

Statistical Modeling of Vigor Ratings in Ruzigrass Breeding

JALES MENDES OLIVEIRA FONSECA

Bayer (United States)

FLAVIA MARIA AVELAR GONÇALVES

Federal University of Lavras

FAUSTO SOUZA SOBRINHO

Brazilian Agricultural Research Corporation

JÚLIO SÍLVIO DE SOUZA BUENO FILHO

Federal University of Lavras

FLÁVIO RODRIGO GANDOLFI BENITES

Brazilian Agricultural Research Corporation

DAVI HENRIQUE LIMA TEIXEIRA

Federal Rural University of Amazonia

JOSÉ AIRTON RODRIGUES NUNES

jarnunes@uf1a.br

Federal University of Lavras

Research Article

Keywords: Urochloa ruziziensis, selection efficiency, Box-Cox transformation, Threshold model, Mixed Model Approach, Forage Breeding

Posted Date: February 22nd, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-3970804/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Ruzigass (*Urochloa ruziziensis*) is a forage crop with high agronomic and nutritional value. Plant breeders often assess ruzigrass phenotypic traits using vigor ratings. The analyses of these categorical data often fail to meet usual statistical assumptions. In this study we compared four fittings of linear models for vigor rating analyses: i) a mixed model for the original scale (LMM), ii) a mixed model for a Box-Cox transformed scale (BCLMM), iii) a multinomial generalized mixed model (GLMM), and iv) a hierarchical Bayesian model (HBM). Additionally, biomass yield was assessed, and indirect selection of high-performing genotypes was evaluated. The experimental design had 2,204 ruzigrass genotypes randomized to augmented blocks. Six graders visually assessed each plot using a rating scale. Fitting methods were sampled from three scenarios, using just one, three, or six graders. A non-null genetic variance component was detected for both traits. Except for BCLMM, methods for analyzing vigor ratings were correlated. The correlations and coincidence indexes for selecting genotypes increased with the number of graders. The analysis of vigor ratings under gaussian approximations is riskier when a single grader evaluates genotypes. GLMM and HBM are more recommendable and suitable analyses of vigor ratings to select high-performing ruzigrass genotypes.

INTRODUCTION

Ruzigass [*Urochloa ruziziensis* (R. Germ. & CM Evrard) Crins.], syn. *Brachiaria ruziziensis*, is a crop widely used in forage breeding programs presenting high agronomic and nutritional value (Marcelino et al. 2020; Nouhoun et al. 2022). Classified as an obligate sexual species due to its ploidy level ($2n = 2x = 18$), it can produce interspecific hybrids when chromosomes are duplicated ($2n = 4x = 36$) (Pereira et al. 2005; Timbó et al. 2014).

The development of a new ruzigass cultivar requires selecting and recombining best-performing genotypes across multiple breeding cycles to accumulate favorable alleles and produce desirable agronomic traits. Assessing these traits in forage crops can be costly, time-consuming, and labor-intensive depending on the stage and the size of trials. Forage breeders have often assigned categorical ratings to discriminate genotypes in field evaluations for some important traits such as resistance to pests (Silva et al. 2013), regrowth capacity (Gouveia et al. 2022), turf quality (Whitman et al. 2022), flowering (Marcon et al. 2021), biomass production (Price and Casler 2014; Teixeira et al. 2020; Dias et al. 2022). Embrapa Dairy Cattle Research Center has used vigor ratings to select at the early stages of the ruzigass breeding program from many genotypes. The vigor rating was devised to adjust biomass rating contemplating many plant aspects like health, size of leaves and stalks, height, regrowth, and tillering. However, the selection of genotypes has been based solely on realized field ratings, with no further statistical analysis.

Some statistical analyses have been used to deal with ratings or scores by plant breeders. Ratings are ordinal categorical variables (Agresti 2002); nonetheless, average ratings have been successfully approximated by Gaussian distributions via central limit theorem (Mood et al. 1974). Sometimes this

assumption does not hold, and non-linear data transformations have been used (e.g. Box and Cox 1964). The search for a model that represents the actual data may improve the interpretation of parameters and predictive ability in general, avoiding spurious results (Jaeger 2008; Bolker et al. 2009; Stroup 2015). For breeding purposes, this could avoid erroneous genotype ranking and deliver more efficient selection.

Plant breeders are continuously searching for flexible methods to analyze data. The new statistical tools available in free software (R packages) have been used to avoid data transformation in count data or proportions (Venables and Ripley 2002; Bates et al. 2015; Christensen 2019; Hadfield 2019; R Core Team 2020). Generalized linear mixed models can be fit to different distributions of the exponential family and are an extension of the linear mixed models (McCulloch and Searle 2004). The exponential family unifies several random variables, among which we could enroll the multinomial distribution and, ordinal categorical data (Nelder and Wedderburn 1972). Those models could be fit either in a frequentist (McCulloch and Searle 2004) or a Bayesian approach (Gianola and Foulley 1983; Gianola and Fernando 1986). In the last case, hierarchical model formulation is needed to build prior distributions, from which joint posterior distributions for parameters and data can be worked out. This results in convenient methods to make inference on marginal distributions for parameters. Gouy et al. (2013) emphasized the suitability of threshold models under a hierarchical Bayesian framework for analyzing rust resistance reaction ratings to select sugarcane clones. Correa et al. (2016) used usual hierarchical formulation for threshold models from animal and plant breeding purposes.

In this study we aim to find flexible models to analyze vigor ratings for efficiently select high-performing ruzigrass genotypes, comparing four methods, including i) a linear mixed model in original scale (LMM), ii) a linear mixed model in transformed Box-Cox scale (BCLMM), iii) a multinomial generalized linear mixed model (GLMM), and iv) a hierarchical Bayesian model (HBM). Analyses will use R packages. The effect of having multiple accessors in final selection of ruzigrass genotypes by different methods is also investigated.

MATERIALS AND METHOD

Location

The study was carried out at the Embrapa Dairy Cattle experimental station situated at 21°33' S latitude and 43°06' W longitude at 410 m above sea level in Coronel Pacheco, MG, Brazil. Soil is predominantly classified as red yellow Argisol (Santos et al. 2006). According to Köppen, the climate is classified as humid subtropical (Cwa type mesothermic) with an average annual temperature of 19°C. Winters are commonly dry and cold, while summers are rainy with moderately high temperatures. The annual average rainfall is 1536 mm.

Experimental Design and Germplasm

In August 2011, the experiment was laid out as an augmented block design with 51 blocks ranging from 28 to 72 genotypes within each block. Nonreplicated treatments consisted of 2,204 ruzigrass genotypes

derived from selected seeds during the second recurrent intraspecific forage breeding program cycle at Embrapa Dairy Cattle. Marandu (*U. brizantha*) and Basilisk (*U. decumbens*) cultivars were used as checks, thus replicated in all blocks. Plants were transplanted and plots were composed of a single plant with an average area of 1.5 m² spaced one meter apart.

Plant fertilization of 350 kg per hectare of the formula 8-28-16 (Nitrogen-Phosphorus-Potassium) was applied. Topdressing fertilization of 1 ton per hectare of the formula 20-05-20 (Nitrogen-Phosphorus-Potassium) was partitioned across each cutting during the rainy season. Manual weeding was performed when necessary, and a cut to homogenize all treatments was carried out at the beginning of December 2011. In 2012, two evaluation cuts were performed, one in January and a second approximately 40 days later. For this study, only the first evaluation cut was analyzed.

Traits Evaluated

Vigor rating and biomass yield were assessed per plot. Visual ratings of the plant vigor ranged from 1 to 5 grade scale as follows: 1 - very bad, 2 - bad, 3 - regular, 4 – good, and 5 - very good (Fonseca et al. 2020). Six different graders performed visual assessments immediately before the evaluation cut. To evaluate biomass yields, plots were cut at 5.0 cm above the soil level, and biomass was weighed using a portable suspension scale in grams.

