$), which used cubic polynomials, did not increase the correlation with $\mathbf{R}2$ compared to RN-L (Fig. 3g and h). Note that the sign of the correlations is arbitrary, as the loadings and regressions in FA-2 and RR-2 can be multiplied by $+1$ or $-$ whilst retaining their meaning (Smith and Cullis 2018). They will be reported as absolute values herein.

Table 4 Absolute value of the mean REML estimates of the positive and negative loadings for the two common factors in FA-2 and RR-2

[a]Model	Common factor	Mean positive loading	Mean negative loading
FA-2	1	0.273	0.022
	2	0.126	0.083
RR-2	1	0.231	0.022
	2	0.105	0.103

[a]*RN-L* Linear reaction norm; *RN-NL* Nonlinear reaction norm; *FA-2* Factor-analytic model with two common factors and specific genetic effects; *RR-2* Reduced-rank factor-analytic model with two common factors and zero specific genetic effects

Relationship between overall performance and stability in different models

In RN-L, the genetic correlation between the intercept and slope (0.76, Table S2) and the correlation between GEBVs for the intercept ($\mathbf{a}_0$) and slope ($\mathbf{a}_1$) (0.73, Fig. 4a) was high. This indicated that genotypes with a larger intercept tended to have larger change in GEBV across the EC and were therefore less stable. When the reaction norm slope was adjusted for scale-type $G \times E$ using the genetic regression ($\mathbf{a}_1^*$), the correlation with overall performance was close to zero (-0.08) (Fig. 4b). This was also the case in FA-2 and RR-2 (Fig. 4c and d). Although this was expected since $\mathbf{a}_1^*$ and $\mathbf{R}2$ were both estimated orthogonally to the overall performance, the correlation between $\mathbf{S}$ and $\mathbf{a}_0$ was also much smaller than 0.73 and made no assumption about the relationship between the intercept and slope (RN-L: 0.40, Fig. 4e).

The correlation between $\mathbf{S}$ and $\mathbf{a}_0$ in the cubic RN-NL was comparably high (0.71, Fig. 4f). However, the individual standardised reaction norms in this model were very noisy (see next section). This could cause the estimates of $\mathbf{S}$ (and therefore the correlation with $\mathbf{a}_0$) to be very sensitive to small changes in the interval which it was evaluated over (t_x and t_y), making interpretation of the correlation difficult.

Reaction norm scale-corrections methods

To explore reaction norm scale-corrections methods in more detail, the reaction norms for RN-L and RN-NL were plotted before and after applying the two correction methods (Fig. 5). The two genotypes with very poor performance at low EC levels (coloured blue and black, Fig. 5a, c) were 'check' varieties with lots of yield records and were not considered outliers in the analysis. Before the applying the correction methods, there was greater variation in GEBVs at higher mean yields and individual reaction norms crossed over, demonstrating both scale-type and rank-type $G \times E$, respectively (Fig. 5a, c and f). When the slope of reaction norm was given by the genetic regression ($\mathbf{a}_1^*$), the variation in GEBVs was more constant across the EC, although there was slightly more variation in the low-yielding environments (Fig. 5b). This was likely because the mean EC (3.56 t/ha), which was the intercept, was slightly larger than the midpoint of the EC range (3.14 t/ha). The slope of regression changed direction for some genotypes.

When the GEBVs were standardised by the genetic variance, the reaction norms for RN-L were no longer linear, whilst the variation of GEBVs was also constant across the EC in both models (Fig. 5d and g). This can be seen more clearly in Fig. S2, which shows the variance of the GEBVs across the EC before and after standardisation. A linear approximation of the standardised reaction

