Received: 24 September 2024 | Revised: 17 November 2024 | Accepted: 30 November 2024

DOI: 10.1002/cche.10860

RESEARCH ARTICLE

Random forest regression to predict Farinograph traits from GlutoPeak output in wheat wild relative backcross lines

John H. Price[1,2] | **Mary J. Guttieri**[2] | **Sydney Stutz**[2,3]

[1]Oak Ridge Institute for Science and Education (ORISE), Manhattan, Kansas, USA

[2]USDA-ARS Hard Winter Wheat Genetics Research Unit, Center for Grain and Animal Health Research, Manhattan, Kansas, USA

[3]Department of Agronomy, Kansas State University, Manhattan, Kansas, USA

Correspondence
Mary J. Guttieri, USDA-ARS Hard Winter Wheat Genetics Research Unit, Center for Grain and Animal Health Research, Manhattan, KS, USA.
Email: mary.guttieri@usda.gov

Funding information
Wheat Genetics Resource Center I/UCRC, Grant/Award Number: NSF IIP-1338897

Abstract

Background and Objectives: Flour quality is a key target of hard winter wheat breeding. The Farinograph is important for assessing quality before cultivar release in the United States, but large sample size requirements and long test times render it impractical for early-stage selection relative to the GlutoPeak. To improve GlutoPeak utility for breeding, we calculated new parameters from device raw output and used random forest regression to predict key Farinograph parameters in a winter wheat population containing wild relative introgressions.

Findings: The key quality parameters of absorption, bake absorption, tolerance stability, and mixing tolerance index were moderately well predicted (R^2 ranging from 0.488 to 0.745). Classification of samples as acceptable or unacceptable for mixing tolerance index and tolerance stability was more accurate than prediction of numeric values.

Conclusions: New features calculated from the GlutoPeak raw data were useful predictors of quality. Prediction accuracies are sufficient to improve breeding populations.

Significance and Novelty: This study is the first to use wheat wild relative introgressions in GlutoPeak Farinograph prediction, the first to generate features from raw data, and is one of the few random forest models for quality prediction. The tools that we provide will improve ability to cull poor-quality lines early in the breeding pipeline can support efficient wheat cultivar development.

KEYWORDS
Farinograph, GlutoPeak, machine learning, rheology

1 | INTRODUCTION

Suitability for bread making is a key characteristic for hard winter wheat (HWW) cultivars. Before release, all HWW cultivar candidates must be subjected to rigorous milling, mixing, and baking tests. Failing to meet key quality thresholds may result in the rejection of the cultivar candidate by end-users, decreasing its desirability to growers. Therefore, evaluating end-use quality early in the cultivar development process may allow breeders to avoid expending trial resources on wheat lines that ultimately will not be commercially viable. Early-stage quality evaluations must have high sample throughput for both time- and cost-effectiveness in a breeding pipeline. To meet this need, the end-use quality technology industry continually seeks to develop tools for high-throughput methods.

In the United States, a key quality evaluation is the Farinograph (Brabender, Anton Paar) test, which measures torque during dough mixing to evaluate a wheat flour's development time, stability time, and water absorption capacity. Four key Farinograph measurements that we will consider are as follows: (1) absorption—the mass of water that must be added to the flour to bring the torque measurement to 500 Brabender Torque Units (BU), expressed as a percentage of the flour mass; (2) bake absorption—absorption plus any additional water added during the Farinograph test to maintain proper dough consistency; (3) Mixing Tolerance Index (MTI)—the reduction in torque 5 min after the dough reaches maximum torque; and (4) tolerance stability (TS), which is the length of time for which the torque remains above 500 BU (Bock, 2022). One drawback of the Farinograph is a large sample size (Wang et al., 2021). Depending on the lab and chosen method, 300 g or more of refined flour from each sample may be necessary to run the test (AACC Method 54-21.02). While breeders typically have sufficient grain samples to spare in the late stages of selection, the early selection stages typically produce a much smaller sample size, insufficient for both Farinograph testing and for subsequent generations of yield testing. Additionally, the time required for the farinograph (up to 30 min per sample, plus setup, and clean-up) test limits daily throughput of testing laboratories. Further constraints are imposed by the time required to prepare refined flour on an experimental mill. The throughput of experimental milling operations in quality labs typically ranges from 8 to 20 samples per workday. Consequently, early in the selection process, when breeders have hundreds to thousands of samples, it is typically impractical to identify those genotypes that will show the desired Farinograph traits. The ability to make quality selections earlier could save breeders the time, labor, and cost of carrying a commercially unacceptable genotype to the later stages of selection. Presently, many in the wheat improvement industry look to the GlutoPeak and its potential to provide an early screening method.

The GlutoPeak is a high-shear flour testing method that uses a dry flour or whole-meal sample, combined with deionized water, to measure the required time for complete gluten aggregation, via the measurement of torque (Melnyk et al., 2011). The GlutoPeak has been shown to have reliable correlations of Farinograph water absorption and stability between white flour and whole wheat meal (Wang et al., 2021). Whole meal requires less time and labor to produce, further contributing to the potential utility. A typical GlutoPeak method will require less than 10 g of sample, potentially allowing breeders to distinguish between desirable and undesirable breeding lines (Sissons & Smit, 2018). The smaller sample requirement also provides the opportunity to use small-scale milling for flour preparation. Small-scale milling procedures (e.g., Brabender Quadrumat Junior) can have throughput of 50–100 samples in an 8-h workday. The shorter run time contributes to the efficiency of the GlutoPeak: a test run on the GlutoPeak typically takes 5–7 min, including setup and clean-up. This is much shorter than the Farinograph, which can take up to 30 min (plus setup and clean-up), depending on the sample size, mixing speed, and dough properties. Therefore, the GlutoPeak could enable breeders to accelerate the selection process to more efficiently identify the genotypes that will produce grain that expresses the desired quality traits in milled flour.