Statistical Analysis

a) Biomass yield

Biomass yield data were checked for outliers and their residuals were tested for normality. The residuals were inspected by both the Shapiro-Wilk test (Shapiro and Wilk 1965) and diagnostic plots (Q-Q plot) using the *shapiro.test* and *qqnorm* functions of the *stats* R package (R Core Team, 2020). The model presented in Fonseca et al. (2020) was fitted as follows:

$$y = X\mu + Z_1b + Z_2g + e$$

1
,

where y is the vector of biomass yield; μ is the intercept; b is the vector of block effects, $b \sim N(0, I\sigma_b^2)$; g is the vector of genotype effects, $g \sim N(0, I\sigma_g^2)$; e is the vector of errors, $e \sim N(0, I\sigma_e^2)$; and X , Z_1 , Z_2 are the design matrices of the effects μ , b , and g , respectively. The estimation of fixed effects (best linear unbiased estimator - BLUE) and the prediction of random effects (best linear unbiased prediction - BLUP) were performed via the solution of Henderson's system of equations (Henderson 1984). Variance components were estimated by the restricted maximum likelihood (REML) method (Patterson and Thompson 1971). The significance of variance components were evaluated by the likelihood ratio test considering a significance level of 5% using the *lmerTest* package (Kuznetsova et al. 2017) available in R (R Core Team 2020). The *lmer* function of the *lme4* R package was used to fit the linear mixed model

(Bates et al. 2015). Repeatability (R) was calculated following Cullis et al. (2006) as presented in Schmidt et al. (2019): $R_{Cullis}^2 = 1 - \frac{\bar{\nu}_{\Delta..}^{BLUP}}{2\sigma_g^2}$, where $\bar{\nu}_{\Delta..}^{BLUP}$ is the mean variance of the difference between two BLUPs and σ_g^2 is the genotypic variance based on biomass.

b) Vigor rating

Vigor ratings were analyzed using linear mixed model in original scale (LMM), a linear mixed model in transformed Box-Cox scale (BCLMM), a multinomial generalized linear mixed model (GLMM), and a multinomial hierarchical Bayesian model (HBM) in three distinct scenarios as follows:

b1) Scenario 1 – One grader

To analyze the vigor rating using the LMM, model (1) was applied, replacing the response variable biomass with vigor rating. Similarly, for the BCLMM, model (1) was applied to adjust vigor ratings under the Box-Cox transformation scale using the following expression:

$$y_t(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases}$$

where λ is the Box-Cox transformation parameter; y is the vector of vigor ratings; and $y_t(\lambda)$ is the vector of transformed vigor ratings. The λ estimate was calculated by maximizing the likelihood function (Box and Cox 1964) of model (1) assuming all effects as fixed. For that, the *boxcox* function available in the *MASSR* package was applied (Venables and Ripley 2002). The normality of the residuals was checked by the Shapiro-Wilk test (Shapiro and Wilk 1965) and Q-Q plot procedure using the respective functions *shapiro.test* and *qqnorm* of the *stats* R package (R Core Team 2020).

To investigate the vigor ratings using the GLMM under a frequentist approach, the data were analyzed via a threshold model assuming a multinomial distribution (Agresti 2002; Agresti 2007) and a probit link function as follows:

$$\Phi^{-1}(\pi_c | \mu, b, g) = X\mu + Z_1b + Z_2g$$

where Φ^{-1} is the probit link function; π_c is the probability associated with each category; c is the number of categories (i.e., $c = 5$ in this study); μ is the intercept; b is the vector of block effects, $b \sim N(0, I\sigma_b^2)$; g is the vector of genotype effects, $g \sim N(0, I\sigma_g^2)$; and X, Z_1, Z_2 are the design matrices of the effects μ, b and g , respectively. Maximum likelihood estimates of parameters and BLUPs of genotypic effects were obtained via Laplace approximation to compute the likelihood function (Pinheiro and Chao 2006). The

clmm function of the *ordinal* R package was used (Christensen 2019). The significance values of variance components were assessed by the likelihood ratio test. Repeatability of vigor ratings under LMM, BCLMM, and GLMM were calculated following Cullis et al. (2006) as previously described.

For the analysis of vigor rating using the HBM, i.e., the GLMM under a Bayesian approach, model (2) was extended to incorporate priors on each effect. For block and genotype, an inverse Gama *prior* was used; for fixed effects, a noninformative *prior* was used; and for the residual, a fixed *prior* of 1 was used (Hadfield 2019). The model ran 500,000 Markov chain Monte Carlo (MCMC) simulation iterations with a burn-in of 100,000 samples and a thin every 200 iterations, resulting in a posterior distribution of 2,000 samples of estimated parameters. Based on the samples, the mean and highest posterior density (HPD) interval were determined. The *MCMCglmm* function of the *MCMCglmm* R package was used (Hadfield 2019). Repeatability was calculated following the standard repeatability procedure as $R = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\phi^2}$, where

σ_g^2 is the genetic variance estimate, and σ_ϕ^2 is the probit distribution variance, which equals to 1. All methods used to assess vigor ratings were applied for each grader individually.

b2) Scenario 2 – Three graders

To analyze vigor ratings using the LMM and BCLMM, model (1) was applied, and an average vigor rating was used as the response variable. Alternatively, for the GLMM and HBM, vigor ratings were analyzed using an extension of model (2) to include the random effect of grader. Thus, these models allowed vigor rating to be assessed under its original scale instead of the average of the vigor ratings. For HBM, an inverse Gama *prior* was used to model the grader effect. In this scenario, the same three randomly selected graders were used in all methods.

b3) Scenario 3 – Six graders

Similar models presented in scenario 2 were developed, considering all six graders involved in this study. Therefore, the only difference between scenarios 2 and 3 is the number of graders used to calculate the average of vigor ratings for LMM and BCLMM; whereas, for GLMM and HBM, the difference is inclusion of three graders within their grader effect.

c) Comparison between vigor ratings and scenarios, and their correspondence to biomass yield

c1) Correlation and selection efficiency

Spearman correlations between biomass yields and vigor ratings were obtained from BLUPs of genotypes for all scenarios and methods. The *cor* function from the *stats* R package was used (R Core Team 2020). The selection efficiency was calculated following a modified version of the coincidence index (*C*) proposed by Hamblin and Zimmermann (1986). Originally, the CI aimed to assess the proportion of coincident genotypes selected under two different planting systems assuming a specific

selection intensity (SI). Herein, CI was used to calculate the selection efficiency of each model to select the top 5% high-performing genotypes using grades as follow:

$$CI = \frac{A - C}{B - C}$$

where A is the number genotypes selected based on grades and biomass simultaneously; B is the top 5% high-performing genotypes based on biomass; and C is the expected number of coincident genotypes selected by chance, i.e., a proportion of B due to chance ($C = B \times 0.05$).

c2) Response to selection simulation

Expected genetic gains based on direct and indirect selection were calculated considering the top 5% BLUP of genotypes. Response to selection were presented on normal and categorical scales.

RESULTS

The genetic variance was significant for biomass yield and vigor ratings in all methods and scenarios (Table 1). The distribution of vigor ratings varied across graders (Fig. 1), impacting the proportion of variance explained among different graders and methods (Fig. 2). The repeatability for biomass yield was 0.46, whereas, for vigor ratings, such magnitudes varied from 0.35 (HBM) to 0.85 (BCLMM) in scenario 1 (one grader). For scenario 2 (three graders), substantial differences occurred in GLMM and HBM methods regarding repeatability, latent variables, and thresholds estimates. Alternatively, LMM and BCLMM methods presented slight changes for repeatability, although the genetic and residual variances varied in magnitude. There were minor variations from scenario 2 to 3, i.e. when all six graders were assessed; an exception was the range of values for the latent variable at HBM.