Fig. 3 Scatterplots between stability GEBVs from RN models ($\mathbf{a}_1$, $\mathbf{a}_1^*$ and $\mathbf{S}$) compared with GEBVs for responsiveness to the second common factor (**R2**) in FA-2 and RR-2. The correlations are printed in each frame. RN-L: linear reaction norm; RN-NL: nonlinear reaction norm; FA-2: factor-analytic model with two common factors and specific genetic effects; RR-2: reduced-rank factor-analytic model with two common factors and zero specific genetic effects
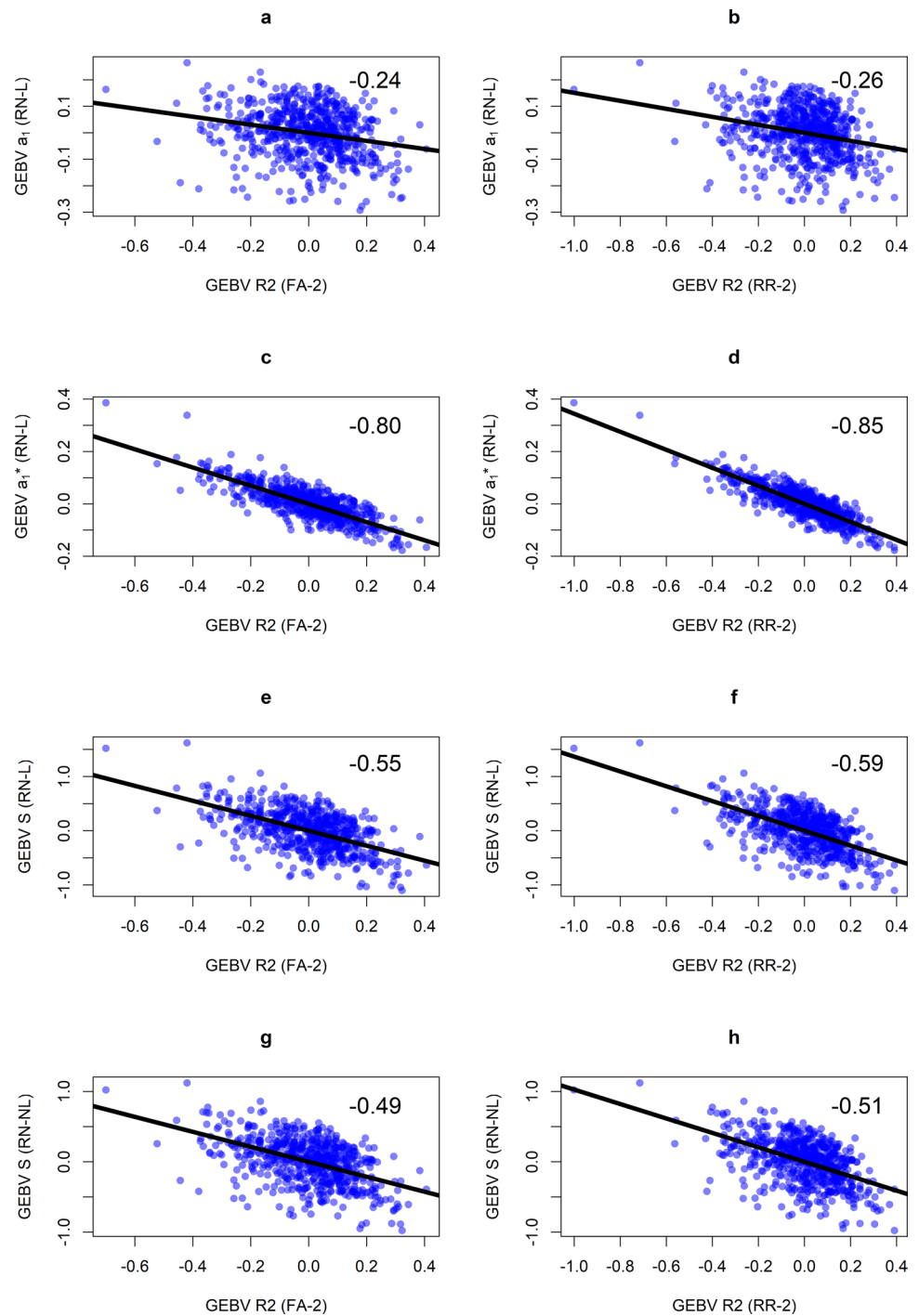
norms, which used the SGEBV when EC = 0 as the intercept and **S** as the slope, revealed that scale-type G × E was greatly reduced for RN-L (Fig. 5e) but to a lesser extent for RN-NL (Fig. 5h). The individual standardised reaction norms in RN-NL tended to be more erratic with very steep curves. This could indicate that the model was attempting to capture more complex G × E effects, rather than the general pattern across environments.