Previous work has sought to predict some Farinograph characteristics from GlutoPeak parameters. These studies have either focused on predicting Farinograph trait values per se (Daba et al., 2021; Marti et al., 2015) or a binary classification based on evenly splitting the data (Malegori et al., 2018). These studies have shown reasonably high prediction accuracies for three key Farinograph parameters in relatively narrow germplasm pools: absorption, bake absorption, and TS.

Here, we build on previous work by calculating new features from short (150 s) GlutoPeak runs to predict the key Farinograph parameters of absorption, bake absorption, TS, and MTI, as well as industry-guided categories for stability and MTI, using a random forest approach (Breiman, 2001). By calculating novel summary statistics ($n = 14$) from the raw torque value at each time point, we add to the small ($n = 9$) number of features calculated by the included GlutoPeak software (peak maximum time, maximum torque, torque 15 s before and after maximum, and area under torque curve between four consecutive milestone curve features), a method that we believe increases the utility of this device. The germplasm pool for this study is an introgression population of the wheat ancestral species wild emmer (*Triticum turgidum* subsp. *dicoccoides*) crossed into a HWW (*Triticum aestivum* L.) background, thus highlighting the utility of this method for trait discovery in diverse

material. Our goal is to expand the utility of the GlutoPeak as a tool for early-generation evaluation of breeding germplasm, improving breeders' ability to cull low end-use quality lines early in the cultivar development process.

2 | MATERIALS AND METHODS

2.1 | Plant material

This study included four Kansas-adapted hard red winter wheat cultivars: Bob Dole (PI 690435), Zenda (PI 683512), KanMark (PI 675456), and KS090387K-20 (pedigree 'Winterhawk'/KS011020-6//'Hitch'), and 13 introgression lines of wild emmer introduced into bread wheat with KanMark or KS090387K-20 as recurrent parents in a first backcross.

Wheat was grown in four Kansas trial locations for harvest in 2022: Ashland Bottoms (39.14216, −96.63222), Colby (39.38689, −101.07709), Hutchinson (37.92983, −98.02888), and Hays (38.85195, −99.33479). Trials were grown using standard management practices for rainfed small plot testing within a wheat improvement program. Trials were harvested with a Zürn 150 (Zürn Harvesting GmbH & Co.) small plot combine or a Hege 140 (Hans-Ulrich Hege Saatzuchtmaschinen GmbH) small plot combine. For Colby, Hays, and Hutchinson and some Ashland Bottoms plots, two field replications of each genotype were composited to produce a 1.5 kg sample for milling and baking evaluation. For the remainder of the genotypes at Ashland Bottoms, the two field replications were processed as independent samples. In practice, a number of genotypes yielded poorly at one or more locations, so all four locations of harvest were not used for all genotypes. A total of 67 samples were evaluated.

2.2 | Sample processing

Milling and Farinograph testing was conducted by the Great Plains Analytical Laboratory (GPAL), following AACC-approved methods (AACC 26-21.02; AACC 54-21.02). Bake absorption was calculated as initial absorption plus any additional water added during the Farinograph test to maintain proper dough consistency: this trait is not outlined in AACC methods but is accepted as a proxy for a baker's experience with the flour (Bock, 2022; T. Fontana, GPAL, personal communication, May 24, 2024). Samples were milled using a Bühler laboratory mill (Bühler MLU-202). Before milling, whole-grain protein and moisture data were collected using near-infrared spectrometry (Foss Infratec 1241 Grain Analyzer). Protein was estimated from the total nitrogen content and adjusted to 12% moisture content.

2.3 | GlutoPeak

Milled flour samples received from GPAL were evaluated using a GlutoPeak (Brabender GlutoPeak, Model 803420). Nine grams of flour was mixed with 9 g of deionized water and run at 2700 rpm and 34°C (Bouachra et al., 2017). Run length varied; all features were calculated from the first 150 s of output from the GlutoPeak, as peak formation was consistently observed before this mark, and increased torque, possibly due to starch gelatinization, was inconsistently observed at longer time points.

2.4 | Feature calculation

Feature calculation and all subsequent data analyses were conducted using the R language (R Core Team, 2022). Other than protein concentration, all features used for model construction were calculated from the raw torque/time output from the GlutoPeak. First, to aid in further calculations, a series of sliding window summary statistics were calculated for each 1 s interval, using the R package "slider" (Vaughan, 2021). These were (1) the maximum torque for a 10-second sliding window centered on the time point and (2) the mean torque of the 7 s sliding window after the time point. These summary statistics were then used to identify peak time points, which were defined as time points where the measured torque is within 95% of the 10 s sliding-window maximum, and the difference between the torque at the time point and the 7 s leading mean is greater than 2.5% of the torque at the time point. These criteria were selected based on visual assessment of time × torque plots to most consistently find peaks, rather than shoulders before true peaks. Typically, either one or two peaks occurred in the 150 s time frame. If a second peak occurred, it was always at a higher torque than the first peak. Thus, these peaks were labeled "first peak" and "max peak," and subsequent features were calculated based on these peaks (Table 1). The area under the curve was calculated as the sum of the average torque between each 1 s interval, effectively providing a trapezoidal estimate of the integral.