Table 1

– Estimates of the means (μ), genetic variances ($\hat{\sigma}_g^2$), error variances ($\hat{\sigma}_e^2$), repeatability (R^2), latent variables, and thresholds for biomass yields and vigor ratings across three scenarios.

Scenario 1 (grader A)								
Parameter	Biomass	Vigor ratings				LI 2.5%	Posterior mean	LS 2.5%
		LMM	BCLMM	GLMM	HBM			
μ	1.2 kg/plant	2.83	14.24	-	-	-	-	-
$^2_{\hat{\sigma}_g}$	0.1418*	0.32*	180.92*	0.55*	0.25	0.56		0.92
$^2_{\hat{\sigma}_e}$	0.1579	0.50	28.03	1.00	-	1.00		-
R^2	0.46	0.38	0.85	0.34	0.20	0.35		0.48
Latent variable	-	-	-	-	1.62	1.92		2.21
Threshold 1	-	-	-	-1.93	1.32	1.49		1.69
Threshold 2	-	-	-	-0.43	2.58	2.89		3.23
Threshold 3	-	-	-	0.97	3.75	4.22		4.72
Threshold 4	-	-	-	2.30	-	-		-
Scenario 2 (graders A,B, and C)								
μ		3.37	71.7	-	-	-		-
$^2_{\hat{\sigma}_g}$		0.21*	3205.0*	3.00*	2.71	2.98		3.29
$^2_{\hat{\sigma}_e}$		0.39	660.3	1.00	-	1.00		-

*Significant by the likelihood ratio test at a p-value < 0.001; LMM: vigor ratings fitted by a linear mixed model; BCLMM: vigor ratings fitted by Box-Cox transformations; GLMM: vigor ratings fitted by a generalized linear mixed model; HBM: vigor ratings fitted by a hierarchical Bayesian model; IL: inferior limit; SL: superior limit.

Scenario 1 (grader A)						
R^2	0.35	0.82	0.88	0.73	0.75	0.76
Latent variable			-	0.53	4.16	7.99
Threshold 1			-4.25	2.14	2.25	2.38
Threshold 2			-1.98	4.39	4.55	4.71
Threshold 3			0.32	6.30	6.49	6.70
Threshold 4			2.26	-	-	-
Scenario 3 (all graders)						
μ	3.07	63.18	-	-	-	-
$\hat{\sigma}_g^2$	0.21*	3354.8*	2.72*	0.54	0.94	1.37
$\hat{\sigma}_e^2$	0.34	484.4	1.00	1.00	1.00	1.00
R^2	0.38	0.86	0.93	0.22	0.51	0.79
Latent variable	-	-	-	2.07	3.49	5.08
Threshold 1	-	-	-3.49	2.28	2.35	2.43
Threshold 2	-	-	-1.14	4.39	4.49	4.59
Threshold 3	-	-	0.99	6.17	6.30	6.43
Threshold 4			2.80	-	-	-
*Significant by the likelihood ratio test at a p-value < 0.001; LMM: vigor ratings fitted by a linear mixed model; BCLMM: vigor ratings fitted by Box-Cox transformations; GLMM: vigor ratings fitted by a generalized linear mixed model; HBM: vigor ratings fitted by a hierarchical Bayesian model; IL: inferior limit; SL: superior limit.						

The normality of the vigor rating residuals was assessed before and after performing a Box-Cox transformation using both a statistical test (Shapiro-Wilk normality test) and diagnostic plots (Q-Q plot). Results from both procedures showed that vigor rating residuals on the original scale did not follow a normal distribution, regardless of the numbers of graders considered in the analysis. Additionally, the

application of the Box-Cox transformation was not effective to ensure normality of the residuals (data not shown).

The correlations (r) between the BLUPs of genotypes for vigor ratings obtained from different methods were overall high, ranging from 0.85 to 0.99 across the scenarios (Figs. 3, 4, and 5). While the inclusion of graders slightly decreased the correlation between BLUPs of genotypes for vigor ratings, the correlation between vigor ratings and biomass increased. Such an increase reflected the coincidence index estimates, whose values were highest for GLMM (0.44) and HBM (0.43) (Fig. 5). The coincidence index between vigor ratings across different methods was overall high, regardless of the scenario, with the exception being BCLMM (Figs. 3, 4, and 5). These results indicated a potential benefit to select high-performing genotypes at earlier stages of ruzigrass breeding using vigor ratings. Nonetheless, the method and the number of graders can limit the benefits to indirectly select genotypes.

Simulations based on the direct response to selection using the top 5% BLUP of genotypes showed a 50% increase in biomass (Table 2), whereas the indirect response to selection reported relatively lower values. In the case of scenario 1, the vigor ratings resulted in an indirect response to selection of a ~ 35% increase in biomass using LMM, GLMM, and HBM, whereas the BCLMM resulted in a ~ 12% increase. Similar results were observed in scenarios 2 and 3, with a significant increase only when BCLMM was applied.

Table 2
– Simulations based on direct and indirect response to selection considering the top 5% BLUP of genotypes.

Response to selection (kg/plot)			
Biomass	0.603		
	Scenario 1	Scenario 2	Scenario 3
LMM	0.418	0.423	0.444
BCLMM	0.140	0.273	0.232
GLMM	0.418	0.442	0.435
HBM	0.418	0.445	0.433

The application of generalized models using the probit link function provided estimates of thresholds between categories, which allowed the representation of response to selection considering the vigor ratings ordinal categorical scale (Fig. 6). Moreover, such threshold estimates enabled estimating the probability of a genotype to be found in a given category. For instance, ~ 80% of the genotypes were likely to be found in categories 2 and 3, whereas 2.71% and 1.05% of genotypes were likely to be present in categories 1 and 5, respectively (red graph). After simulating the selection of the top 5% BLUP of genotypes based on vigor ratings, there was an increase in the likelihood of finding genotypes in higher categories. For example, the probabilities of finding genotypes in categories 4 and 5 were 40.04% and

8.45%, respectively (blue graph), while such probabilities were significantly lower in the original population (red graph).

DISCUSSION

The utilization of visual ratings in forage breeding is justified due to their feasibility and efficiency for indirectly selecting superior genotypes. Notwithstanding, the genetic architecture of each trait should be considered because it influences selection efficiency. For instance, qualitative traits are controlled by fewer loci and are less influenced by the environment than quantitative traits; thus, assessing qualitative traits using ratings (i.e. grade scale) can provide efficient decisions. Nevertheless, ratings for quantitative traits can also be useful whether the correlations between the trait of interest and their corresponding ratings are relatively high. For instance, Silva et al. (2013) used ratings to assess the genetic merit of resistant ruzigrass genotypes to *Collaria oleosa*. Souza Sobrinho et al. (2010) indirectly selected ruzigrass genotypes that were resistant to two species of spittlebugs (*Mahanarva spectabilis* and *Deois schach*) using visual ratings. Moreover, Riday (2009) reported on the efficiency of visual ratings to select high-yielding red clover genotypes. Alternative to its efficiency on selecting superior genotypes, visual ratings can also assist breeders in eliminating undesirable genotypes (Atkins 1964; Burton 1982).

