Received: 20 December 2018 | Accepted: 21 July 2019 | Published online: 16 October 2020

DOI: 10.1002/csc2.20133

Crop Science

ORIGINAL RESEARCH ARTICLE

Crop Breeding & Genetics

Unraveling complex traits in wheat: Approaches for analyzing genotype × environment interactions in a multienvironment study of falling numbers

Stephanie M. Sjoberg[1] | **Arron H. Carter**[1] | **Camille M. Steber**[2] | **Kimberly A. Garland-Campbell**[2]

[1]Dep. of Crop and Soil Sciences, Washington State Univ., PO Box 646420, Pullman, WA 99164, USA

[2]USDA-ARS, Wheat Health, Genetics and Quality Research Unit, and Dep. of Crop and Soil Sciences, Washington State Univ., PO Box 646420, Pullman, WA 99164, USA

Correspondence
Kimberly A. Garland-Campbell, USDA-ARS, Wheat Health, Genetics and Quality Research Unit, and Dep. of Crop and Soil Sciences, Washington State Univ., PO Box 646420, Pullman, WA 99164, USA.
Email: kim.garland-campbell@ars.usda.gov

Funding information
Washington Grain Commission, Grant/Award Numbers: 7336, 7599; Agricultural Research Service; National Institute of Food and Agriculture, Grant/Award Number: 2015-05798

The last two authors contributed equally to this work.

Assigned to Associate Editor Jochum Wiersma.

Abstract

Multienvironment trials provide useful information about highly variable, complex plant traits like yield and quality but are difficult to analyze due to their frequently unbalanced nature, with genotypes and locations varying from year to year. Our objective was to use multiple approaches, including joint regression and principal components analysis, to characterize patterns in the genotype × environment interactions across an unbalanced 3-yr multienvironment wheat (*Triticum aestivum* L.) variety trial dataset, examining falling number (FN) test results in wheat. The FN test measures the decrease in flour gelling capacity resulting from starch digestion by the enzyme α-amylase. Low FN and high-α-amylase grain is discounted because it is associated with poor end-use quality. Low FN can be caused by susceptibility either to preharvest sprouting when it rains before harvest or to late-maturity α-amylase induction by temperature fluctuations during grain maturation. The most effective and visually intuitive approach for selecting varieties with high FN across variable environments was a combination of joint regression, such as Finlay–Wilkinson and Eberhart and Russell, with biplot methods such as the additive main effects and multiplicative interaction model (AMMI) and the genotype main effects and genotype × environment interaction model (GGE). We identify stable lines for FN resistance and provide a means to analyze unbalanced, multienvironment data from breeding and variety trials.

1 | INTRODUCTION

The wheat (*Triticum aestivum* L.) industry uses the Hagberg–Perten falling number (FN) test to measure starch degradation caused by α-amylase enzyme activity in flour (Perten, 1964; Yu et al., 2015). Alpha-amylase is produced during seed germination to mobilize stored reserves to fuel seedling growth. However, high grain α-amylase resulting in low FN is associated with poor-quality baked goods (Farrand, 1964; Kruger & Lineback, 1987). Thus, farmers receive steeply discounted prices for low-FN grain. To help farmers choose varieties less prone to low FN, >12,000 FN data points have been collected by researchers from Washington State University (WSU) Extension Cereal Variety Testing and the USDA-ARS since 2013 and made publicly available (www.smallgrains.wsu.edu,

Abbreviations: AMMI, additive main effects and multiplicative interaction model; BLUP, best linear unbiased predictor; FN, falling number; G × E, genotype × environment; GGE, genotype main effects and genotype × environment interaction model; LMA, late-maturity α-amylase; PC, principal component; PCA, principal component analysis; PHS, preharvest sprouting; WSU, Washington State University.

www.steberlab.org). Within this large dataset, the relative ranking of varieties by FN value varied greatly with location and year, in part because low FN can be caused by genetic susceptibility to two different forms of environmental stress, preharvest sprouting (PHS) and late-maturity α-amylase (LMA).

The primary goal of the current study was to explore multiple approaches to assess the relative performance of varieties while considering the complex genotype × environment (G × E) interactions controlling FN. The secondary goal was to assess these simple and visually intuitive approaches for their utility in helping scientists and farmers to see patterns in the data. The WSU Extension Cereal Variety Testing program provides comprehensive adaptation and performance information on small-grain varieties (comprising released cultivars and advanced breeding lines, which we will hereafter call either varieties or genotypes) across growing regions of eastern Washington to growers, industry, and breeding programs. However, interpretation of data for traits like FN is further complicated by the highly unbalanced nature of variety trials. Locations and genotypes change from one year to the next as the breeders submit new varieties and growers' preferences shift. Analysis methods for such multienvironment trials have evolved over the years. The original and most basic method of multienvironment trial data analysis used an ANOVA to split the variance into four components: genotype, environment, G × E interaction, and within-trial variation. Although this analysis is useful to partition the total variance into components due to genotype, environment, and the G × E interaction, it suffers from a lack of insight into how specific varieties respond to specific environments or to groups of similar environments (Kempton, 1984).

To more fully examine the G × E interaction, Finlay and Wilkinson (1963) developed a joint regression analysis where a genotype's performance in individual environments was regressed onto an environmental index of the overall effects of the environments that make up the trial. The index was usually defined as the mean performance of the genotypes in that environment. Further breakdown of the G × E interaction was introduced by Eberhart and Russell (1966), who defined genotype stability as the residual variance of a genotype from the slope of the Finlay–Wilkinson regression. Lin, Binns, and Lefkovitch (1986) described three types of stability:

Type 1: A genotype is considered stable if the variance is small among environments.

Type 2: A genotype is considered stable if its response to environments is parallel to the mean response of all genotypes in the trial.

Type 3. A variety is considered stable if its residual mean square from the regression model on the environmental mean is small.

Finlay–Wilkinson regression can be interpreted as Type 1 or 2, depending on the definition of a standard stable genotype. If stable genotypes are defined by having a slope of one, then Type 2 is implied, but if they are defined as a slope of zero, then Type 1 is implied. The Eberhart and Russell method measures Type 3 stability. Type 3 stability measures the unpredictable component of stability. Although Eberhart and Russell (1966) argue that this is the true definition of stability, Lin et al. (1986) explain that without using actual environmental factors as an environmental index in a prediction model, Type 3 measures of stability are least likely to reflect actual genotypic stability.

More recently, it has been common practice to group environments within large multienvironment trials by mega-environments defined by some phenotypic or environmental pattern. The mega-environment concept has been used to describe global target environments (Braun, Rajaram, & Ginkel, 1996; Crespo-Herrera et al., 2017; Rajaram, van Ginkel, & Fischer, 1995). If mega-environments are considered to be environmental factors, then it is valid to use the Eberhart and Russell method as an estimate of stability within mega-environmental groups.

In unbalanced datasets, there are two problems with traditional linear regression approaches to stability analyses such as the Finlay–Wilkinson regression. Environmental effects may be determined with bias when different genotypes are grown and then used to compare environments. In addition, a similar problem of bias arises when using genotype means to estimate environmental effects and then using those effects to determine genotype stability parameters. Bias in the estimation of fixed genotype and environment effects was solved by using a Bayesian approach. Using this approach to the Finlay–Wilkinson regression, genotypes and environments are treated as random rather than fixed variables, incorporating estimates of genotypic and environmental covariance and experimental error into the estimates for genotype stability and environment indices (Lian & de los Campos, 2016; Su et al., 2006). The additive main effects and multiplicative interaction model (AMMI) and the genotype main effects and genotype × environment interaction model (GGE) used principal components (PCs) to discern patterns in the G × E interaction (Gauch, 1992; Kempton, 1984). Graphical representations of PC results provided an instinctual way to understand G × E interactions (Yan, Hunt, Sheng, & Szlavnics, 2000).

These statistical methods are useful for comparing how the FN of specific genotypes respond to the environment because low FN and high α-amylase can result from genetic susceptibility to two environmental problems, PHS and the developmental defect LMA (reviewed by Mares & Mrva, 2014). The FN test is used to measure the presence of α-amylase in wheat. In LMA, α-amylase is induced in response to cold or high temperature shock during grain filling (approximately 24–28 d past anthesis). Preharvest sprouting, the

germination of mature grain on the mother plant when cool, rainy conditions occur before harvest, is also associated with elevated α-amylase. Although advanced cases of PHS obviously result in the emergence of a seedling root, mildly sprouted grain and LMA-affected grain cannot be detected by the naked eye. Alpha-amylase can be produced early in the germination process, before seedling growth is obvious (Lunn, Major, Kettlewell, & Scott, 2001). The performance of varieties in an LMA environment is not predictive of how they will perform in a PHS environment. Moreover, ANOVA of the entire FN dataset suggested that genetics accounted for a very low proportion of the total variability (Garland Campbell, unpublished data, 2017). This is likely due to the G × E interaction and the fact that PHS and LMA have different mechanisms of resistance (Mares & Mrva, 2008). With good experimental design and analysis to control and identify environmental effects, this is a trait that has genetic variation that can be identified and manipulated.