2.5 | Applying industry-relevant thresholds for quality categorization

In consultation with industry and USDA cereal chemists, it was determined that categorizing flour samples as "unacceptable", "acceptable," or "excellent" for MTI and TS would be valuable. The threshold values that define these categories are presented in Table 2.

TABLE 1 Description of all model features calculated from raw GlutoPeak torque/time output.

Features	Description
A01*	Area under the curve before the first peak.
A12*	Area under the curve from the first peak to trough.
First_peak_time & First_peak_torque	The first peak is defined as the first time point where the measured torque is within 95% of the maximum torque in a 10 s sliding window around the point, and the difference between the torque at the time point and the mean of the torque 7 s after that time point is greater than 2.5% of the torque at the time point. These criteria were selected based on visual assessment of time × torque plots to most consistently find the first peak, rather than shoulders before the peak.
Firstpeakarea	Area under the curve around the first peak, defined as all points contiguous with the first peak where the difference between the torque at the time point and the mean of the torque in a 15 s sliding window around the time point is positive.
Max_before60_torque	Maximum torque before 60 s.
Max_dif	Maximum difference between the torque at a time point and the mean of the torque in a 15 s sliding window around the time point.
Max_dif_time	Time where "max_dif" occurs.
Max_time*	Time of maximum torque.
Max_torque*	Maximum torque.
Maxpeakarea	Area under the curve around the max peak, defined the same way as the area under the first peak.
Peak_slope	Slope from the first peak to trough.
PMslope	Slope from the first peak to 15 s after the peak.
Time_near_max	Number of seconds where torque remains within 80% of max.
Torque_at_150	Torque at 150 s.
Torque_decrease	Difference between the first peak torque and torque at 150 s.
Trough_time and trough_torque	Trough is defined as the lowest torque after the first peak.

Note: Traits denoted with an (*) are also provided in standard GlutoPeak summary statistics from Brabender software (Brabender, Metabridge Glutopeak BMB: 2.2.0 CV 2).

TABLE 2 Category thresholds for mixing tolerance index (MTI) and tolerance stability.

Class	MTI threshold (BU)	Tolerance stability threshold (min)
Unacceptable	Value ≥ 40	Value < 7.5
Acceptable	20 ≤ value < 40	7.5 ≤ Value < 14
Excellent	Value < 20	Value ≥ 14

2.6 | Model construction and evaluation

For each trait of interest, random forest regression or classification was carried out using the R packages "randomForest" and "caret" (Kuhn, 2022; Liaw & Wiener, 2002). For all traits, fivefold cross-validation was used to evaluate model accuracy. For the tuning parameter "mtry" (number of variables randomly selected at each tree split), all values from 2 to 14, advancing by 2, were tested. To obtain the most repeatable values possible for accuracy and feature importance, the reported values for each trait are the median of 30 independent replications of the model construction and cross-validation. Generally, reported accuracy statistics are the out-of-bag (OOB) estimates generated during cross-validation referring to the subset of samples that are held out as a test set during model training. During cross-validation, each sample is held out at least once, allowing for the calculation of prediction accuracies for the population while still allowing every sample to be used in the final model construction. Following model construction, the least-informative predictor was then held out, and model construction was repeated. If OOB accuracy increased, this process was repeated until removing features decreased model accuracy. The set of features that yielded the maximum OOB accuracy was then used.

For each trait, all calculated GlutoPeak parameters listed in Table 1, as well as flour protein concentration, were

supplied as potential predictors for random forest regression or classification. For traits that were predicted through regression, the median OOB root mean square error (RMSE) and R^2 of the 30 model constructions are reported. Feature importance is reported as the increase in node purity (reduction in RMSE) due to the inclusion of the feature, again reporting the median of 30 model constructions. For classification traits, the median OOB accuracy of 30 model constructions and the Spearman rank correlation between observed class predicted from one randomly selected model construction are reported. A single model construction was used for observed versus predicted evaluations to better replicate the way such a model would be used in practice by breeders. Feature importance is reported as mean decrease in the Gini index, which measures the accuracy of classification based on the inclusion of a given variable (Menze et al., 2009).

2.7 | Calculation of Best Linear Unbiased Estimates (BLUES)

To summarize trait values for each genotype across locations, BLUEs were calculated for both observed and predicted Farinograph traits (Henderson, 1975). Here, BLUEs are the estimated fixed effect for each genotype in the linear mixed model Trait = Genotype + Location, where Location is treated as a random effect.

3 | RESULTS

3.1 | Summary of key parameters

For several Farinograph traits, introgression lines increased the range of trait values that would be available for breeding (Table 3). This observation highlights the merit of evaluating larger diversity panels for these traits.