Discussion

The MET dataset contained large differences in the genetic variance (scale-type G × E) and low genetic correlations (rank-type G × E) between environments across the EC, which is expected in populations where G × E is significant. This level and type of G × E is typical of METs

Fig. 4 Relationships between overall performance (a_0 and **OP**) and stability (a_1, a_1^*, **S** and **R**2) within the RN and FA models. The correlations are printed in each frame. RN-L: linear reaction norm; RN-NL: nonlinear reaction norm; FA-2: factor-analytic model with two common factors and specific genetic effects; RR-2: reduced-rank factor-analytic model with two common factors and zero specific genetic effects
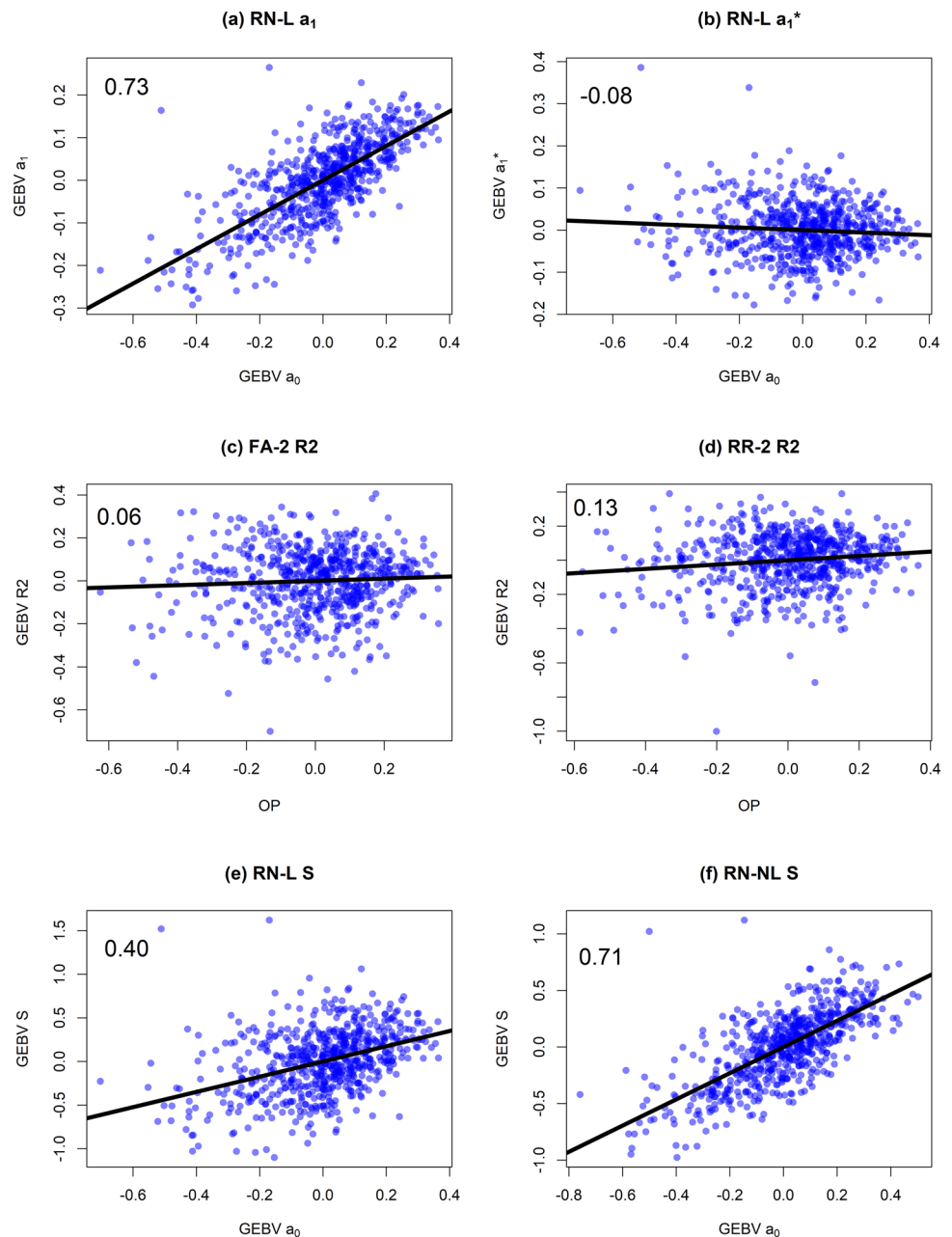
collected in Australian plant breeding programs (Smith et al. 2021). Hence, the data were a good resource for testing the scale-correction methods in this study.