It should be noted that the vigor rating analyses presented herein in three different scenarios have their specific utilities given the populations under investigation and the selection phase of the breeding pipeline. Breeding programs have distinct evaluation/selection phases that ultimately give rise to new cultivars. As these phases progress and selections are performed, genetic variances among genotypes diminish; hence, the utilization of more graders increases the power of the applied method to detect real differences between similar genotypes, thereby increasing genetic gains (Fonseca et al. 2020). Nonetheless, a single grader might be considered for selecting high-performing genotypes (or discarding poor ones) in the early stages of selection cycles where the genetic variances between genotypes are highest. This study focused on the evaluation of ruzigrass genotypes at early stages of a selection cycle following the forage breeding pipeline at Embrapa Dairy Cattle. Broad genetic variation in biomass yields and vigor ratings were observed, allowing for the selection of superior genotypes. However, low repeatability estimates for both traits indicated that assessments of the true genotypic merit might be challenging either because of i) a pronounced environmental effect associated with the experimental design used (augmented block design) or ii) the quantitative genetic architecture of the traits under field evaluation (biomass yields and plant vigor ratings). Therefore, forage breeding programs that follow the Embrapa Dairy Cattle should avoid using a single grader to perform biomass selections via vigor ratings. Although it is possible to visually screen ruzigrass under field evaluations, the effect of a single grader can vary considerably and ultimately affect decisions. Fonseca et al. (2020) proposed a method to calculate the optimized number of graders to assess genotypes visually, and for ruzigrass it is recommended at least three.

Besides the variation related to the number of graders, additional factors can influence on the estimates of important parameters. For instance, the variations in the repeatability estimates across different

methods and scenarios reflected the scale assumed to perform the analysis. While the GLMM and HBM presented similar magnitudes, as both estimates were presented in a probit scale, the LMM assumed a normal scale and the BCLMM assumed a scale derived from its lambda, thus resulting in different magnitudes. The high repeatability estimates for the BCLMM in scenarios 1 and 2 should be cautiously interpreted. Besbes et al. (1993) also found an increase in heritability when Box-Cox transformations were applied to transform data describing egg production in laying hens. Such high magnitudes presented here become questionable when the model adjustments between methods are compared against each other. The BCLMM under scenarios 1 and 2 showed a noticeably high deviance (data not showed), thus reinforcing the idea that data transformation can lead to spurious results (Jaeger 2008; Bolker et al. 2009; Stroup 2015).

Correlations between the BLUPs of biomass yields and the BLUPs of vigor ratings showed a strong positive relationship, supporting the efficiency of indirect selection (Figs. 3, 4, and 5). Moreover, such positive relationships increased in scenarios 2 and 3 for all methods. A deeper analysis of the BCLMM scatterplot in scenario 1 (Fig. 1) permits identifying the discretization of its BLUPs, which corresponds to the same number of categories (i.e., five) used to assess genotypes visually. Such discretization, nonetheless, dissolved in following scenario due to an approximation of the normal distribution of the vigor ratings as the number of graders increased, supporting what the central limit theorem suggests (Mood et al. 1974). However, the approximation derived from the Box-Cox transformation did not assure normality in any scenario; thus, its application should be carried out with caution. Alternatively, the LMM provided comparable results when contrasted to the more flexible models (i.e., GLMM and HBM). This is likely due to the high quantity of the genotypes involved in this trial (2004 unreplicated genotypes plus two checks), indicating that the central limit theorem is a robust principle and can support the analysis of vigor ratings using LMM even though some assumptions are violated. Further, LMM is simpler and less computationally demanding. Nonetheless, it is expected that GLMM and HBM provide better estimates when trials investigate fewer genotypes.

Although the correlations between methods followed similar positive results, the CI estimates showed that BCLMM did not identify the top 5% high-performing ruzigrass genotypes; thus, the simulated increase in biomass derived from that analysis is likely due to random sampling. However, the LMM, GLMM, and HBM generated a significant response to selection (~ 37%) when comparing the response to selection based on biomass per se (~ 50%). The increments in biomass presented herein derived from simulations and should not be compared to the actual response to selections documented elsewhere. These results had the intention to compare the effects of different methods to analyze vigor ratings and assess the benefits of using grade scales to evaluate forage biomass.

Although different models provided similar responses to selection, vigor rating adjustments via generalized models are more appropriate for statistical judgment (Jaeger 2008; Bolker et al. 2009; Stroup 2015). Generalized models are more flexible due to their abilities to model response variables via distributions that share similar properties unified by the exponential family (Nelder and Wedderburn 1972). However, because of their intrinsic complexities, such models are disregarded in plant breeding.

This study presented an application of the GLMM and HBM by incorporating a multinomial distribution and a probit link function using packages available in R (R Core Team 2020). Multinomial distributions better describe ordinal categorical variables (e.g., vigor ratings) (Jansen 1991; Agresti 2002; Agresti 2007) than other distributions and enable threshold estimates between categories (Agresti 2002; Gianola and Fernando 1986); thus, results can be reported on their original scale. Falconer and Mackay (1996) reported the benefits of using a threshold model for classifying genotypes according to their genetic merits. As Fig. 6 demonstrated, there was a nearly 80% chance of finding a genotype in categories 2 and 3, while this probability was only 1% for category 5 (red graph). After simulating the selection of the top 5% BLUPs of vigor ratings, there were higher probabilities of finding genotypes in categories 4 (40.04%) and 5 (8.45%) (blue graph). Thus, selecting genotypes with higher ratings might increase the probability of finding genotypes in higher categories in future breeding cycles.

CONCLUSION

There are statistical alternatives to analyze vigor ratings including linear mixed models and generalized linear mixed models. However, some models can generate spurious results and should be cautiously used. The number of graders can affect decisions and ultimately limit the response to selection; one grader should not be used to visually assess ruzigrass biomass. Although there were no remarkable differences among LMM, GLMM and HBM, generalized models are most appropriate and informative for reporting vigor ratings due to their flexibility in incorporating distributions that effectively describe ordinal categorical scales. Vigor ratings can be used to select the best genotypes of ruzigrass for optimizing biomass yields during the early stages of a breeding program.

Declarations

Competing Interests

All Authors declare they have no financial interests.

Funding

The authors declare that the CNPq grants, or other support were received during the preparation of this manuscript.

Author Contribution

All authors contributed to the study's conceptualization and design. Material preparation and data curation were performed for FMAG, FSS, DHLT, and FRGB. The analyses were performed by JMOF, JARN, and JSSBF. The first draft of the manuscript was written by JMOF, FMAG, FSS, JSSBF, and JARN and all

authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgment

The authors thank Embrapa Dairy Cattle, the National Council for Scientific and Technological Development (CNPq), Coordination for the Improvement of Higher Education Personnel (CAPES) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for their financial support of this research.

References

1. Atkins, RE (1964) Visual selection for grain yield in barley. **Crop Science** 4:494-497. <https://doi.org/10.2135/cropsci1964.0011183X000400050018x>
2. Agresti, A (2007) An introduction to categorical data analysis (2^a ed.). New York: John Wiley & Sons.
3. Agresti, A. (2002) Categorical data analysis (2^a ed.). New York: John Wiley & Sons.
4. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. **Journal of Statistical Software** 67:1-48. <https://doi.org/10.18637/jss.v067.i01>
5. Besbes B, Ducrocq V, Foulley JL, Protais M, Tavernier A, Tixierboichard M, Beautnont C (1993) Box-Cox transformation of egg-production traits of laying hens to improve genetic parameter estimation and breeding evaluation. **Livestock Production Science** 33: 313–326. [https://doi.org/10.1016/0301-6226\(93\)90010-F](https://doi.org/10.1016/0301-6226(93)90010-F)
6. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. **Trends in Ecology Evolution**. 24:127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
7. Box GEP, Cox DR (1964) An analysis of transformations. **Journal of the Royal Statistical Society**. 26:211–243. <https://doi.org/10.1111/J.2517-6161.1964.TB00553.X>
8. **Science** 14:831–835. <https://doi.org/10.2135/cropsci1974.0011183x001400060016x>.
9. Burton GW (1982) Improved recurrent restricted phenotypic selection increases Bahiagrass forage yields. **Crop Science** 22:1058–1061. <https://doi.org/10.2135/cropsci1982.0011183x002200050040x>.
10. Christensen RHB (2019) Regression models for ordinal data [R package ordinal version 2019.12-10].
11. Correa FM, Silva JW, Ferreira DF, Bueno Filho JSS (2016) Bayesian algorithms for analysis of categorical ordinal data. **Brazilian Journal of Biometrics** 34:597–620.
12. Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. **Journal of Agricultural Biological and Environmental Statistics** 11:381–393. <https://doi.org/10.1198/108571106X154443>.