This study aims to utilize joint regression and biplot analyses, as described, (i) to interpret the G × E interaction contributing to the variation in FN, (ii) to determine the relative importance of sources of variation in the FN data, and (iii) to provide farmers and breeders with information on which to base their variety selections. By using these methods to understand FN, we provide the basis to apply it to other unbalanced, extremely variable datasets.

2 | MATERIALS AND METHODS

2.1 | Plant material and cultural data

One hundred and thirty-three unique soft white winter wheat genotypes consisting of released and unreleased cultivars and breeding lines from the University of Idaho, WSU, Oregon State University, Limagrain, WestBred/Monsanto, USDA-ARS, and Syngenta were evaluated in 2013, 2014, and 2016. The genotypes were grown by the WSU Cereal Variety Testing program as described by Guy, Jitkov, Lauver, and Horton (2013, 2014) and Higginbotham, Jitkov, and Horton (2016). Briefly, the field trials were conducted as winter cropping systems (planted in fall and harvested in late summer) at 17 locations in 2013 and 18 locations each in 2014 and 2016. These were years in which Pacific Northwest farmers experienced economic losses due to low FN. Field locations (Supplemental Figure S1) were spread across different climatic regions of Washington State based on average annual rainfall. Varieties were grown under rainfed conditions except for two irrigated locations, Moses Lake and Pasco, in all 3 yr. At each location, grain yield was evaluated in an α-lattice design with three replications. Field plot size ranged between 16 and 29 m^2 with sowing density from 56 to 106 kg ha^{-1} and row spacing of 15 to 38 cm, depending on the location. All locations were planted according to standard recommended agronomic practices for winter wheat in the inland Northwest (http://smallgrains.wsu.edu). All plots were planted with small plot planters and harvested with a Wintersteiger plot combine. Trial location and management details, plus yield and agronomic data including heading and harvest dates, and location notes are available at http://smallgrains.wsu.edu.

2.2 | Falling number test

The FN test detects the digestion of starch in wheat meal based on the resulting decrease in starch gelling capacity, by measuring the time it takes (in seconds) for a stirrer to fall through a heated flour–water slurry. The FN values were determined with the Hagberg–Perten FN Apparatus 1800 (Perten Instruments) using the ICC 107/1 method (1995) with minor modifications as described in Martinez et al. (2018). Grain was harvested from each plot, and 40-g samples were milled into whole meal on an Udy sample mill (Udy Corporation). A whole grain meal sample corresponding to 7 g at 14% moisture was used to run one FN test per sample. The number of plots sampled per genotype and environment varied between one and three biological replicates, such that there were three replicates per genotype in 14 environments, two in 36, and one in three environments. The number of replications tested in each environment was dictated by the resources available at the time of data collection (Supplemental Table S1). The nature of the sampling was such that the α-lattice design of the field was not maintained in the testing of FNs, so replication was the only factor fitted into the models described below. For all instances where there was only one replication, that one replication was used in the analysis.

2.3 | Mega-environment characterization

Although mega-environments are generally defined as areas with similar biotic and abiotic stresses, this study adapted the concept to categorize the trial mega-environments likely to cause low FN based on weather data suggesting conditions associated with LMA and PHS. Three mega-environments were identified:

1. No Event: trials where no weather event associated with PHS or LMA was observed,
2. PHS: trials where precipitation occurred when grain was mature (up to 2 wk prior to harvest).
3. LMA: trials experiencing heat shock or cold shock during late grain maturation between 28 and 35 d after heading, approximately 23 to 30 d past anthesis.

When both LMA and PHS weather conditions were observed, the trial was categorized as PHS. Weather data including maximum and minimum daily temperature (°C) and daily precipitation (cm) from WSU's agricultural weather station, AgWeatherNet (www.agweathernet.com), were used in combination with heading date and harvest date to categorize each trial into one of the three mega-environments. For example, Creston 2016 was determined to be affected by LMA because the temperature dipped and then spiked abruptly during the susceptible window of LMA induction, 28 to 35 d after heading (Supplemental Figure S2).

2.4 | Mixed linear model and heritability estimates

The full dataset was analyzed initially using a standard linear model as

$$Y_{ijkl} = \mu + V_i + E_j + \mathrm{VE}_{ij} + R(E)_{k(i)} + e_{ijk} \quad (1)$$

where Y was the plot FN, μ was the mean (intercept), V_i was the genotype effect, E_j was the environment (location × year) effect, VE_{ij} was the environment × genotype interaction, $R(E)_{k(i)}$ was the replication effect within each environment, and e was the residual variance (Supplemental Table S2). All effects were considered random. Best linear unbiased predictors (BLUPs) were obtained for all random effects, further serving to account for the unbalanced nature of the data (Piepho & Mohring, 2007).

Then, an extended model was fit that incorporated mega-environment as a fixed effect:

$$Y_{ijkl} = \mu + M + V_i + E_j + \mathrm{VE}_{ij} + \mathrm{ME}_i + R(E)_{k(i)} + e_{ijk} \quad (2)$$

where ME_i was the environment × mega-environment interaction. All components were considered to be random as above, except M, which was the fixed effect of the mega-environment. A Wald test was used to test the significance of mega-environment as an explanatory variable (Kenward & Roger, 1997). Predictions were obtained for genotype response within each mega-environment, and pairwise differences were calculated.

Variance was estimated for all components within genotype and environment using the standard model (Eq. [1]), and the extended model was used to incorporate mega-environment as a fixed effect (Eq. [2]). Packages used for these analyses included the *asremlPlus* package in *ASReml-R*, *lme4*, *agricolae*, *fw*, *coda*, and *gge* in R version 3.5.1.

Heritability was estimated based on BLUPs using the generalized method of Cullis, Smith, and Coombes (2006). This estimate uses the concept of "effective error variance" introduced by Cochran and Cox (1957) and was computed as

$$\bar{H}_c^2 = 1 - \frac{\bar{v}_{\mathrm{BLUP}}}{2\sigma_g^2} \quad (3)$$

where $\bar{v}_{\mathrm{BLUP}}$ was the mean variance of a difference between two BLUPs, and σ_g^2 was the variance component attributed to the variety effect, both calculated from Eq. (1).

2.5 | Stability analysis

For this unbalanced dataset, the Bayesian method for the Finlay Wilkinson regression as proposed by Su et al. (2006) was used because it reduced potential biases. The model,

$$p(y|\theta) = \prod_{ijk} N\left(\mu + V_i + E_j + b_i E_j, \sigma_e^2\right) \quad (4)$$

where $p(y|\theta)$ was the conditional distribution of the data, given the parameters; θ represented the collection of unknowns ($\theta = \{\mu, V, b, E, \sigma_V^2, \sigma_b^2, \sigma_E^2, \sigma_e^2\}$), $1 + b_i$ was the expected change in performance of the ith variety per unit change in the environment effect, and σ_V^2, σ_b^2, σ_E^2, and σ_e^2 were the variance components for each parameter defined above. The arithmetic mean FN was calculated for each genotype per environment and used as the response variable in the model. The Gibbs sampler method available in the FW package was used to estimate features of the posterior distribution. Further details about the model are described in Lian and de los Campos (2016). From this model, the slope and intercept were taken as a measure for adaptability and for general performance, respectively (defined above as the Finlay–Wilkinson stability parameters Type 1 and Type 2). The residual variance of a genotype from the regression line, b_i, indicated how stable a genotype is (defined above as the Eberhart and Russell stability parameter, Type 3). Thus, a Bayesian analysis of the data produced estimates of Finlay–Wilkinson and Eberhart and Russell stability parameters for the combined dataset and within each mega-environment. The Finlay–Wilkinson model was used to analyze the entire dataset first, then with only environments characterized as LMA or PHS affected.