3.2 | Predicting Farinograph absorption and bake absorption

Both Farinograph absorption and bake absorption were relatively well predicted (Figure 1). Importance measures for all features in all prediction models are reported in Supporting Information S1: Table 1. Random forest regression is not intended to prove causality; any mechanistic inferences drawn from the relative importance different model features are potential insights to guide future work, but should not be understood directly as causal relationships. In addition, Pearson's correlation coefficients between all traits and predictors are reported in Supporting Information S1: Table 2: while these correlations do not relate directly to feature importance in random forest regression, they may aid in interpretation. Across 30 independent

TABLE 3 Mean, range, and variance for key traits.

Trait	Germplasm group	Mean	Range	Variance
Absorption (%)	Combined	64.09	58.2–72.8	7.09
	Checks	63.5	58.2–68	7.41
	Introgression lines	64.3	60.4–72.8	7.01
Bake absorption (%)	Combined	66.23	62.5–75	5.23
	Checks	65.9	62.5–70	5.89
	Introgression lines	66.3	63–75	5.09
Mixing Tolerance Index (BU)	Combined	26.18	6–63	172.45
	Checks	20.9	6–53	146.25
	Introgression lines	27.8	7–63	171.93
Tolerance stability (min)	Combined	13.12	3.37–30[a]	49.71
	Checks	16.2	5.05–30	48.69
	Introgression lines	12.1	3.37–30	47.02
Protein (%)	Combined	13.54	11.44–18.18	1.83
	Checks	12.8	11.5–15.4	1.14
	Introgression lines	13.8	11.4–18.2	1.84

Note: Summary statistics are given for the combined population of all lines, as well separately for the set of check cultivars and set of introgression lines.

[a]The Farinograph measurement was ended at 30 min, artificially capping all tolerance stability measures at this time point.

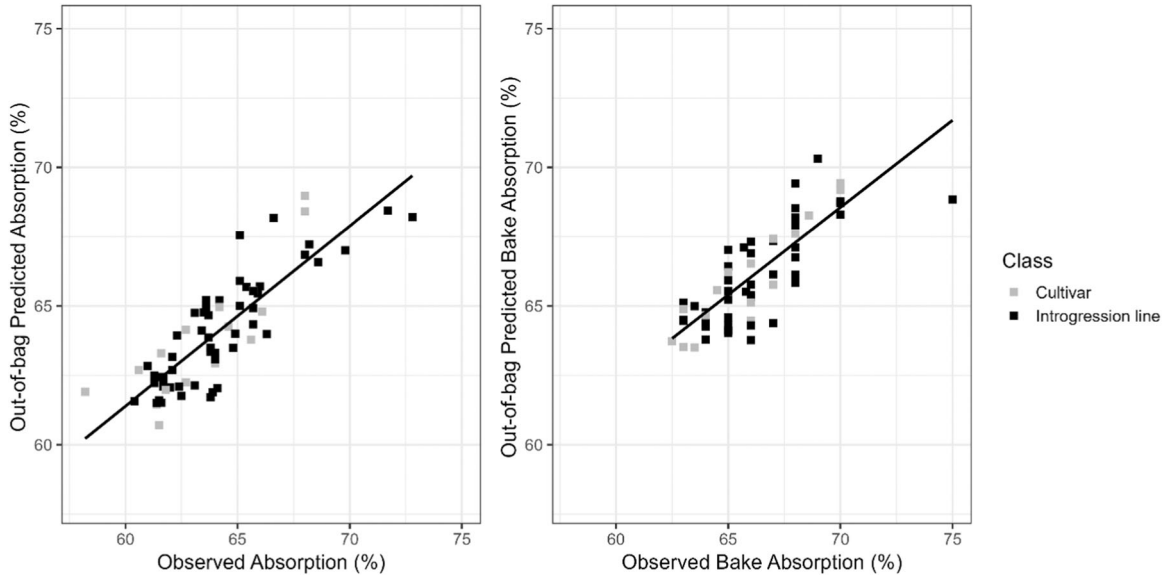

FIGURE 1 Observed versus out-of-bag predicted absorption and bake absorption values for all samples. Predicted values are extracted from a single random forest model construction (rather than a median of 30 independent constructions). Color is used to differentiate bread wheat cultivars from bread wheat lines containing introgressions from wild emmer wheat. The trendlines through data points represent the linear regression of predicted values on observed values.

model constructions, the median OOB RMSE for Farinograph absorption was 1.43 percentage points or 10% of the observed range for the trait. The median model R^2 was 0.745. For bake absorption, the median OOB RMSE across 30 model constructions was 1.26, again 10% of the range for the observed values of the trait. The median model R^2 was 0.726. For Farinograph absorption, the two most important predictors, both with more than double the increase in node purity relative to the third most important feature, were PMslope (the slope from the first peak to 15 s after the first peak), and A12, the area under the curve from the first peak to the trough. For bake absorption, the two most important predictors were trough_torque (the lowest torque value observed after the first torque peak) and Max_torque (the maximum torque value observed).

3.3 | Predicting the MTI

3.3.1 | Numeric trait

The MTI was predicted both as a numeric trait through regression (Figure 2) and with classification of the samples into three categories (Table 2). For the numeric trait, the median RMSE was 9.47 Brabender torque units (BU), or 15.8% of the range observed for the trait. The median model R^2 was 0.521. The two most important predictors were the PMslope and A12.