The standard estimate of stability in a RN model is the breeding value for the slope. The slope GEBVs $\left(a_1\right)$ were very lowly correlated (0.24 to 0.26) with stability GEBVs estimated in the FA models (**R**2) (Fig. 3). This is interesting, as the RN and FA models would produce vastly different rankings of genotypes based on stability. Likewise, the SNP effects that underly the GEBVs would also be very different, which could lead to conflicting accounts of the genetic architecture of stability. However, correcting the slope GEBVs using a genetic regression (a_1^*) tripled the correlation with

R2 (0.80 to 0.85). Because **R**2 captured mainly rank-type G×E (Table 4), we can conclude that 1) a_1 was lowly correlated with **R**2 because it contained variation due to scale-type G×E and 2) a_1^* was highly correlated with **R**2 because the variation due scale-type G×E in a_1 was successfully removed by the genetic regression.

The GEBVs estimated using the standardisation method (**S**) were also more highly correlated with **R**2 compared to a_1, although the increase was less than with a_1^*. Therefore, it appears that the genetic regression is a more suitable approach for accounting for scale-type G×E in linear reaction norms. The correlation with **R**2 was higher when **S** was estimated using the linear RN model (RN-NL) compared to

the cubic RN model (RN-L). However, we would expect the linear reaction norm to be more closely related to $R2$ than a cubic reaction norm, because $R2$ is also a linear regression. Therefore, it was difficult to evaluate the merit of estimating S in either RN-L or RN-NL based on these results. Nevertheless, the standardisation method should still be useful to account for scale-type $G \times E$ in situations which require curvilinear reaction norms.

The intercept and slope GEBVs were highly correlated in RN-L (Fig. 4a, Table S2), indicating that genotypes with a larger intercept tended change more in GEBV across the EC and were therefore less stable. However, the correlation was close to zero when scale-type $G \times E$ was accounted for using a genetic regression and factor-analytic models (Fig. 4b, c and d), indicating there was no relationship between overall performance and stability due to rank-type $G \times E$. This is similar to other analyses of stability which account for scale-type $G \times E$ (Smith and Cullis 2018; Tolhurst et al. 2019). This highlights the importance of understanding how stability is defined when interpreting its relationship with overall performance in different studies.

The additional benefit gained by applying the scale corrections is that the remaining slope variation should represent rank-type $G \times E$, which arises when the genetic correlation between environments is less than unity. This could be used to explore how the genetic architecture of a trait varies across an EC. This phenomenon was highlighted in a reaction norm study in Australian sheep (Waters et al. 2022), where a seemingly pleiotropic quantitative trait loci affected both the intercept and slope but disappeared when a genetic regression was used. Instead, a new region which contained a group of genes previously associated with environmental sensitivity in livestock was identified. It would be interesting to see if similar reports of pleiotropic genes affecting overall performance and stability in plants using reaction norms (e.g. Li et al. 2019) also change if additional analysis to account for scale-type $G \times E$ is used.

There are exciting prospects for the improvement of yield stability in breeding programs using the genomic models in this study. Genomic breeding values derived from a training population such as routine METs could be used in combination with a speed-breeding program to fast track the creation of stable genotypes (Watson et al. 2018). In MET data, stability GEBVs derived from factor-analytic models would be recommended over the reaction norm models due to their superior ability for capturing the multifactorial causes of $G \times E$ in these settings (Kelly et al. 2007). This was realised in our analysis based on the AIC, BIC and the log likelihood. Additionally, FA-2 estimated between 30 and 45% more genetic variance within environments compared to the remaining models, indicating that it could disentangle $G \times E$ more efficiently. Despite this, the GEBVs from both models were highly correlated for overall performance

(0.93–0.95) and stability (0.80 to 0.85) after scale-type G x E was accounted for. This is important, considering that the linear reaction norm required approximately two minutes to converge, compared to approximately 1 h 10 min (RR-2) and 6 h 4 min (FA-2) for the factor-analytic models. As the number of genotypes and environments in routine analysis grow, the value of using genomic reaction norms for their speed also could increase. Examples of reaction norm models being used to efficiently fit dense genomic analyses with hundreds of thousands of genotypes can be found in livestock and human studies (Carvalheiro et al. 2019; Lee and van der Werf 2016; Ni et al. 2019).