Although the Finlay–Wilkinson regression approach used staged fitting, the AMMI model provided joint estimates of its parameters by combining ANOVA and principal component analysis (PCA) (Crossa, 1991; Gollob, 1968). The AMMI model was used to estimate the variety effects in each environment. After the variety and environment effects were fitted, the PCA was used to fit multiplicative effects for G × E interaction. The AMMI model was

$$Y_{ij} = \mu + V_i + E_j + \sum_{k=1}^{n} \lambda_k \alpha_{ik} \gamma_{jk} + e_{ij} \quad (5)$$

where λ_k was the square root of the eigenvalue of the PCA axis k, α_{ik} and γ_{ik} were the PC scores for PCA axis k of the ith genotype and the jth environment, respectively. An F test at 0.05 probability calculated using the mean square of each PC was used to identify the PCA axes that were significant (Crossa, 1990; Purchase, 1997).

The AMMI stability value was calculated using Eq. (6), adapted from Purchase (1997).

$$\text{AMMI Stability Value} = \sqrt{\frac{\text{SSPC1}}{\text{SSPC2}}(\text{PC1})^2 + (\text{PC2})^2} \quad (6)$$

In this formula, SS was the sum of squares, PC1 was the interaction of PCA 1, and PC2 was the interaction of PCA 2. Higher AMMI's stability values indicated more stability.

Although the AMMI model produced another measure of stability across environments, the GGE biplot summarized the genotype and G × E interaction effects graphically (Kempton, 1984). The GGE biplot model was standardized using a scaling factor, s_j, which was the standard error in environment j:

$$\mathbf{P}_{ij} = \frac{y_{ij} - \mu - V_j}{s_j} = \frac{E_i + \text{VE}_{ij}}{s_j} \quad (7)$$

In this case, y_{ij} was defined by Eq. (1) and $\mathbf{P}_{ij}$ was a matrix that was subjected to PCA, which broke down a two-way table of genotypes and environments into respective PCs. The arithmetic mean FN was calculated for each genotype per environment and used as the response variable in the model. The R packages *gge* and *agridat* were used to produce the biplots (Wright & Laffont, 2018). The fit of the GGE biplot was diagnosed using patterns described in Yan, Kang, Ma, Woods, and Cornelius (2007), who explained that the adequacy of the biplot is reflected in the presence of clear patterns. The GGE biplot made with the complete dataset was complicated to analyze. High correlation was observed between LMA and PHS mega-environments, as evident in the angle between environmental vectors within a biplot (Yan & Tinker, 2006). Vector lengths of PHS mega-environments were longer compared with LMA mega-environments in the combined biplot, indicating that more effective selection for resistance to low FN can be done within PHS mega-environments. With that in mind, interpretations from a GGE biplot containing only PHS mega-environment data were reported here.

To compare the various stability statistics calculated for this dataset, the methods used in this study were compared by constructing a biplot of the first two components of a PCA of the variety stability estimates from each model.

3 | RESULTS

3.1 | Phenotypic data analysis

The distribution of FN data by location showed that those environments characterized as LMA or PHS had widespread incidence of low FN (Figure 1, Table 1). The range in FN values within environments was large (Supplemental Table S3), but when observed by mega-environment, the ranges were generally greater for the LMA and PHS environments than the No Event environments. Visual inspection revealed that the FN range in PHS environments seemed to extend lower than in the LMA environments, and pairwise comparisons of means between PHS and LMA mega-environments indicated that there was no significant difference between them ($p = 0.603$) (Table 2). Only a small increase, twofold, was observed in error variance among environments, demonstrating that the majority of environments had a similar amount of unexplained error and that differences in coefficient of variation were primarily due to differential genetic variability within environments. The dataset was highly unbalanced, with pairs of environments sharing between 4 and 60 varieties, an average of 24, and a median of 19 varieties (Supplemental Figure S3). This low connectivity likely affected our calculation of variance for G × E interaction.

3.2 | Variance components

To examine the effects of two models (Eq. [1] and [2]) on the distribution of variation in these experiments, variance components were estimated, and their percentage of total variation was compared (Table 2). All variance components were highly significantly different from zero ($p < 0.001$), and for the extended model, the fixed effect, mega-environment, was also significant ($p < 0.001$). Most of the phenotypic variation was attributed to the differences among environments, representing almost half of the total phenotypic variance in the standard model. The genotypic variance component was the next most important and increased slightly from 18.2% in the standard model to 21.1% in the extended model. The reduction in magnitude of the environmental variance component in the extended model was due to the incorporation of some of the environmental variance into the mega-environment fixed effect. Variance due to the G × E interaction was lower in both models than the genotype and environment main effects but still represented 14 to 16% of the total phenotypic variance. Although the residual variance was slightly lower in the standard model (−5.8%), the generalized heritability improved from 0.55 in the standard to 0.64 in the extended model due to the incorporation of mega-environment effects in the extended model.

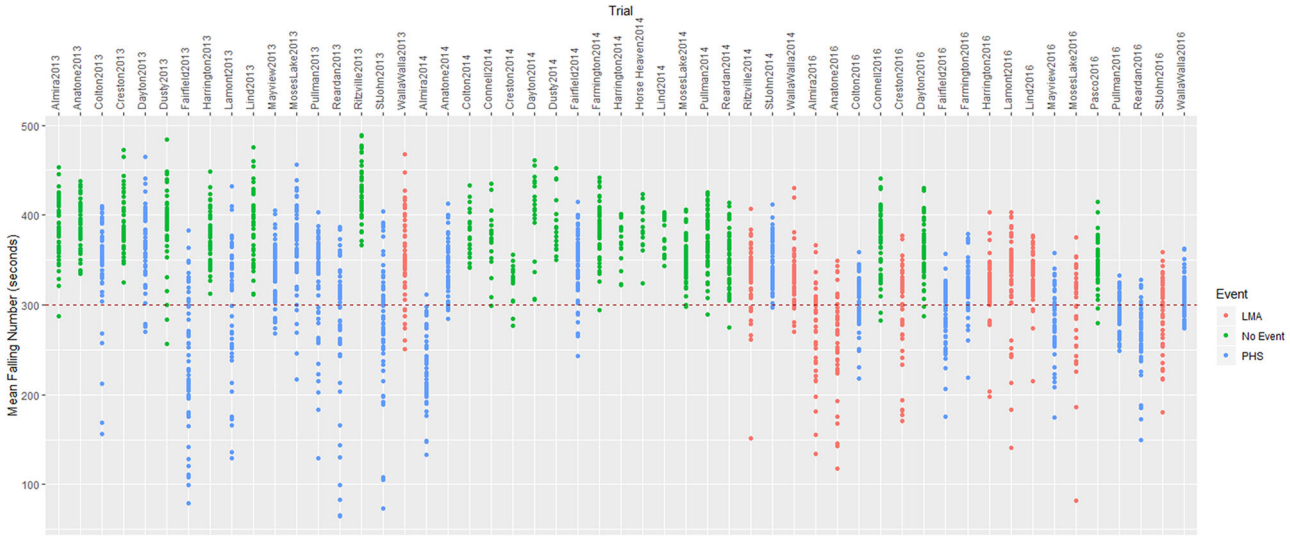

FIGURE 1 Distribution of falling number data by environment, colored by mega-environment (red = late-maturity α-amylase [LMA], blue = preharvest sprouting [PHS], and green = No Event). The red dashed line indicates 300 s, the point below which discounts on grain are usually incurred

TABLE 1 Decomposition of Washington State University Cereal Variety Testing Program winter wheat trials data based on mega-environment and incidence of low falling number (FN)

Year	No. of locations	No. of varieties	No. of locations with FN below 300 s	No. of locations with PHS[a]	No. of locations with LMA[b]
2013	17	65	12	9	1
2014	18	72	10	4	2
2016	18	78	18	7	8

[a]PHS, preharvest sprouting.
[b]LMA, late-maturity α-amylase.