13. Dias JA, Rosado LR, Benites F, Souza Sobrinho F, Nunes JAR, Gonçalves FMA (2022) Efficiency of indirect selection for green biomass production of *Urochloa ruziziensis*. **Crop Breeding and Applied Biotechnology** 22: e417822212. <https://doi.org/10.1590/1984-70332022v22n2a22>
14. Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. 4th ed. Longman, Essex, England SE - xiii.
15. Fonseca JMO, Nunes JAR, Gonçalves FMA, Souza Sobrinho F, Benites FRG, Teixeira DHL (2020) Predictive approach to optimize the number of visual graders for indirect selection of high-yielding *Urochloa ruziziensis* genotypes. **Crop Breeding and Applied Biotechnology** 20: e329220314. <http://dx.doi.org/10.1590/1984-70332020v20n3a48>
16. Gianola D, Fernando RL (1986) Bayesian methods in animal breeding Theory. **Journal of Animal Science** 63:217–244. <https://doi.org/10.2527/jas1986.631217x>
17. Gianola D, Foulley J (1983) Sire evaluation for ordered categorical data with a threshold model. **Genetics Selection Evolution** 15(2): 201-224. <https://doi.org/10.1186/1297-9686-15-2-201>
18. Gouveia B T, Mateus RG, Barrios SCL, Valle CB, Bueno Filho JSS, Rios EF, Dias AM, Nunes, JAR (2022). Combining ability and selection for agronomic and nutritional traits in *Urochloa* spp. hybrids. **Grass and Forage Science** 77:33– 44. <https://doi.org/10.1111/gfs.12555>
19. Gouy M, Rousselle Y, Bastianelli D, Lecomte P, Bonnal L, Roques D, Efile J-C, Rocher S, Daugrois J, Toubi L, Nabeneza S, Hervouet C, Telismart H, Denis M, Thong-Chane A, Glaszmann JC, Hoarau J-Y, Nibouche S, Costet L (2013) Experimental assessment of the accuracy of genomic selection in sugarcane. **Theoretical and Applied Genetics** 126:2575–2586. <https://doi.org/10.1007/s00122-013-2156-z>.
20. Hadfield J (2019). Markov chain Monte Carlo generalised linear mixed models - Course Notes. <https://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf> (accessed 2 December 2020).
21. Hamblin J; Zimmermann MJO (1986) Breeding common bean for yield in mixtures. **Plant Breeding Reviews** 4:245-272.
22. Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph.
23. Jaeger TF (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. **Journal of Memory and Language** 59:434–446. <https://doi.org/10.1016/j.jml.2007.11.007>.
24. Jansen J (1991) Fitting regression models to ordinal data. **Biometrical Journal** 33:807–815. <https://doi.org/10.1002/bimj.4710330707>.
25. Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: tests in linear mixed effects models. **Journal of Statistical Software** 82:1-26. <https://doi.org/10.18637/jss.v082.i13>
26. Marcelino L L, Moreira GR, Souza Sobrinho F, Almeida MIV, Cóser AC, Cunha GM, Benites F RG (2020) Nutritive value of improved populations *Brachiaria ruziziensis*. **Semina: Ciências Agrárias** 41:323-334. <https://doi.org/10.5433/1679-0359>

27. Marcon F, Brugnoli EA, Nunes JAR, Gutierrez VA, Martínez EJ, Acuña CA (2021) Evaluating general combining ability for agromorphological traits in tetraploid bahiagrass. **Euphytica**, 217:1-11. <https://doi.org/10.1007/s10681-021-02942-5>
28. McCulloch CE, Searle SR (2004) Generalized, linear, and mixed models. John Wiley & Sons.
29. Nelder JA, Wedderburn RWM (1972) Generalized linear models. **Journal of the Royal Statistical Society** 135:370-384. <https://doi.org/10.2307/2344614>.
30. Mood AM; Graybill FA; Boes DC (1974) Introduction to the Theory of Statistics. 3rd McGraw Hill.
31. Nouhoun Z, Traoré TC, Sawadogo ETBP, Ayantunde A, Prasad KVS, Blummel M, Balehegn M, Rios E, Dubeux JC, Boote K, Adesogan AT (2022). Herbage accumulation and nutritive value of cultivar Mulato II, Congo grass, and Guinea grass cultivar C1 in a subhumid zone of West Africa. **Agronomy Journal** 114:138-147. <https://doi.org/10.1002/agj2.20861>
32. Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. **Biometrika** 58:545-554. <https://doi.org/10.1093/biomet/58.3.545>.
33. Pereira AV, Souza Sobrinho F, Valle CB, Léo FJS, Botrel MA, Oliveira JS, Xavier DF (2005) Selection of interspecific *Brachiaria* hybrids to intensify milk production on pastures. **Crop Breeding Applied Biotechnology** 5:99-104. <https://doi.org/10.12702/1984-7033.v05n01a13>.
34. Pinheiro JC, Chao EC (2006) Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. **Journal of Computational and Graphical Statistics** 15:58-81. <https://doi.org/10.1198/106186006X96962>
35. Price DL, Casler MD (2014) Divergent selection for secondary traits in upland tetraploid switchgrass and effects on sward biomass yield. **Bioenergy Research** 7:329-337. <https://doi.org/10.1007/s12155-013-9374-8>.
36. R Core Team (2020) R: A language and environment for statistical computing. <https://www.r-project.org/>.
37. Riday H (2009). Correlations between visual biomass scores and forage yield in space planted red clover (*Trifolium pratense* L.) breeding nurseries. **Euphytica** 170:339-345. <https://doi.org/10.1007/s10681-009-9991-7>.
38. Schmidt P, Hartung J, Bennewitz J, Piepho HP (2019) Heritability in plant breeding on a genotype-difference basis. **Genetics** 212:991-1008. <https://doi.org/10.1534/genetics.119.302134>.
39. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (Complete Samples). **Biometrika** 52(3/4): 591. <https://doi.org/10.2307/2333709>.
40. Silva DM, Moraes JC, Auad AM, Fonseca MG, Souza Sobrinho F (2013) Genetic variability of *Brachiaria ruziziensis* clones to *Collaria oleosa* (Hemiptera: Miridae) based on leaf injuries. **American Journal of Plant Sciences** 4:2418-2424. <https://doi.org/10.4236/ajps.2013.412300>.
41. Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Oliveira JB, Coelho MR, Lumberras JF, Cunha TJF (2006) Sistema brasileiro de classificação de solos. Embrapa, Rio de Janeiro.