It is increasingly common for METs to collect environmental information to understand the causal factors underlying $G \times E$ (Cooper et al. 2021). A common way to incorporate the large number of resulting ECs is through (co) variance structures which describe the relationships between the environments based on the ECs (Jarquín et al. 2014; Raffo et al. 2022). Because these models implicitly regress phenotypes on ECs, the have also been interpreted as reaction norm models. Whilst these models have demonstrated increased predictive ability over models which lack ECs (Cuevas et al. 2017; Lopez-Cruz et al. 2015), it is unclear whether they can reveal the stability of genotypes across the ECs due to the implicit nature of their fit. If methods to extract the regressions of genotypes across the ECs become available, the impact scale-type $G \times E$ should be explored when estimating stability.

Another promising example for modelling this type of data integrates measured ECs into factor-analytic models (Tolhurst et al. 2022). In this study, the ECs explained mostly rank-type $G \times E$, whilst the unobserved common factors explained scale-type $G \times E$. Therefore, genotypes could be ranked based on their stability to the important ECs in the model. These models also outperformed equivalent reaction norms, which fit implicit random regressions on the individual environmental factors. Despite this, reaction norms should be the most efficient method for analysing stability when the EC is known and controlled (e.g. Houshmandfar et al. 2019; Paschke et al. 2003; Sadras et al. 2012), or the main source of $G \times E$ for the trait (Li et al. 2018; Millet et al. 2019).