3.3.2 | Classification

Samples were categorized into one of three MTI categories. "Unacceptable" samples had an MTI equal to or above 40 BU, "acceptable" samples had an MTI below 40 BU and equal to or greater than 20 BU, and "excellent" samples had an MTI below 20 BU. This method produced a relatively high prediction accuracy, with a median OOB prediction accuracy of 0.734 across 30 independent model constructions. A confusion matrix of the model predicted classification for each sample is shown in Table 4. The Spearman's rank correlation between the observed and predicted classification was lower than the OOB prediction accuracy, at 0.647. Crucially, no samples that were observed to have an "unacceptable" MTI were classified as "excellent" and vice versa. Therefore, either culling all "unacceptable" lines or only retaining "excellent" lines would not result in the inclusion of the opposite extreme class in the culled or retained material. The two best predictors for this trait were PMslope and A01, or the area under the torque curve before the first peak. A12, the second most important predictor for the numeric value of the trait, was the third best predictor.

3.4 | Predicting Farinograph TS

3.4.1 | Numeric trait

Like MTI, Farinograph TS was predicted both as a numeric and as a categorical trait (Figure 2). For the

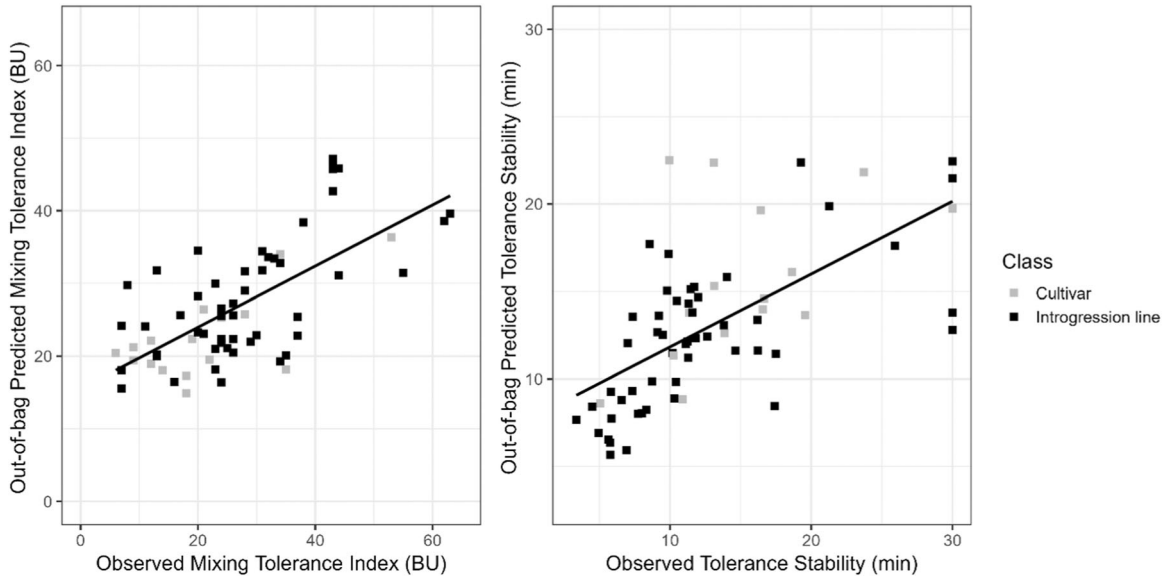

FIGURE 2 Observed versus out-of-bag predicted tolerance stability and mixing tolerance index for all samples. Predicted values are extracted from a single random forest model construction (rather than a median of 30 independent constructions). Color is used to differentiate bread wheat cultivars and bread wheat lines containing introgressions from wild emmer wheat. The trendlines through data points represent the linear regression of predicted values on observed values.

TABLE 4 Confusion matrix for mixing tolerance index classification.

Observed class	Predicted class	Count
Unacceptable	Unacceptable	7
Unacceptable	Acceptable	3
Acceptable	Unacceptable	2
Acceptable	Acceptable	32
Acceptable	Excellent	3
Excellent	Acceptable	10
Excellent	Excellent	10

numeric trait, the median OOB RMSE was 5.13 min, or 19.5% of the range observed. The median model R^2 was 0.486. The two most important predictors were PMslope and A12.

3.4.2 | Classification

Samples were categorized into one of three TS categories. "Unacceptable" samples had a TS below 7.5 min, "acceptable" samples had a TS above or equal to 7.5 min and below 14 min, and "excellent" samples had a TS equal to or above 14 min. This method was relatively accurate, with a median OOB prediction accuracy of 0.634 across 30 independent model constructions. A confusion matrix of the predicted classification for each sample is shown in Table 5: the Spearman's rank correlation between the observed and predicted classification was lower than the OOB prediction accuracy, at 0.56. In this case, one sample that was observed to have an "unacceptable" TS was classified as "excellent." The two most important predictors for this trait were PMslope and A01, the same as the best predictors for MTI classification. Again, as for MTI classification, A12 was the third best predictor.

Following industry guidance, a second classification method was attempted, adding a fourth "Long" category, for samples over 20 min. The overall accuracy of this classification was similar to the three-class TS classification, with a median OOB accuracy of 0.64, and a correlation between predicted and observed classes of 0.585. However, this method only identified one sample in the relatively narrow "Excellent" class between 14 and 20 min (Table 6).