42. Souza Sobrinho F, Auad AM, Léo FJS (2010) Genetic variability in *Brachiaria ruziziensis* for resistance to spittlebugs. **Crop Breeding and Applied Biotechnology** 10:83-88.
43. Stroup WW (2015) Rethinking the analysis of non-normal data in plant and soil science. **Agronomy Journal** 107:811-827. <https://doi.org/10.2134/agronj2013.0342>.
44. Teixeira DHL, Gonçalves FMA, Nunes JAR, Souza Sobrinho F, Benites FRG, Dias KOG (2020) Visual selection of *Urochloa ruziziensis* genotypes for green biomass yield. **Acta Scientiarum. Agronomy** 42:4-12. <https://doi.org/10.4025/actasciagron.v42i1.42444>
45. Timbó ALO, Souza PNC, Pereira RC, Nunes JD, Pinto JEBP, Souza Sobrinho F, Davide LC (2014). Obtaining tetraploid plants of *Urochloa ruziziensis* (*Brachiaria ruziziensis*). *Revista Brasileira Zootecnia* 43:127-131. <https://doi.org/10.1590/S1516-35982014000300004>.
46. Venables WN, Ripley BD (2002) *Modern applied statistics with S*. Springer New York, New York, NY.
47. Whitman B, Iannone BV, Kruse JK, Unruh JB, Dale AG (2022) Cultivar blends: A strategy for creating more resilient warm season turfgrass lawns. *Urban Ecosystems* 25:797-810 <https://doi.org/10.1007/s11252-021-01195-3>

Figures

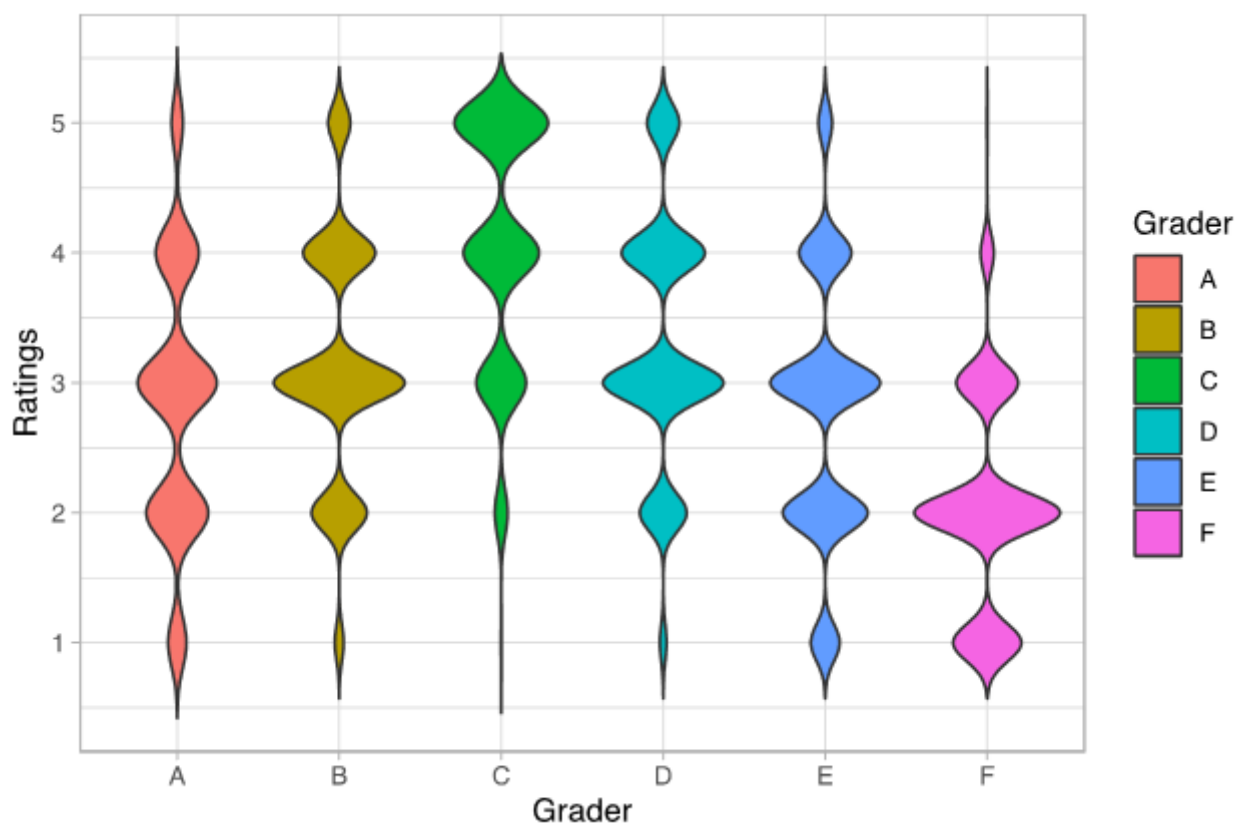


Figure 1

Distribution of vigor ratings across graders when assessing ruzigrassgenotypes visually

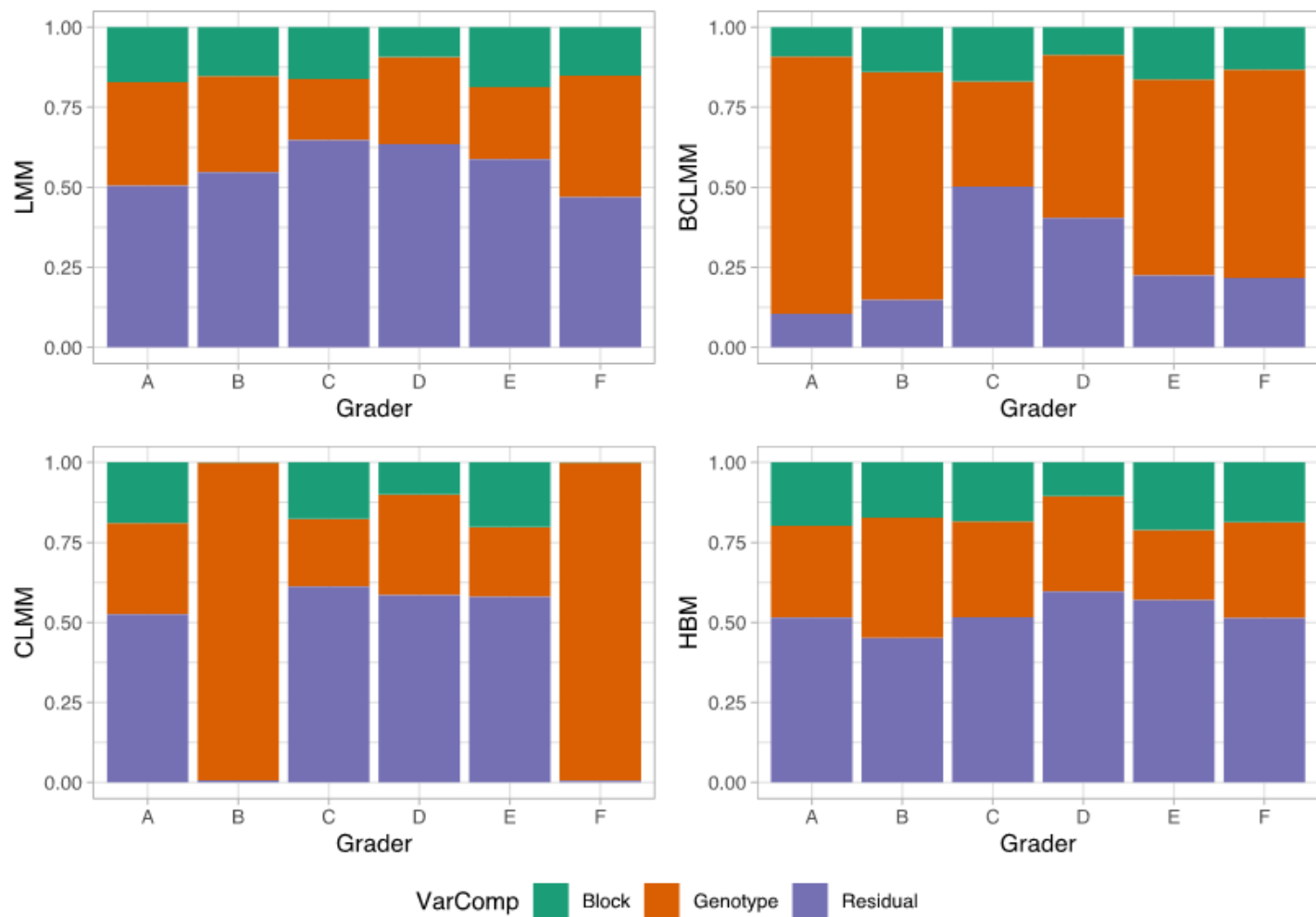


Figure 2

The proportion of variance components based on different graders and methods.