Conclusion

This study investigated the ability of two methods to partition variation in the slope of a reaction norm model into the two types of $G \times E$ (rank-type and scale-type) which underly it. Stability estimated from factor-analytic models, which explicitly disentangle rank-type and scale-type $G \times E$, was used for comparison. The two methods substantially increased the correlation with stability estimated from the
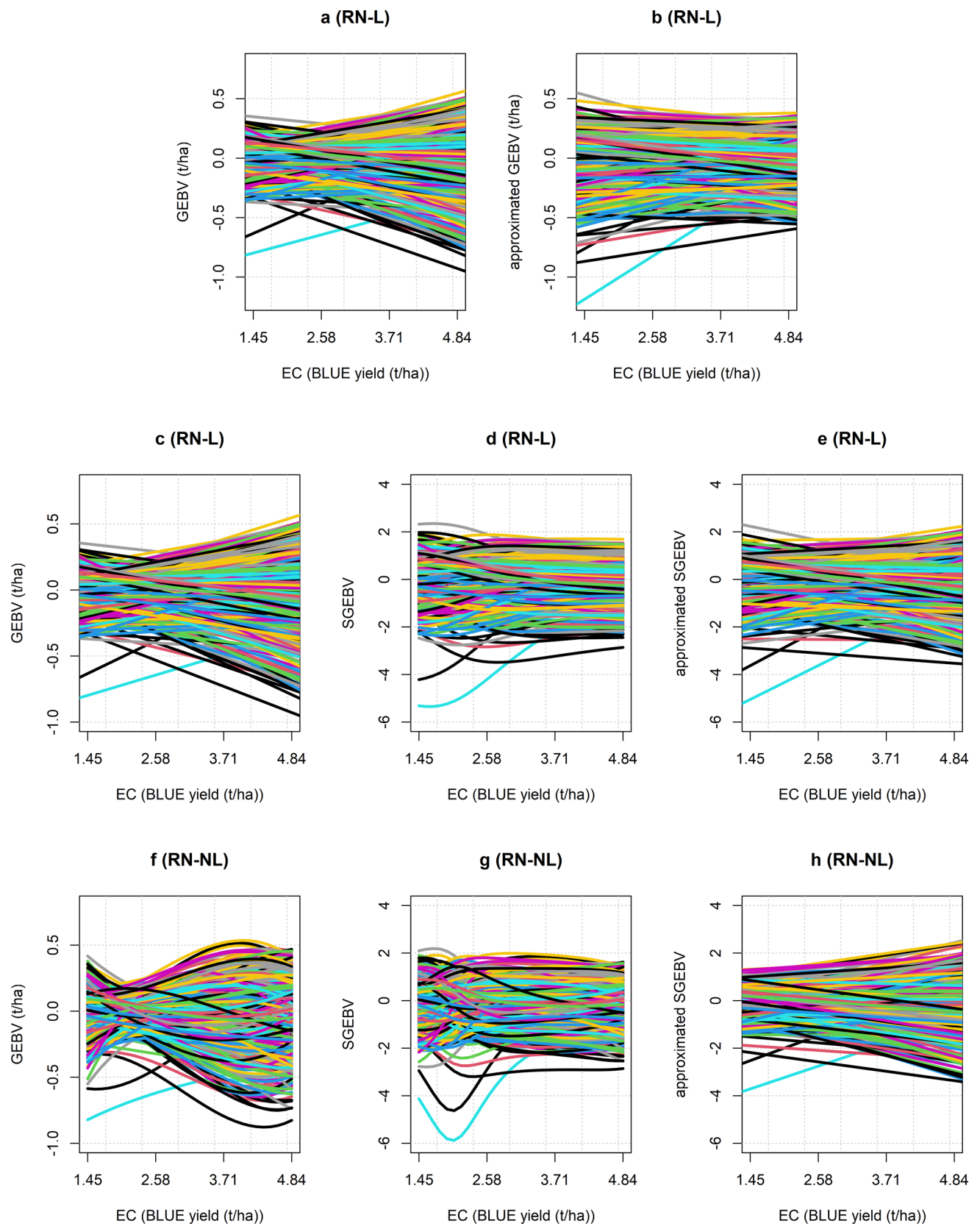
◄**Fig. 5** Reaction norms before and after applying the scale corrections. The top row shows the RN-L before **a** and after **b** applying the genetic regression, where the slope of each genotype is given by a_1 and a_1^*, respectively. The middle row shows RN-L before **c** and after **d** standardisation, and a linear approximation of the standardised reaction norm **e** where the slope of each genotype is given by their average change in SGEBV across the entire EC. The bottom row shows RN-NL before (**f**) and after (**g**) standardisation, and a linear approximation of the standardised reaction norm (**h**) where the slope of each genotype is given by their average change in SGEBV across the entire EC. RN-L: linear reaction norm; RN-NL: nonlinear reaction norm

factor-analytic models, indicating that they removed variation in the reaction norm slope that originated from scale-type G×E. The genetic regression method appeared to be the most effective, yielding estimates of stability that were highly correlated (0.80—0.85) to those estimated from the factor-analytic models. Although the standardisation method had a more modest increase, it could still be useful in situations where curvilinear RN models are required. Analyses which use reaction norms should consider implementing the scale corrections outlined in this study to gain additional understanding of the nature of G×E and stability in their population.

Declarations

References

Butler DG, Cullis BR, Gilmour AR, Gogel BJ, Thompson R (2018) ASReml-R Reference Manual Version 4. *ASReml-R Reference Manual*.

Calderini DF, Slafer GA (1999) Has yield stability changed with genetic improvement of wheat yield? Euphytica 107(1):51–59. https://doi.org/10.1023/A:1003579715714

Carvalheiro R, Costilla R, Neves HHR, Albuquerque LG, Moore S, Hayes BJ (2019) Unraveling genetic sensitivity of beef cattle to environmental variation under tropical conditions. Genet Sel Evol. https://doi.org/10.1186/s12711-019-0470-x

Cooper M, Voss-Fels KP, Messina CD, Tang T, Hammer GL (2021) Tackling G × E × M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. Theor Appl Genet 134(6):1625–1644. https://doi.org/10.1007/s00122-021-03812-3

Cuevas J, Crossa J, Montesinos-López OA, Burgueño J, Pérez-Rodríguez P, De Los Campos G (2017) Bayesian genomic prediction with genotype × environment interaction kernel models. G3 Genes, Genom, Genet. https://doi.org/10.1534/g3.116.035584