3.5 | Combined analysis of the MTI and TS

TS and MTI are related traits and can be considered together to provide a more complete picture of quality. When the two are combined, 55% of samples are placed in the correct category for both traits and 76% of samples

TABLE 5 Confusion matrix for tolerance stability classification.

Observed class	Predicted class	Count
Unacceptable	Unacceptable	10
Unacceptable	Acceptable	3
Unacceptable	Excellent	1
Acceptable	Unacceptable	4
Acceptable	Acceptable	23
Acceptable	Excellent	5
Excellent	Acceptable	12
Excellent	Excellent	9

TABLE 6 Confusion matrix for tolerance stability (TS) classification, including a fourth class of "Long" sample, with a TS value over 20 min.

Observed class	Predicted class	Count
Unacceptable	Unacceptable	10
Unacceptable	Acceptable	4
Acceptable	Unacceptable	3
Acceptable	Acceptable	26
Acceptable	Excellent	1
Acceptable	Long	2
Excellent	Acceptable	12
Long	Acceptable	5
Long	Excellent	3
Long	Long	1

are classified correctly for at least one of the two traits (Table 7).

3.6 | Correlation of BLUES

For the purposes of model construction and evaluation, all samples have been treated independently. However, for breeders, the ultimate goal is to make selections among genotypes, using information gained from observations of the same genotype set in multiple environments. To this end, we then calculated BLUEs of both observed and predicted trait values for the 17 genotypes used in this study (Supporting Information S1: Table 3). Spearman's rank correlations were then calculated between the BLUEs for observed and predicted traits. These correlations were all high and statistically significant (Table 8).

TABLE 7 Combined confusion matrix for mixing tolerance index (MTI) and tolerance stability (TS) classification.

Observed class (MTI/TS)	Predicted class (MTI/TS)	Count
Unacceptable/ Unacceptable	Unacceptable/ Unacceptable	7
Unacceptable/ Unacceptable	Acceptable/Unacceptable	3
Acceptable/Unacceptable	Acceptable/Acceptable	3
Acceptable/Unacceptable	Acceptable/Excellent	1
Acceptable/Acceptable	Acceptable/Acceptable	21
Acceptable/Acceptable	Unacceptable/ Unacceptable	2
Acceptable/Acceptable	Acceptable/Unacceptable	2
Acceptable/Acceptable	Acceptable/Excellent	3
Acceptable/Acceptable	Excellent/Acceptable	1
Acceptable/Acceptable	Excellent/Excellent	1
Acceptable/Excellent	Acceptable/Acceptable	2
Acceptable/Excellent	Excellent/Acceptable	1
Excellent/Acceptable	Acceptable/Acceptable	1
Excellent/Acceptable	Excellent/Excellent	1
Excellent/Excellent	Acceptable/Acceptable	9
Excellent/Excellent	Excellent/Excellent	9

TABLE 8 Spearman rank correlation coefficients between BLUEs for observed and predicted trait. All values are statistically significant ($p < .01$).

Trait	Correlation
Absorption	0.799
Bake absorption	0.792
MTI (numeric)	0.865
MTI (categorical)	0.659
TS (numeric)	0.775
TS (categorical)	0.766

Abbreviations: MTI, Mixing Tolerance Index; TS, tolerance stability.

4 | DISCUSSION

4.1 | Comparison to previous studies

Since 2015, efforts have been made to relate GlutoPeak measurements to key Farinograph parameters (Marti et al., 2015). These methods have typically made use of a GlutoPeak run time between 5 and 10 min, as opposed to the

150 s measurement used here. Marti et al. (2015) reported a great deal of success in predicting absorption and TS ($R^2 = 0.96$ and 0.88) in 120 commercial Italian wheat varieties using a partial-least-squares regression model. Further studies have obtained lower prediction accuracies, comparable to or somewhat higher than those obtained in the present study (Daba et al., 2021; Rakita et al., 2018; Zawieja, Makowska & Gutsche, 2020). Other authors have also adopted a classification approach to predicting Farinograph values from GlutoPeak parameters. Malegori et al. (2018) obtained a high prediction accuracy for stability on splitting samples into "high" and "low" stability groups; however, the threshold values for these two categories were determined by evenly splitting the training data rather than being determined by industry standards, and thus may be less immediately relevant to breeding efforts.

Although the prediction accuracies reported here are not as high as others in the literature, we believe that this study makes several valuable contributions to this area. This is the first model relating GlutoPeak and Farinograph characteristics in a wild relative introgression population. We predict the key Farinograph parameter of MTI and present methods for using the raw output of GlutoPeak runs. A second advantage of this study is the use of runs terminated at a standardized 150 s, rather than the standard 5–10 min. This change allows for increased sample throughput, further increasing the utility of this method for early-generation testing. Finally, this is the first report to use industry-derived thresholds for MTI and TS classification, increasing relevance to breeding efforts.

4.2 | Utility for breeding

End-use quality is a key set of parameters in HWW cultivar development. In many cases, this characteristic may best be thought of as a threshold selection trait. Rather than seeking a maximal value, a released variety should consistently be above a minimum level of quality (Seabourn, 2006). This prediction method, particularly categorizing MTI and TS, is well suited to this paradigm. Combining the categorical classification for MTI and TS is one scenario in which a breeder could apply this method for selection. In this data set, 55% of samples were correctly classified for both traits and 76% were correctly classified for at least one trait.