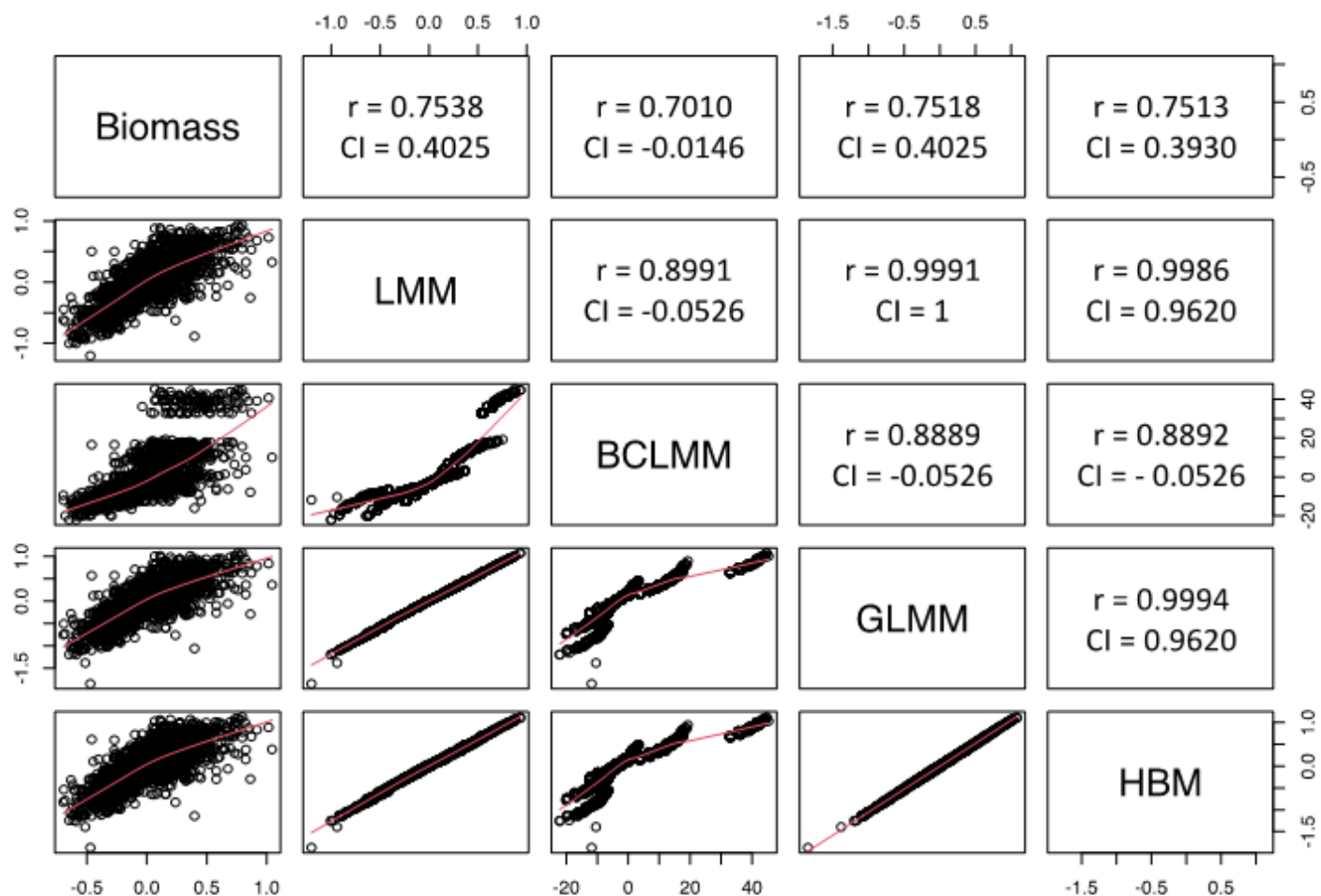


Figure 3

Spearman correlations and coincidence indices between the BLUPs of various genotypes for the biomass yields and vigor ratings of different methods for scenario 1 (one grader).

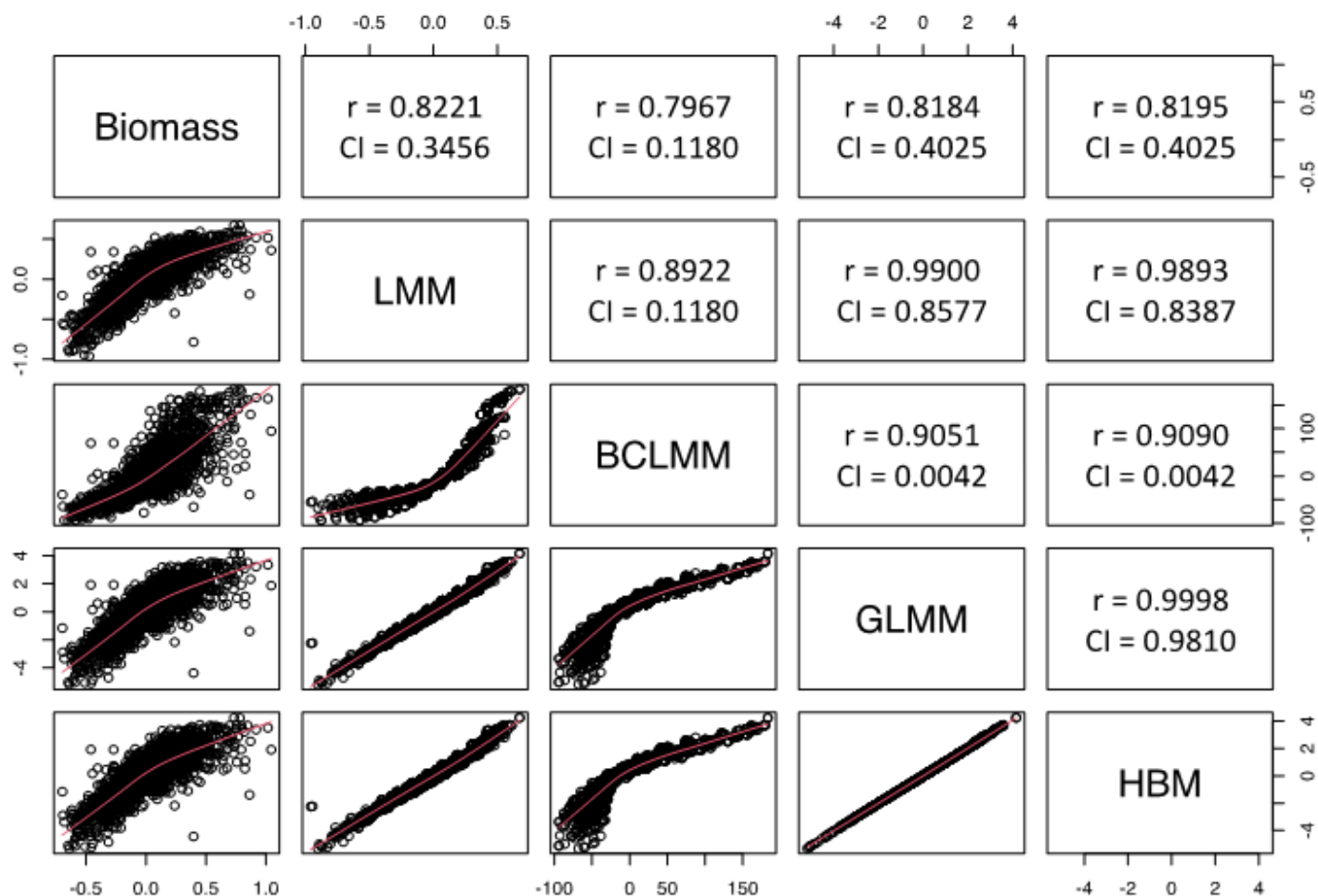


Figure 4

Spearman correlations and coincidence indices between the BLUPs of the biomass yields and the BLUPS of the average vigor ratings across different methods for scenario 2 (three graders).