De Souza MH, Pereira Júnior JD, Steckling SD, Mencalha J, Dias FD, Rocha JR, Carneiro PC, Carneiro JE (2020) Adaptability and stability analyses of plants using random regression models. PLoS ONE 15(12):e0233200. https://doi.org/10.1371/journal.pone.0233200

Eberhart SA, Russell WA (1966) Stability parameters for comparing varieties. Crop Sci 6(1):36–10. https://doi.org/10.2135/cropsci1966.0011183x000600010011x

Falconer DS (1952) The problem of environment and selection. Am Naturalist 86(830):293–298. https://doi.org/10.1086/281736

Falconer DS (1990) Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. Genet Res 56(1):57–70. https://doi.org/10.1017/S0016672300028883

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14(6):742–754. https://doi.org/10.1071/AR9630742

Houshmandfar A, Rebetzke GJ, Lawes R, Tausz M (2019) Grain yield responsiveness to water supply in near-isogenic reduced-tillering wheat lines – An engineered crop trait near its upper limit. Eur J Agronomy. https://doi.org/10.1016/j.eja.2018.11.003

James JW (2009) Genotype by environment interaction in farm animals. Adapt Fitness Anim Popul. https://doi.org/10.1007/978-1-4020-9005-9_10

Jarquín D, Crossa J, Lacaze X, du Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Pérez P, Calus M, Burgueño J, De Los Campos G (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127(3):595–607. https://doi.org/10.1007/s00122-013-2243-1

Keeble-Gagnère G, Pasam R, Forrest KL, Wong D, Robinson H, Godoy J, Rattey A, Moody D, Mullan D, Walmsley T, Daetwyler HD, Tibbits J, Hayden MJ (2021) Novel design of imputation-enabled SNP arrays for breeding and research applications supporting multi-species hybridization. Front Plant Sci. https://doi.org/10.3389/fpls.2021.756877

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(3):1063–1070. https://doi.org/10.2135/cropsci2006.08.0540

Kraakman ATW, Niks RE, van den Berg PMMM, Stam P, van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168(1):435–446. https://doi.org/10.1534/genetics.104.026831

Kusmec A, Srinivasan S, Nettleton D, Schnable PS (2017) Distinct genetic architectures for phenotype means and plasticities in Zea mays. Nature Plants 3(9):715–723. https://doi.org/10.1038/s41477-017-0007-7

Lacaze X, Hayes PM, Korol A (2009) Genetics of phenotypic plasticity: QTL analysis in barley Hordeum Vulgare. Heredity 102(2):163–173. https://doi.org/10.1038/hdy.2008.76

Lee SH, van der Werf JHJ (2016) MTG2: An efficient algorithm for multivariate linear mixed model analysis based on genomic information. Bioinformatics 32(9):1420–1422. https://doi.org/10.1093/bioinformatics/btw012

Li X, Guo T, Mu Q, Li X, Yu J (2018) Genomic and environmental determinants and their interplay underlying phenotypic plasticity. Proc Nat Acad Sci U S Am 115(26):6679–6684. https://doi.org/10.1073/pnas.1718326115

Li C, Wu X, Li Y, Shi Y, Song Y, Zhang D, Li Y, Wang T (2019) Genetic architecture of phenotypic means and plasticities of kernel size and weight in maize. Theor Appl Genet 132(12):3309–3320. https://doi.org/10.1007/s00122-019-03426-w

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 Genes, Genom, Genet 5(4):569–582. https://doi.org/10.1534/g3.114.016097

Ly D, Huet S, Gauffreteau A, Rincent R, Touzy G, Mini A, Jannink JL, Cormier F, Paux E, Lafarge S, le Gouis J, Charmet G (2018) Whole-genome prediction of reaction norms to environmental stress in bread wheat (Triticum aestivum L.) by genomic random regression. Field Crop Res 216:32–41. https://doi.org/10.1016/j.fcr.2017.08.020

Meyer K (2009) Factor-analytic models for genotype × environment type problems and structured covariance matrices. Genet Sel Evol 41(1):54354. https://doi.org/10.1186/1297-9686-41-21

Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, Charcosset A, Welcker C, van Eeuwijk F, Tardieu F (2019) Genomic prediction of maize yield across European environmental conditions. Nat Genet 51(6):952–956. https://doi.org/10.1038/s41588-019-0414-y

Ni G, van der Werf J, Zhou X, Hyppönen E, Wray NR, Lee SH (2019) Genotype–covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. Nat Commun. https://doi.org/10.1038/s41467-019-10128-w

Paschke M, Bernasconi G, Schmid B (2003) Population size and identity influence the reaction norm of the rare, endemic plant Cochlearia bavarica across a gradient of environmental stress. Evolution 57(3):496–508. https://doi.org/10.1111/j.0014-3820.2003.tb01541.x

Perkins JM, Jinks JL (1968) Environmental and genotype-environmental components of variability III Multiple lines and crosses. Heredity 23(3):339–356. https://doi.org/10.1038/hdy.1968.48

Powell N, Ji X, Ravash R, Edlington J, Dolferus R (2012) Yield stability for cereals in a changing climate. Funct Plant Biol 39(7):539–552. https://doi.org/10.1071/FP12078

Raffo MA, Sarup P, Andersen JR, Orabi J, Jahoor A, Jensen J (2022) Integrating a growth degree-days based reaction norm

methodology and multi-trait modeling for genomic prediction in wheat. Front Plant Sci. https://doi.org/10.3389/fpls.2022.939448

Sadras VO, Montoro A, Moran MA, Aphalo PJ (2012) Elevated temperature altered the reaction norms of stomatal conductance in field-grown grapevine. Agric Forest Meteorol. https://doi.org/10.1016/j.agrformet.2012.06.005

Sandhu KS, Mihalyov PD, Lewien MJ, Pumphrey MO, Carter AH (2021) Genomic selection and genome-wide association studies for grain protein content stability in a nested association mapping population of wheat. Agronomy 11(12):2528. https://doi.org/10.3390/agronomy11122528

Sjoberg SM, Carter AH, Steber CM, Garland-Campbell KA (2020) Unraveling complex traits in wheat: approaches for analyzing genotype × environment interactions in a multienvironment study of falling numbers. Crop Sci 60(6):3013–3026. https://doi.org/10.1002/csc2.20133

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

Smith A, Cullis B, Gilmour A (2001) Applications: the analysis of crop variety evaluation data in Australia. Aust New Zealand J Statis 43(2):129–145. https://doi.org/10.1111/1467-842X.00163

Smith A, Norman A, Kuchel H, Cullis B (2021) Plant variety selection using interaction classes derived from factor analytic linear mixed models: models with independent variety effects. Front Plant Sci. https://doi.org/10.3389/fpls.2021.737462

Tan Y, Zhou J, Wang J, Sun L (2020) The genetic architecture for phenotypic plasticity of the rice grain ionome. Front Plant Sci. https://doi.org/10.3389/fpls.2020.00012

Tolhurst DJ, Mathews KL, Smith AB, Cullis BR (2019) Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. J Anim Breed Genet 136(4):279–300. https://doi.org/10.1111/jbg.12404

Tolhurst DJ, Gaynor RC, Gardunia B, Hickey JM, Gorjanc G (2022) Genomic selection using random regressions on known and latent environmental covariates. Theor Appl Genet. https://doi.org/10.1007/s00122-022-04186-w

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

Waters DL, Clark SA, Moghaddar N, van der Werf JH (2022) Genomic analysis of the slope of the reaction norm for body weight in Australian sheep. Genet Sel Evol. https://doi.org/10.1186/s12711-022-00734-6

Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Asyraf MD, Hatta M, Hinchliffe A, Steed A, Reynolds D, Adamski NM, Breakspear A, Korolev A, Rayner T, Dixon LE, Riaz A, Martin W, Ryan M, Edwards D, Hickey LT (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. Nature Plants 4(1):23–29. https://doi.org/10.1038/s41477-017-0083-8

Woltereck R (1913) Weitere experimentelle untersuchungen über Artänderung, speziell über das Wesen quantitativer Artunterschiede bei Daphniden. Z Indukt Abstamm Vererbungsl 9(1):146. https://doi.org/10.1007/BF01876686