One complication in using the results of this study to illustrate applications to breeding is the small number of unique genotypes in the training data set: 17 genotypes replicated across four environments. The high degree of genetic and quality diversity represented by these lines also add to the challenges for prediction. In practice, a breeder applying this method to early generation selection would be selecting from a much larger pool of genotypes, using only one or two locations of data. In the development of prediction models for this study, neither the genotype nor the source location of a flour sample was ever used as a predictor. Because quality traits can vary significantly by location, even within genotypes, there is a limited extent to which illustrating the application of this method to a breeding population by treating each sample as a unique "line" is useful. Scenarios 1 and 2 below follow this approach. In Scenario 3, BLUE values calculated for each of the 17 genotypes for predicted and observed trait values are used to illustrate the utility of this method where it applied to the training data set as actually constructed.

4.2.1 | Scenario 1: High selection intensity

A realistic selection criterion could be to advance lines that are classified as "Excellent" for both traits. Here, 18 of 67 samples were observed to be "Excellent" for both traits, while 11 samples were predicted to meet this criterion, and so would be selected. Of these 11 samples, 9 were in fact also observed to be "Excellent" for both traits. One remaining sample was "Acceptable" for both traits and one sample was "Excellent" for one trait and "Acceptable" for the other. However, this also means that nine of the samples observed to be "Excellent" for both traits would not have been selected. In this instance, the selection criterion would have a low rate of false positives (18%) but a relatively high rate of false negatives (50%). This selection scheme would remove 84% of the population from the breeding program and it would increase the frequency of samples that are "Excellent" for both traits from 0.269 to 0.81. This level of selection intensity may be advantageous in a recurrent selection or parental development context, or when breeding for a high-quality market niche.

4.2.2 | Scenario 2: Low selection intensity

Conversely, this method could be used to cull lines that are unacceptable for either trait. In this scenario, 14 of 67 samples would be culled. Of these, 12 were actually observed to be "Unacceptable" for a trait, while the remainder were observed to be "Acceptable" for both traits, meaning that no "Excellent" samples would be removed. This would also remove all samples that were observed to be "Unacceptable" for both traits, and it would leave four samples that were "Unacceptable" for one trait. This removal of 20% of the population would reduce the proportion of the population with an "Unacceptable" trait from 0.21 to 0.07, while not removing any "Excellent" samples. This relatively low selection intensity could be appropriate to cull lines

early in the breeding pipeline. Alternatively, it could be used as a first threshold selection step late in the breeding pipeline, with remaining lines then selected for agronomic performance, disease resistance, and other traits more immediately relevant to growers' variety selection decisions.

4.2.3 | Scenario 3: Selection among genotypes using BLUE values for predicted traits

The training data used for this experiment comprised of 17 genotypes, replicated over four environments. Realistically, a breeder seeking to make quality selections in this context would create bulked samples, combining multiple environments, and evaluate them using a Farinograph. However, given data for each line × environment combination, BLUEs could be calculated to compare genotype values (Supporting Information S1: Table 3).

Here, the correlations between BLUEs for observed and predicted trait values were actually higher for numeric TS and MTI values, rather than the categorical trait. This, combined with the lack of statistical clarity of linear estimates calculated for ordinal, categorical trait values, suggests that BLUEs for the numeric trait values should be used for selection. In this example, selecting only lines with a BLUE for predicted TS or MTI that falls in the "Excellent" category would result in the selection of 10 of the 17 genotypes. This selection scheme would result in keeping all lines with an "Excellent" BLUE for the observed value of either MTI or TS; in this selected subset, the maximum BLUE for observed MTI is 24.4 BU and the minimum BLUE for observed TS is 13.14 min. Both of these values are near the threshold between "Acceptable" and "Excellent" for their respective traits, further highlighting the ability of the prediction method to identify high-preforming lines.

This scenario helps to highlight an advantage of this approach, relative to Farinograph evaluation in breeding populations. Because of the time savings offered by milling much smaller grain samples and by much shorter tests, we estimate that it would be possible to use our method to obtain several environments' worth of data on a set of breeding lines in less time than would be required to obtain Farinograph measurements on an equivalent number of lines in one location. This would enable the routine calculation of BLUEs for predicted Farinograph parameters, allowing better estimates of end-use quality for breeding lines: this high correlation between BLUEs for observed and predicted Farinograph values further highlights the potential value of this approach.

4.3 | Developing an updating model for Farinograph parameters

Because the Farinograph more directly replicates the commercial baking process than the GlutoPeak, measurements from this tool will likely remain the predominant quality parameters for wheat variety acceptance in the United States for the foreseeable future. Therefore, rather than directly adopting GlutoPeak parameters to select for end-use quality in the final evaluation stages of wheat breeding, we believe that continuing to build on this prediction procedure is worthwhile. To this end, we envision developing a "testing/training" breeding method similar to that used for genomic selection. For a given breeding program, we envision that a small number of representative lines (checks, important parents, elite progeny) would be evaluated using both the Farinograph and GlutoPeak each year, while a larger test set of lines would be evaluated with only the GlutoPeak. These Farinograph data, combined with GlutoPeak data for the tested lines, would be used by breeders to continuously update and evolve the prediction models for their own breeding germplasm in their particular testing environments.