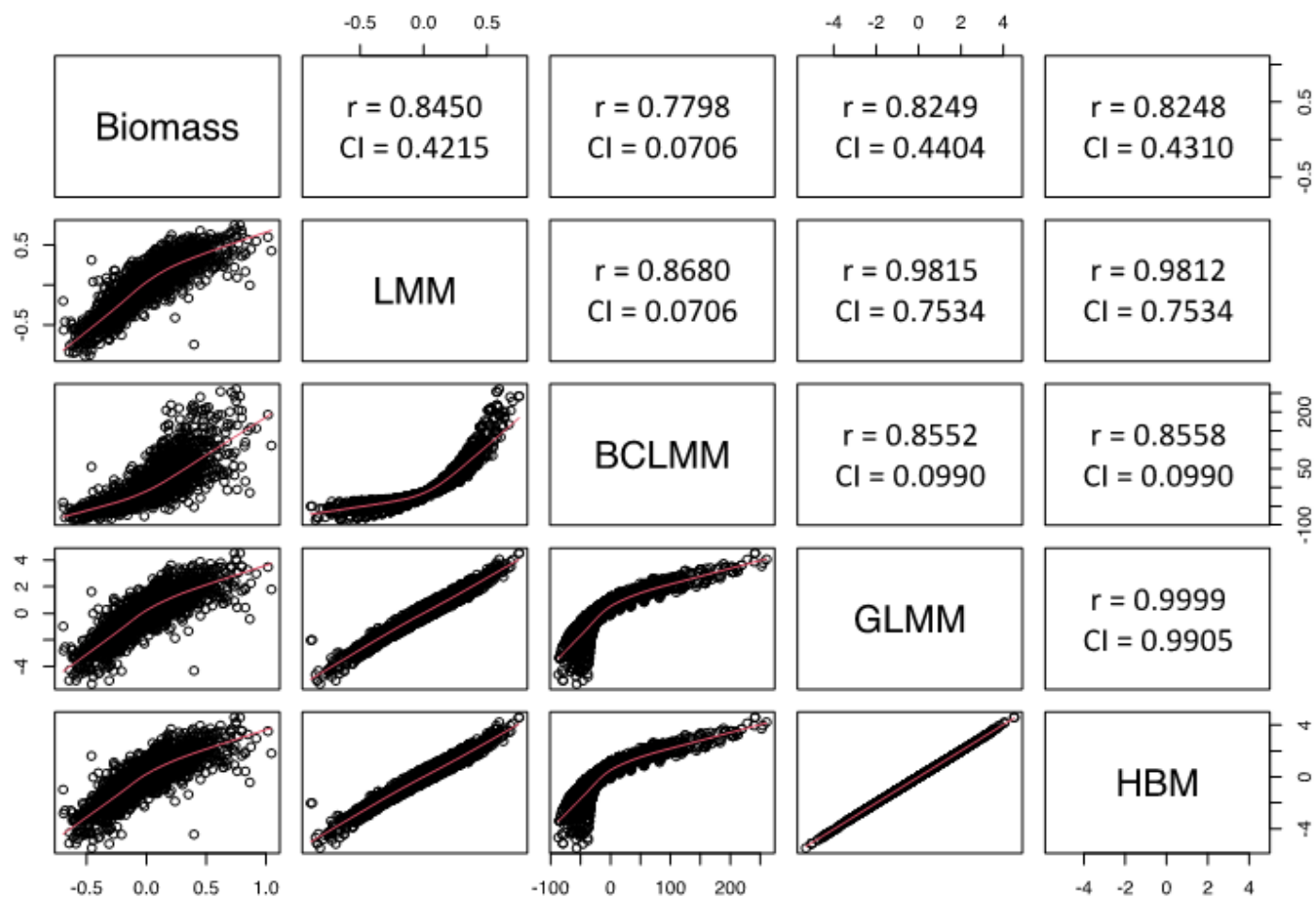


Figure 5

Spearman correlations and coincidence indices between the BLUPs of the biomass yields and the BLUPS of the vigor ratings across different methods for scenario 3 (six graders).

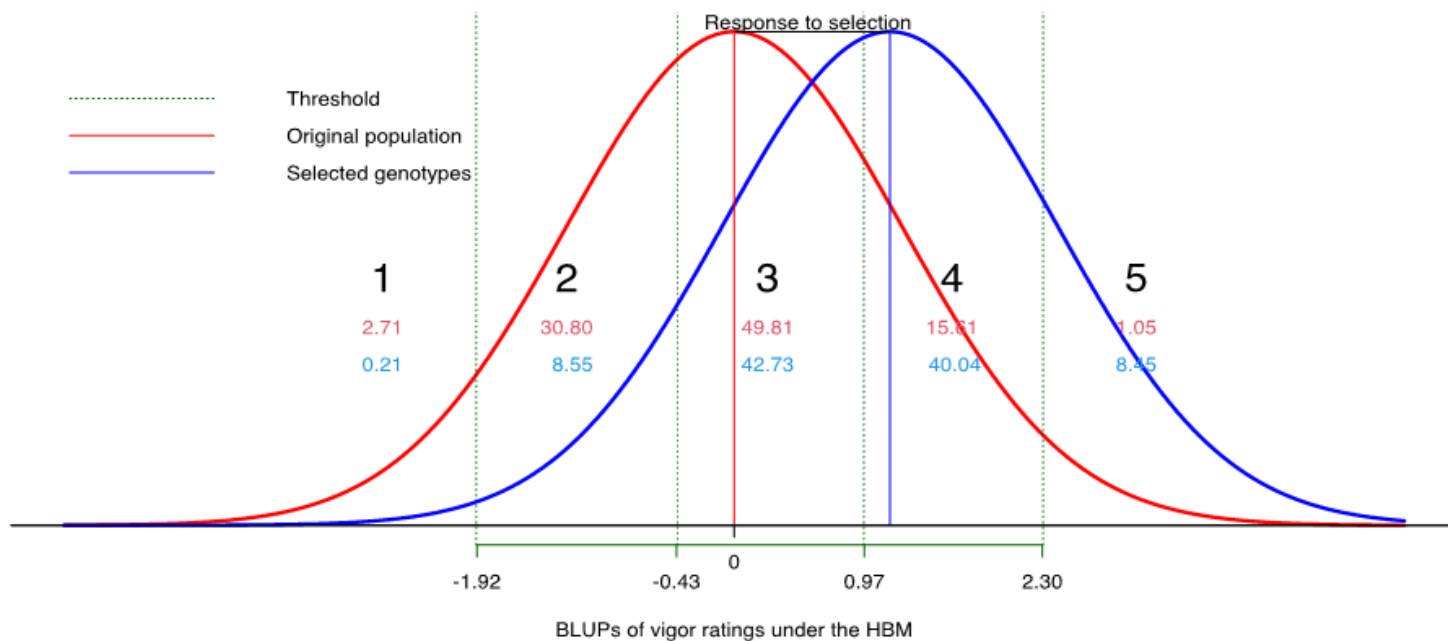


Figure 6

Predictions of genetic gains based on the rating in scenario 1 for the HBM.