3.3 | Finlay–Wilkinson regression

A Finlay–Wilkinson regression was performed to examine G × E interaction by assessing the performance of individual genotypes in relation to environment effects. This approach fits the best linear regression of the performance of a variety on the mean environment value over all genotypes in that environment. The slope of the line is used to approximate genotype-specific environmental stability. Variance components obtained from the Finlay–Wilkinson analysis (Eq. [4]) revealed that the majority of phenotypic variation was due to residual (error) regardless of whether the dataset was combined or categorized into mega-environment (Supplemental Table S4). The credibility intervals (analogous to confidence intervals) and mean for the genotypic and G × E variances were larger within the PHS and LMA mega-environments than in the No Event or combined dataset because environmental conditions in the PHS and LMA mega-environments resulted in separation of varieties for response to the FN test. When the varieties with the three highest (WB 456, Coda, and WB 1376CLP) and lowest (4J071246-1C, WA 8251, and WA 8226) BLUP values (y intercept in the graphs in Figure 2) for genotype effect were compared across the entire set of environments (Figure 2a), the range in slope was greater among the varieties with the highest BLUP values. Two of those varieties, WB 456 and WB 1376CLP, had a slope of less than one, indicating low genotypic sensitivity or better ability to maintain high FN even under LMA and/or PHS conditions. However, the other variety, Coda, had a slope of greater than one. The three varieties with the lowest BLUP values had similar slopes of greater than one. In contrast, when the No Event data were removed from the analyses (Figure 2b), the three varieties with the highest BLUP values (KWS 040, WB 456, and WB 1376CLP) had slopes less than one, and the three varieties with the lowest BLUP values (WA 8251, WA8226, and 4J071246-1) had slopes much greater than one. Thus, varieties can be characterized by their overall genetic effect (high or low BLUP values, y intercept) and by their responsiveness (slope).

3.4 | Eberhart and Russell stability

The Eberhart and Russell method was used to estimate stability, values that showed a broad range across mega-

TABLE 2 Mixed model analysis of 2460 falling number (FN) data points (estimated variance components from Equations [1] and [2]). The percentage of phenotypic variation explained by the model is given (% of phenotypic)

Term	Decomposition	Standard	Extended[a]	Standard	Extended[a]
		s^2		% of phenotypic	
Variety	variety	789.8***	788.8***	18.2	21.1
Mega-environment	PHS	–	−69.01***[b]	–	–
	LMA	–	−62.91[b]	–	–
Environment	Environment	1809.2***	784.7***	41.7	23.9
	Environment(rep)[c]	278.5***	271.8***	6.4	8.4
Variety.mega-environment[d]	Variety.mega-environment	–	147.7***	–	4.5
Variety.environment[e]	variety.environment	614.5***	531.5***	14.2	16.3
Error	error	845.3***	841.0***	19.5	25.8

***Significant at the 0.001 probability level.
[a]Pairwise difference from the No Event mega-environment.
[b]No significant difference between preharvest sprouting (PHS) and late-maturity amylase (LMA) mega-environments.
[c]Replication effect within each environment.
[d]Mega-environment × variety interaction.
[e]Environment × variety interaction.

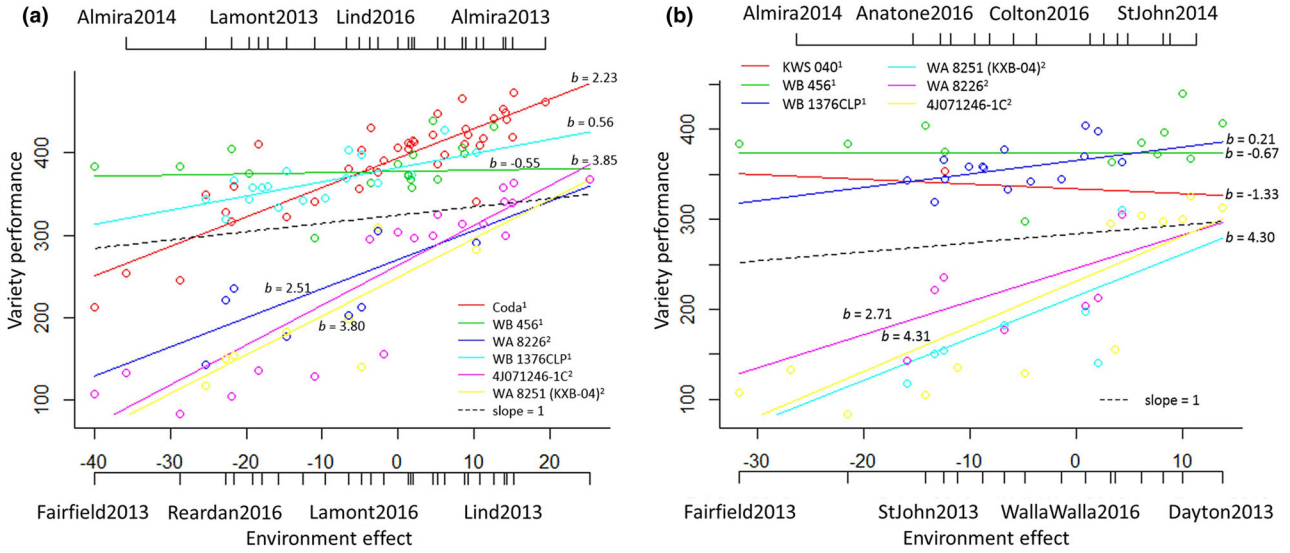

FIGURE 2 The performance of six varieties selected based on genotype values from Eq. [1] (a) with the whole dataset and (b) with the No Event locations removed. The six varieties are composed of the three highest and lowest (a) best linear unbiased predictor (BLUP) and (b) best linear unbiased estimate (BLUE) values to reduce complexity in making inferences about individual varieties. Fitted values are represented by the lines and cell means of genotype, and environment combinations are represented by the circles, each color indicating a variety. The dashed line has a slope equal to one. The slope (*b*) of each variety is given

environments (Supplemental Table S5). The greater the value, the less stable a variety was. The largest mean, standard deviation, and range was observed for PHS environments, whereas the No Event environments had the smallest values for all three measurements. This result was unsurprising because the distribution of FN in PHS environments was greater than in both other mega-environments (Figure 1). Genotypic values, BLUPs, were not predictive of the Eberhart and Russell stability or the *y* intercept in the Finlay–Wilkinson regression. When the three mega-environments were combined, the top three varieties according to BLUPs did have lower, and therefore more stable, Eberhart and Russell stability values. That same trend was not observed in the PHS mega-environment datasets. In fact, the two most stable varieties in the PHS mega-environment had the lowest genotypic values. This suggests that these Type 3 stability values do not also reflect the overall performance of a variety.

Although the measures of stability derived from the Finlay–Wilkinson regression provide some insight, its utility is limited to only a linear fit of the complex G × E interaction. If the

residual variance of the model is large, then the linear approximation does not explain most of the G × E interaction.

3.5 | AMMI analysis

The AMMI analysis was used to take our exploration of G × E interactions one step further by visualizing these interactions in a biplot and to provide a different estimate of stability. The first axis of the AMMI biplot can be compared to the Finlay–Wilkinson regression because it describes the largest amount of environment interaction. In contrast with the Finlay–Wilkinson regression, the AMMI model is then able to incorporate the next best-fitting effect as the second axis in the biplot, which accounts for the largest amount of genetic interaction. All effects in the model (environments, replications, genotypes, and G × E interaction) were highly significant ($p < 0.001$) (Table 3). There was not an expectation that this model would outperform the variance component model (Eq. [1]); we were more interested in visualizing the nonlinear G × E patterns within the PC piece of the AMMI model. The interaction matrix was decomposed by extracting 52 PCs, 32 of which were significant ($p < 0.05$). The arithmetic mean FN scores for varieties and for environments were plotted against the PC1 scores, accounting for 30.4% of the interaction sum of squares (Figure 3). The varieties and environments on the right side of the origin showed above average FN values. Varieties with the highest average FN were WB 456, WB 1376CLP, and Coda, whereas the varieties with the lowest average FN were WA 8251 (KXB-04), 4J071246-1C, and WA 8226. This corresponded precisely with the rankings based on BLUPs. The AMMI plot allowed us to conclude that of these six, 4J071246-1, Coda, and WB 456 showed a high interaction with the environment vs. the other three, which had PC1 scores near zero, indicating low environment interaction. More information about the G × E interaction can be extracted by plotting PC1 against PC2, accounting for an additional 8.3% of the interaction (Supplemental Figure S4a). Varieties and environments closest to the origin were more variable and difficult to place, whereas those near the external parts of the graph showed positive interaction. When zoomed in closer to the origin (Supplemental Figure S4b), trends are more visible. CuriosityCL+ had the highest FN value in Mayview 2016, and Puma had the highest FN values in Pullman 2013. Both of these environments were PHS mega-environments yet were negatively correlated with each other as they fell in opposing quadrants.