In all, we believe that the tools provided here have the potential to increase the utility of the GlutoPeak for breeding programs. Standardizing all GlutoPeak runs to 150 s would increase sample throughput, relative to the previous standard of 300 s (Bouachra et al., 2017). This would allow for relatively large breeding populations and multiple environments to be evaluated for end-use quality through the prediction of Farinograph characteristics. The application of these methods can contribute to the development of wheat cultivars that combine excellent end-use quality with high yield, disease resistance and stress tolerance, serving wheat growers and processors.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Moses Nyine (former: Kansas State University, current: IITA) for his assistance in identifying germplasm of interest for this study, Rhett Kaufman (USDA-ARS), Lindsey Thies (Grupo Bimbo), and Dr. Mike Tilley (USDA-ARS) for their feedback, particularly in developing thresholds for MTI and TS classification, and Justin Hodge (Kansas State University) for his assistance in running GlutoPeak samples. This project was supported by the WGRC I/UCRC, which was partially funded by an NSF grant contract (IIP-1338897). This research was supported in part by an appointment to the Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the U.S.

Department of Agriculture (USDA). ORISE is managed by ORAU under DOE contract number DE-SC0014664. All opinions expressed in this poster are the author's and do not necessarily reflect the policies and views of USDA, DOE, or ORAU/ORISE. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

DATA AVAILABILITY STATEMENT

R Scripts to recreate data analysis, apply prediction models to new data, and update random forest models with new data may be found at https://github.com/jhprice/Predicting_Farino_from_GP_RF. GlutoPeak and Farinograph data sets are available at AgDataCommons. Dataset DOI: 10.15482/USDA.ADC/26487862.

ORCID

John H. Price http://orcid.org/0000-0002-8306-6448
Mary J. Guttieri http://orcid.org/0000-0002-8099-0686

REFERENCES

AACC Approved Methods of Analysis. *Method 26-21.02. Bühler method for hard wheat*. 11th ed. Approved November 3, 1999. Cereals & Grains Association. https://doi.org/10.1094/AACCIntMethod-26-21.02

AACC Approved Methods of Analysis. *Method 54-21.02. Rheological behavior of flour by Farinograph: Constant flour weight procedure*. 11th ed. Approved January 6, 2011. Cereals & Grains Association. https://doi.org/10.1094/AACCIntMethod-54-21.02

Bock, J. E. (2022). The Farinograph: Understanding Farinograph curves. In J. E. Bock, & C. Don (Eds.), *The Farinograph handbook* (4th ed., pp. 33–41). Woodhead Publishing.

Bouachra, S., Begemann, J., Aarab, L., & Hüsken, A. (2017). Prediction of bread wheat baking quality using an optimized GlutoPeak®-Test method. *Journal of Cereal Science*, *76*, 8–16.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Daba, S. D., Simsek, S., & Green, A. J. (2021). Predictive ability of four small-scale quality tests for dough rheological properties and baking quality in hard red spring wheat. *Cereal Chemistry*, *98*(3), 660–672.

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

Kuhn, M. (2022). *caret: Classification and regression training*. R package version 6-0-93. https://CRAN.R-project.org/package=caret

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Malegori, C., Grassi, S., Ohm, J. B., Anderson, J., & Marti, A. (2018). GlutoPeak profile analysis for wheat classification: Skipping the refinement process. *Journal of Cereal Science*, *79*, 73–79.

Marti, A., Ulrici, A., Foca, G., Quaglia, L., & Pagani, M. A. (2015). Characterization of common wheat flours (*Triticum aestivum* L.) through multivariate analysis of conventional rheological parameters and gluten peak test indices. *LWT-Food Science and Technology*, *64*(1), 95–103.

Melnyk, J. P., Dreisoerner, J., Bonomi, F., Marcone, M. F., & Seetharaman, K. (2011). Effect of the Hofmeister series on gluten aggregation measured using a high shear-based technique. *Food Research International*, *44*(4), 893–896.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*, *10*, 213.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rakita, S., Dokić, L., Dapčević Hadnađev, T., Hadnađev, M., & Torbica, A. (2018). Predicting rheological behavior and baking quality of wheat flour using a GlutoPeak test. *Journal of Texture Studies*, *49*(3), 339–347.

Seabourn, B. (2006). *Recommended quality targets for hard red winter wheat*. Hard Red Winter Wheat Quality Assessment Committee.

Sissons, M., & Smit, J. (2018). Small-scale methods to assess the gluten properties of durum wheat. *Cereal Chemistry*, *95*(3), 456–468.

Vaughan, D. (2021). *slider: Sliding Window functions*. R package version 0.2.2. https://CRAN.R-project.org/package=slider

Wang, K., Sangha, J., Cuthbert, R., & Fu, B. X. (2021). Effectiveness and biochemical basis of wholemeal GlutoPeak test in predicting water absorption and gluten strength of Canadian hard red spring wheat. *Cereal Chemistry*, *98*(4), 878–890.

Zawieja, B., Makowska, A., & Gutsche, M. (2020). Prediction of selected rheological characteristics of wheat based on GlutoPeak test parameters. *Journal of Cereal Science*, *91*, 102898.

SUPPORTING INFORMATION

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

How to cite this article: Price, J. H., Guttieri, M. J., & Stutz, S. (2025). Random forest regression to predict Farinograph traits from GlutoPeak output in wheat wild relative backcross lines. *Cereal Chemistry*, 1–11. https://doi.org/10.1002/cche.10860