The AMMI stability value ranged from 0.01 to 20.63 with an overall mean of 3.07 across the 133 varieties. The lower AMMI stability value indicated higher stability. The most unstable varieties were Bruehl (15.9), ARS-Selbu (19.7), and Xerpha (20.6), whereas the most stable were WA 8201 (0.01), WA 8203 (0.02), and WA 8232 (0.03). The AMMI stability value for the six varieties selected as the top and bottom three based on BLUPs (Table 2) revealed little correlation, meaning that BLUPs for FN are not necessarily indicative of stability according to the AMMI model.

A wide range of correlation values was observed between stability estimates previously discussed: BLUP, variety mean FN, AMMI stability value, Finlay–Wilkinson regression coefficient, and Eberhart and Russell stability based on either the combined or mega-environment datasets (Supplemental Figure S5). The BLUP and variety mean FN were most closely correlated ($p < 0.0005$), but when G × E interaction parameters were incorporated into measures of stability, the correlations to BLUP and variety mean started to decrease. When Finlay–Wilkinson and Eberhart and Russell stability values were calculated using the entire dataset, they did correlate significantly with BLUP and variety mean FN ($p < 0.0005$), but once split by mega-environment, the relationship shifted drastically. Across two of the mega-environments, PHS and LMA, correlations between Finlay–Wilkinson and Eberhart and Russell stability values were significant and positive (0.79 and 0.84, respectively), whereas correlations between Finlay–Wilkinson stability and BLUP (−0.89 and −0.72) and Eberhart and Russell stability and BLUP (−0.66 and −0.67) were significant and negative. The same was not true of the No Event mega-environments. Finlay–Wilkinson and Eberhart and Russell stability measures within the No Event mega-environments were positively correlated with BLUP (0.64 and 0.27), and only the correlation between Finlay–Wilkinson and BLUP was significant.

Although the AMMI model gives an additional measure of stability related to G × E interaction that is missing from the Finlay–Wilkinson regression, it does have some limitations. The AMMI model only explains the G × E interaction well if the first few terms of the PCA represent true structure. If the first few terms are very similar, then it is difficult to identify patterns.

3.6 | GGE biplots

We used GGE biplots to examine the combined genotype plus G × E interactions on a single biplot. This contrasts with the AMMI biplot, which only incorporates the G × E interactions. The GGE biplot allows the user to quickly identify optimal varieties for each environment by looking at their relative length and angles. It is also possible to identify the best-performing variety overall or in each environment. However, it should be noted that all biplots are limited by the fact that they are a visual tool, not a statistical test. If most of the variation is not explained by the first two PCs, then the picture can be misleading.

Blue concentric circles center around a point known as the ideal test environment, equal to the point on the

TABLE 3 ANOVA table for the additive main effects and multiplicative interaction model (AMMI) model applied to entire dataset

Term	Decomposition	*F* value	df	GEI SS (cumulative)[a] %
Environment	environment	26.0***	52	–
Replication (rep)	Environment(rep)	7.0***	188	–
Variety	variety	31.6***	132	–
Variety.environment[b]	variety.environment	2.1***	2266	–
	PC1	48.1***	183	30.4
	PC2	13.3***	181	8.3 (38.7)
	PC3	9.6***	179	5.9 (44.6)
	…[c]	–	–	–
	PC33	1.2	119	0.5 (96.0)
Error	error	***	1559	–

***Significant at the 0.001 probability level.

[a]GEI, genotype × environment interaction; SS, sum of squares.

[b]Environment × variety interaction.

[c]Principal components PC3–PC32 are all significant at $p < 0.05$.

average environment axis indicated by the dashed line that is the length of the longest environment vector away from the origin (Figure 4a). Based on this, four environments were determined to be ideal test environments for FN—Colton, Fairfield, Mayview, and Farmington 2016, all characterized as PHS mega-environments. The performance of a variety was estimated to be better than average in that environment if the angle between its vector and the environment was <90°. Two varieties are highlighted in Figure 4a: Bruehl (blue dashed line) performed poorly overall and Masami (red dashed line) performed better than average in PHS environments.

The six varieties selected based on BLUPs in the previous analyses and an additional 14 random varieties were used to make a 20-variety GGE biplot more manageable for visual inspection (Figure 4b). As was observed in the GGE biplot with all data (Supplemental Figure S6), LMA and PHS mega-environments were closely correlated to each other, and more loosely correlated to No Event mega-environments.

The black dashed line drawn through the PHS mega-environment, the axis for PHS, allowed for ranking of varieties based on performance in that mega-environment. Rankings were visualized more clearly with green solid lines perpendicular to the PHS axis connected to each variety. This was done for all mega-environments with the AMMI stability value for comparison (Supplemental Table S6). Rankings in the three mega-environments were significantly correlated with each other ($p < 0.005$), and with all other measures displayed in Supplemental Figure S5 except for the AMMI stability value and the Eberhart and Russell stability in PHS mega-environments (Supplemental Figure S7). When rankings were done within LMA and PHS mega-environments, the same varieties remained in the top and bottom 10, although the rank shifted slightly. Alternatively, there was no correspondence in variety ranking between No Event mega-environment, and the LMA and PHS entered mega-environment rankings. A which-won-where polygon connected the varieties furthest away from the origin with an equality line (gray solid). Red dashed lines perpendicular to each equality line and going through the origin split the biplot into six sectors. The equality line between WB 1376CLP and WB 456 indicate that WB 456 had higher FN in PHS and LMA mega-environments than WB 1376CLP. The PHS and LMA mega-environments fall into one sector, and No Event falls in an adjacent sector. WB 456 is the winner in the PHS and LMA sector, whereas Coda and WB 1376CLP are both winners in the No Event sector. WA 8251 was the worst performing variety overall according to the which-won-where polygon.

4 | DISCUSSION

This study has compared various statistical methods of analysis to determine the most suitable procedure to evaluate a complex trait measured using a highly unbalanced dataset. Assessing the suitability of these methods for estimating stability in such a trait can serve as an example for future use in other complex traits and crops.

Falling number is a complex trait because it depends on multiple genotypic, developmental, and environmental factors. Resistance to low FN can result from genetic resistance to PHS and to LMA. However, higher FN can also result from higher grain protein concentration (Ross, Flowers, Zemetra, & Kongraksawech, 2012). The G × E interaction is complicated by the fact that susceptibility to LMA and PHS are dependent on the timing of weather events relative to developmental events. Late-maturity α-amylase susceptibility is limited to a developmental window during grain maturation, such that

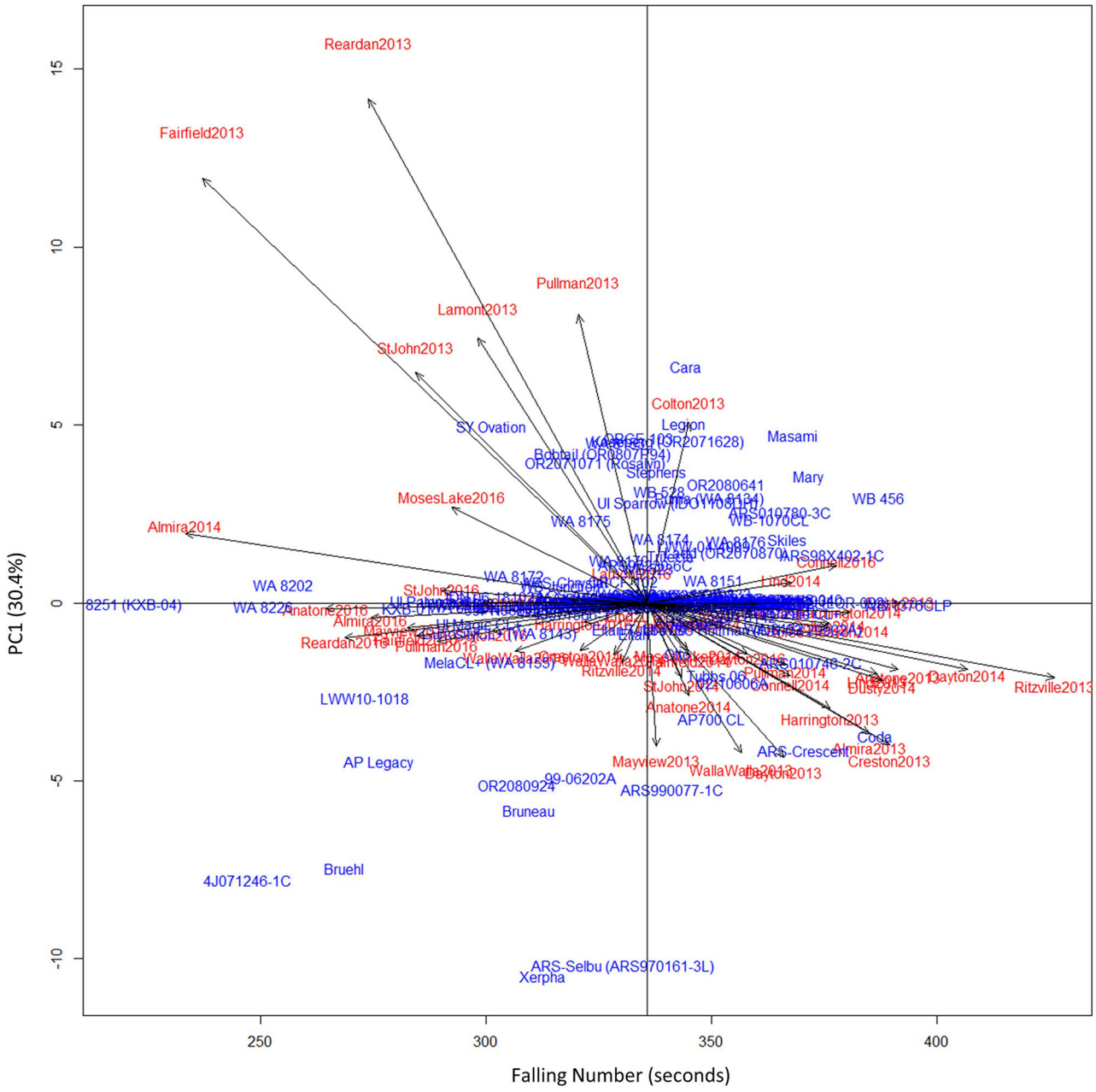

FIGURE 3 Biplot of the unadjusted mean falling number and the first principal component analysis (PCA) axis for interaction of 133 genotypes and 53 environments. Varieties are in blue and environments in red. The vertical lines represent the grand mean of the experiment

a temperature shock must occur between approximately 23 and 28 d past anthesis. Tolerance to PHS mainly results from higher seed dormancy, the inability of mature seeds to germinate under favorable conditions. Since dormancy is gradually lost after physiological maturity through dry after-ripening, differences in PHS resistance depend on how much time has passed since the grain first reached physiological maturity. Moreover, initial dormancy varies with environmental conditions during grain maturation. For example, higher dormancy can result from drier or cooler conditions during grain maturation (Biddulph, Plummer, Setter, & Mares, 2007, 2008; Nakamura et al., 2011). Our ability to account for all of these environmental impacts was limited by the nature of the data available. The physiological data collected were limited to heading date and harvest date, so extrapolating using only those agronomic indicators paired with basic weather data limited precision in defining mega-environments. Moreover, some PHS environments may also have experienced LMA. This may explain why most of the phenotypic variation was explained by the residual (error) in the Finlay–Wilkinson regression, and why residual errors were considerable in other analyses.

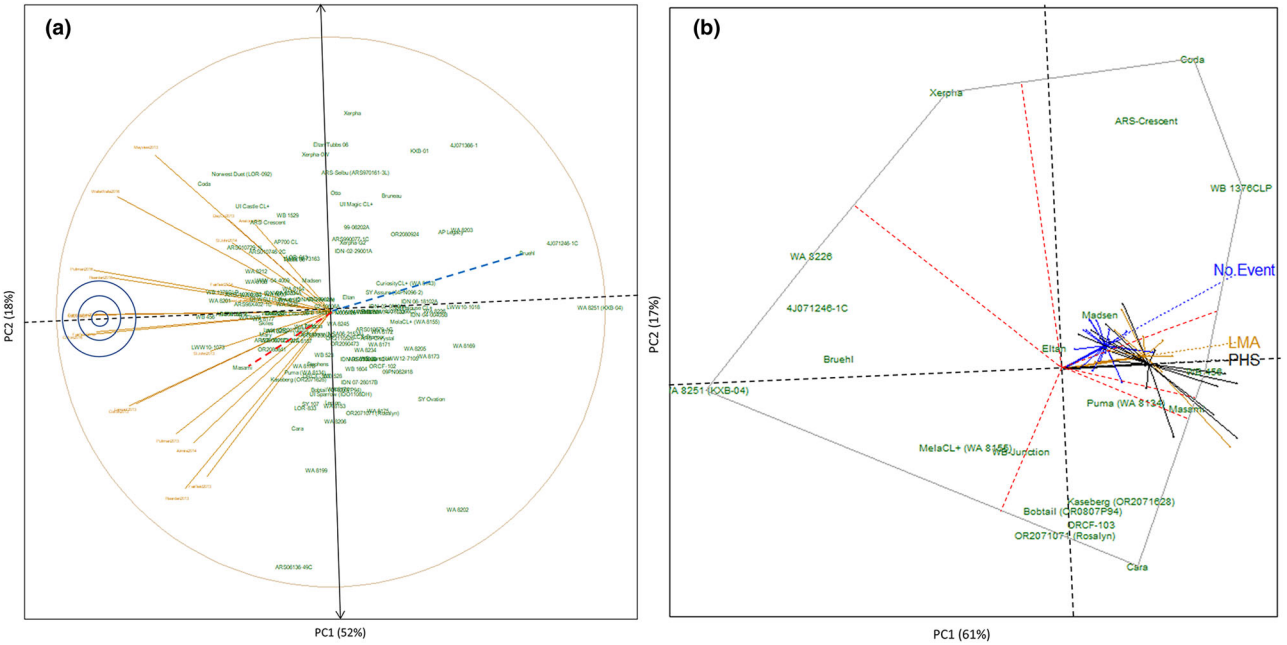

FIGURE 4 (a) Genotypic main effect plus genotype × environment interaction (GGE) biplot showing the ideal test environment for the falling number (FN) in preharvest sprouting (PHS) mega-environment dataset (represented by the center of concentric circles), and two variety performance vectors (Bruehl is blue, Masami is red). The gold unit circle serves as a point to where environment vectors are perfectly represented in the two-dimensional plane. (b) GGE biplot for FN for 20 varieties (6 selected based on BLUPs and 14 selected randomly). Grouping in three mega-environments are shown (blue vectors are No Event [NE], gold vectors are late-maturity α-amylase [LMA], and black vectors are PHS), and sectors (red dashed line) are defined by which-won-where polygon (gray line) made by connecting equality lines between winning varieties at vertices. The black dashed line going through PHS allows ranking of varieties based on performance in PHS environments (green solid lines)

Heritability as measured by the "effective error variance" was greatly improved by using mega-environment as a fixed effect in the model. Piepho and Mohring (2007) found this measure of response to selection using BLUPs to work "remarkably well" for very unbalanced data. This indicates that response to selection within a mega-environment is much higher than it is without first splitting environments into mega-environments. This suggests that selection for higher FN within a breeding program may be more effective if done within trials where LMA or PHS triggering conditions are present.

In this study, Type 1 stability was desirable. Type 1 stability reflects homeostasis, where a variety is less sensitive to environmental factors. This is not always desirable in the breeding program, especially in regard to yield, but with FN, the best varieties will have some degree of homeostasis across environments, both under PHS and LMA. Most of the most stable varieties according to AMMI's stability value do not have the highest mean FN or rank consistently stable based on the Finlay–Wilkinson or Eberhart and Russell methods. However, those varieties with the worst AMMI's stability value ranking were also among those with the lowest mean FN and least stability according to the Finlay–Wilkinson and Eberhart and Russell method, in most cases. This indicates that AMMI's stability value may be best used for culling bad varieties. This finding is consistent with the suggestion of Sabaghnia, Sabaghpour, and Dehghani (2008) that AMMI's stability value is a useful tool for selecting both for yield and stability.

The principal challenge of this study was to recommend the most appropriate methods or combination of methods to be used in selecting superior wheat varieties for breeder and farmer use. The relationships between variety stability estimates were compared on a biplot (Figure 5). The first PC (PC1), explaining 36.3% of the variation, separated the overall genotype performance estimates, FN and BLUP, from all estimates of stability ranging from Type 2 stability (PC1 < 0) to Type 1 stability (PC1 > 0). As discussed by Flores, Moreno, and Cubero (1998), Type 2 stability was traditionally desired by agronomists who define a stable genotype as one that corresponds to the quality of the environment. In contrast, Type 1 stability is considered desirable by plant breeders who define a stable genotype as one that performs consistently under changing environments. Group 1, variety mean FN and BLUP, fall clearly into the Type 2 category of stability, whereas Groups 3 and 4 fall clearly into the Type 1 category. Groups 3 and 4 contain Finlay–Wilkinson, Eberhart and Russell, and AMMI's stability values calculated from all datasets including PHS and LMA mega-environment trials. Interestingly, Finlay–Wilkinson and Eberhart and Russell stability estimates for the No Event mega-environments

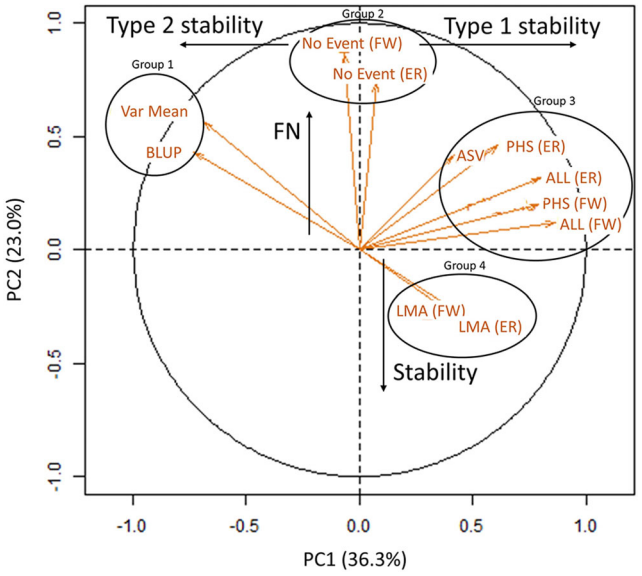

FIGURE 5 Plot of two principal components (PC1 vs. PC2) among ranks by best linear unbiased predictor (BLUP), variety mean falling number (Var Mean), AMMI's stability value (ASV), Eberhart and Russell stability (ER), and Finlay–Wilkinson stability (FW) of 133 winter wheat varieties. Each vector represents one of the aforementioned measures of stability and/or performance. The abbreviation PHS refers to preharvest sprouting and LMA refers to late-maturity α-amylase

fall almost directly on the vertical axis. This indicates that data from the No Event mega-environments alone provided no information regarding Type 1 or Type 2 stability. Thus, environments that do not trigger a PHS or LMA event cannot predict relative variety performance in environments that do trigger PHS or LMA, even though significant genetic variation for FN existed in the No Event environments.

Group 4, containing estimates from LMA datasets only, was a predictor of stability but not of mean FN, and the vector lengths were shorter in that group, indicating that these data were variable. This suggests that it is more effective to select for high FN in PHS than LMA environments. In addition, PHS seems to cause a lower FN in susceptible varieties than LMA in general (Figure 1; Mares & Mrva, 2014). Although it may be difficult to breed for LMA resistance, breeders must continue to select for this trait because even a small decrease in FN can result in steep discounts for wheat producers (Steber, 2017). The BLUP and genotype means were directly negatively correlated with the LMA environments, indicating that they may prove useful for selecting for higher general FN in all environments. Based on our biplot analysis, if the objective is to select varieties with stable and high FN, then Group 3 methods will be the best choice. Of these five methods, the ones falling near the center of the quadrant (Eberhart and Russell in PHS and the AMMI stability value) will maximize accuracy by selecting varieties with Type 1 stability and high FN or, alternatively, culling varieties that are unstable.

The AMMI and GGE biplots allow the researcher to graphically visualize phenotypic performance and interactions. In contrast with AMMI biplots, which explain solely G × E interaction, GGE biplots approximate the combined genetic and G × E interaction effects of varieties. Furthermore, the GGE biplot evaluated environment utility and ranking of varieties by mega-environment. The subtle difference in ranking of varieties in LMA vs. PHS mega-environments agreed with previous reports that different genes independently contribute PHS and/or LMA resistance (Mares & Mrva, 2014). Using a combination of GGE ranking and Eberhart and Russell stability in low-FN environments and AMMI's stability value provides a relatively unique measure of FN value and stability. The main limitation of these biplots is that it is difficult to examine many varieties in a single plot. It may be necessary to first filter the data by choosing to look at varieties with the highest or lowest BLUPs. When we applied these combined selection methods to our dataset, the clear winner was variety, WB 1376CLP, whereas the clear loser was variety 4J071246-1C.

Stability and absolute performance, as measured by BLUPs of FN, are not the same thing. For example, one variety may stably perform poorly, whereas another may perform poorly only with increasing environmental pressure. Even so, we have been able to use a combination of BLUPs, mega-environment distinction, and PCA to identify the best- and worst-performing varieties for FN in these trials. Methods that provide a measure of phenotypic stability, like Finlay–Wilkinson and Eberhart and Russell analyses, build on mixed model theory and can be expanded to provide a way to evaluate genotype response to specific increasing environmental pressure such as weather conditions with higher incidence of PHS-inducing rain or a greater number or magnitude of LMA-inducing temperature fluctuations.

Future work will need to examine whether it is possible to model multiple factors simultaneously to provide farmers with a statistically valid way to rank varieties for complex traits like FN. This could include using the two-stage multienvironment trial analysis described by Cullis, Thomson, Fisher, Gilmour, and Thompson, 1996a, 1996b) or the factor analytic model and the prediction components described by Smith and Cullis (2018), but these will only be effective with large datasets, and currently, software for analysis of factor analytic models are not open source.

Results from GGE biplot and Finlay–Wilkinson regression are easy to interpret and can be used when a moderate amount of G × E data have been collected even without replicated trials. Plant breeders often plant large single-replication nurseries for initial yield trials. The models that we explored can be used to obtain information about G × E interactions from single-replication breeding trials. Second, these results confirmed the fact that locations where FN was 300 or greater

were not useful in making decisions regarding varieties' low FN performance.

Frequently, researchers are trying to make better sense of data after it has been collected or obtained. The methods explored in this paper are useful in analyzing G × E performance from experiments that often are out of the control of the investigator in terms of sampling and execution from the beginning. This is a challenge, but it is still important to find a valid way to analyze the data and make it interpretable to farmers and the public.

AUTHOR CONTRIBUTIONS

K. Garland Campbell and C.M. Steber contributed to the experimental design. S.M. Sjoberg and K. Garland Campbell contributed to the statistical design. S.M. Sjoberg conducted the statistical analysis. The paper was written by S.M. Sjoberg, C.M. Steber, A.H. Carter, and K. Garland Campbell. Funding was provided by C.M. Steber, K. Garland Campbell, and A.H. Carter.

ACKNOWLEDGMENTS

The authors thank all collaborators who conducted the trials of WSU's Cereal Variety Testing Program, including Vadim Jitkov, Stephen Guy, and Ryan Higginbotham. The authors thank Tracy Harris, Rehana Parveen, and Xavian Thompson for expert technical assistance. Thanks are due to members of the Steber and Garland Campbell laboratory and Michael O. Pumphrey for helpful suggestions and comments on the manuscript. This study was supported by the Washington Grain Commission Projects 7599 and 7336 (to C.M. Steber, A.H. Carter, and Michael O. Pumphrey), by the USDA-ARS (to K. Garland Campbell and C.M. Steber), and by USDA Agriculture and Food Research Initiative (AFRI)–National Institute of Food and Agriculture (NIFA) Project 2015–05798 (to A.H. Carter and C.M. Steber).

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Stephanie M. Sjoberg https://orcid.org/0000-0003-0351-7518
Arron H. Carter https://orcid.org/0000-0002-8019-6554
Camille M. Steber https://orcid.org/0000-0001-6255-7670
Kimberly A. Garland-Campbell https://orcid.org/0000-0003-4747-3270

REFERENCES

Biddulph, T. B., Plummer, J. A., Setter, T. L., & Mares, D. J. (2007). Influence of high temperature and terminal moisture stress on dormancy in wheat (*Triticum aestivum* L.). *Field Crops Research*, *103*, 139–153. https://doi.org/10.1016/j.fcr.2007.05.005

Biddulph, T. B., Plummer, J. A., Setter, T. L., & Mares, D. J. (2008). Seasonal conditions influence dormancy and preharvest sprouting tolerance of wheat (*Triticum aestivum* L.) in the field. *Field Crops Research*, *107*, 116–128. https://doi.org/10.1016/j.fcr.2008.01.003

Braun, H. J., Rajaram, S., & Ginkel, M. (1996). CIMMYT's approach to breeding for wide adaptation. *Euphytica*, *92*, 175–183. https://doi.org/10.1007/BF00022843

Cochran, W. G., & Cox, G. M. (1957). *Experimental design*. New York: John Wiley & Sons.

Crespo-Herrera, L. A., Crossa, J., Huerta-Espino, J., Autrique, E., Mondal, S., Velu, G., … Singh, R. P. (2017). Genetic yield gains In CIMMYT's International elite spring wheat yield trials by modeling the genotype × environment interaction. *Crop Science*, *57*, 789–801. https://doi.org/10.2135/cropsci2016.06.0553

Crossa, J. (1990). Additive main effect and multiplicative interaction analysis of two international maize cultivar trials. *Crop Science*, *30*, 493–500. https://doi.org/10.2135/cropsci1990.0011183X003000030003x

Crossa, J. (1991). AMMI adjustment for statistical analysis of an international wheat yield trial. *Theoretical and Applied Genetics*, *81*, 27–37. https://doi.org/10.1007/BF00226108

Cullis, B. R., Smith, A. B., & Coombes, N. E. (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

Cullis, B. R., Thomson, F. M., Fisher, J. A., Gilmour, A. R., & Thompson, R. (1996a). The analysis of the NSW wheat variety database. I. Modelling trial error variance. *Theoretical and Applied Genetics*, *92*, 21–27. https://doi.org/10.1007/BF00222947

Cullis, B. R., Thomson, F. M., Fisher, J. A., Gilmour, A. R., & Thompson, R. (1996b). The analysis of the NSW wheat variety database. II. Variance component estimation. *Theoretical and Applied Genetics*, *92*, 28–39. https://doi.org/10.1007/BF00222948

Eberhart, S. A., & Russell, W. A. (1966). Stability parameters for comparing varieties. *Crop Science*, *6*, 36–40. https://doi.org/10.2135/cropsci1966.0011183X000600010011x

Farrand, E. A. (1964). Flour properties in relation to the modern bread processes in the United Kingdom, with special reference to alpha-amylase and starch damage. *Cereal Chemistry*, *41*, 98–111.

Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, *14*, 742–754. https://doi.org/10.1071/AR9630742

Flores, F., Moreno, M. T., & Cubero, J. I. (1998). A comparison of univariate and multivariate methods to analyze G×E interaction. *Field Crops Research*, *56*, 271–286. https://doi.org/10.1016/S0378-4290(97)00095-6

Gauch, H. G. (1992). *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier, Amsterdam, the Netherlands:.

Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, *33*, 73–115. https://doi.org/10.1007/BF02289676

Guy, S. O., Jitkov, V. A., Lauver, M. A., & Horton, A. (2013). Cereal and grain legume variety performance trials 2013. Rep. 225. Dep. Crop Soil Sci., Washington State Univ., Pullman, WA.

Guy, S. O., Jitkov, V. A., Lauver, M. A., Horton, A., & Higginbotham, R. (2014). Cereal and grain legume variety performance trials 2014. Rep. 220. Dep. Crop Soil Sci., Washington State Univ., Pullman, WA.

Higginbotham, R., Jitkov, V., & Horton, A. (2016). Cereal variety performance trials 2016. Rep. 190. Dep. Crop Soil Sci., Washington State Univ., Pullman, WA.

Kempton, R. A. (1984). The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science*, *103*, 123–135. https://doi.org/10.1017/S0021859600043392

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*, 983–997. https://doi.org/10.2307/2533558

Kruger, J. E., and Lineback, D. R.Carbohydrate-degrading enzymes in cereals. In J. E. Kruger (Ed.), et al. (1987). *Enzymes and their role in cereal technology* (pp. 117–139). St. Paul, MN: Am. Assoc. Cereal Chem.

Lian, L., & de los Campos, G. (2016). FW: An R package for Finlay–Wilkinson regression that incorporates genomic/pedigree information and covariance structures between environments. *G3: Genes, Genomes, Genetics*, *6*, 589–597. https://doi.org/10.1534/g3.115.026328

Lin, C. S., Binns, M. R., & Lefkovitch, L. P. (1986). Stability analysis: Where do we stand? *Crop Science*, *26*, 894–900. https://doi.org/10.2135/cropsci1986.0011183X002600050012x

Lunn, G. D., Major, B. J., Kettlewell, P. S., & Scott, R. K. (2001). Mechanisms leading to excess alpha-amylase activity in wheat (*Triticum aestivum* L.) grain in the U.K. *Journal of Cereal Science*, *33*, 313–329. https://doi.org/10.1006/jcrs.2001.0369

Mares, D., & Mrva, K. (2008). Late-maturity α-amylase: Low falling number in wheat in the absence of preharvest sprouting. *Journal of Cereal Science*, *47*, 6–17. https://doi.org/10.1016/j.jcs.2007.01.005

Mares, D. J., & Mrva, K. (2014). Preharvest sprouting and late maturity alpha-amylase. *Planta*, *240*, 1167–1178. https://doi.org/10.1007/s00425-014-2172-5

Martinez, S. A., Godoy, J., Huang, M., Zhang, Z., Carter, A. H., Garland Campbell, K. A., & Steber, C. M. (2018). Genome-wide association mapping for tolerance to preharvest sprouting and low falling numbers in wheat. *Frontiers in Plant Science*, *9*, 141. https://doi.org/10.3389/fpls.2018.00141

Nakamura, S., Abe, F., Kawahigashi, H., Nakazono, K., Tagiri, A., Matsumoto, T., … Miura, H. (2011). A wheat homolog of MOTHER OF FT AND TFL1 acts in the regulation of germination. *Plant Cell*, *23*, 3215–3229. https://doi.org/10.1105/tpc.111.088492

Perten, H. (1964). Application of the falling number method for evaluating alpha-amylase activity. *Cereal Chemistry*, *41*, 127–140.

Piepho, H. P., & Mohring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics*, *177*, 1881–1888. https://doi.org/10.1534/genetics.107.074229

Purchase, J. L. (1997). Parametric analysis to describe genotype × environment interaction and yield stability in winter wheat. Ph.D. diss., Univ. Orange Free State, Bloemfontein, South Africa.

Rajaram, S., van Ginkel, M., & Fischer, R. A. (1995). CIMMYT's wheat breeding mega-environments (ME). Paper presented at the 8th International Wheat Genetics Symposium. Beijing, China. 19–24 July 1994.

Ross, A. S., Flowers, M. D., Zemetra, R. S., & Kongraksawech, T. (2012). Effect of grain protein concentration on falling number of unger-minated soft white winter wheat. *Cereal Chemistry*, *89*, 307–310. https://doi.org/10.1094/CCHEM-07-12-0085-RC

Sabaghnia, N., Sabaghpour, S. H., & Dehghani, H. (2008). The use of an AMMI model and its parameters to analyse yield stability in multi-environment trials. *Journal of Agricultural Science*, *146*, 571–581. https://doi.org/10.1017/S0021859608007831

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

Steber, C. M. (2017). Avoiding problems in wheat with low falling numbers. *Crops & Soils*, *50*, 22–25. https://doi.org/10.2134/cs2017.50.0208

Su, G., Madsen, P., Lund, M. S., Sorensen, D., Korsgaard, I. R., & Jensen, J. (2006). Bayesian analysis of the linear reaction norm model with unknown covariates. *Journal of Animal Science*, *84*, 1651–1657. https://doi.org/10.2527/jas.2005-517

Wright, K., & Laffont, J.L. (2018). gge: Genotype plus genotype-by-environment biplots. R package version 1.4. Comprehensive R Arch. Network. Retrieved from https://CRAN.R-project.org/package=gge (accessed 1 Oct. 2019).

Yan, W., Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science*, *40*, 597–605. https://doi.org/10.2135/cropsci2000.403597x

Yan, W., Kang, M. S., Ma, B., Woods, S., & Cornelius, P. L. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science*, *47*, 643–655. https://doi.org/10.2135/cropsci2006.06.0374

Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian Journal of Plant Science*, *86*, 623–645. https://doi.org/10.4141/P05-169

Yu, N., Laurenz, R., Siler, L., Ng, P. K. W., Souza, E., & Lewis, J. M. (2015). Evaluation of α-amylase activity and falling number around maturity for soft white and soft red wheat varieties in Michigan. *Cereal Research Communications*, *43*, 672–681. https://doi.org/10.1556/0806.43.2015.026

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Sjoberg SM, Carter AH, Steber CM, Garland-Campbell KA. Unraveling complex traits in wheat: Approaches for analyzing genotype × environment interactions in a multienvironment study of falling numbers. *Crop Science*. 2020;60:3013–3026. https://doi.org/10.1002/csc2.20133