Received: 11 August 2017 | Accepted: 20 March 2018

DOI: 10.1002/cche.10049

RESEARCH ARTICLE

Falling number of soft white wheat by near-infrared spectroscopy: A challenge revisited

Stephen R. Delwiche[1] | **Ryan W. Higginbotham**[2] | **Camille M. Steber**[3]

[1]Food Quality Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, USDA, Beltsville, Maryland

[2]Cereal Variety Testing Program, Washington State University, Pullman, Washington

[3]Wheat Health, Genetics and Quality Research Unit, Agricultural Research Service, USDA, Pullman, Washington

Correspondence
Stephen R. Delwiche, Food Quality Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, USDA, Beltsville, MD.
Email: stephen.delwiche@ars.usda.gov

Funding information
USDA-ARS; Washington Grain Commission

Background and objectives: Wheat Hagberg falling number (FN) is a long-standing quality test that, by means of measuring the viscosity of a heated water-meal or water-flour mixture, characterizes the activity of endogenous α-amylase, the enzyme primarily responsible for starch hydrolysis. The accuracy, time requirement, and cost of this test have come under heightened scrutiny, particularly in seasons when weather conditions have been favorable to preharvest sprouting or late maturity amylase. Near-infrared (NIR) spectroscopy, an analytical approach routinely used in the grain industry to measure contents of protein and moisture, was reexamined as a possible alternative to the FN procedure.

Findings: Partial least squares (PLS) regression quantitative models developed on a genetically diverse set of Washington grown white wheat demonstrated low accuracy, with standard errors of performance ranging from 40 to 77 s. Alternatively, linear discriminant analysis and PLS discriminant analysis (PLSDA) qualitative models, developed and tested using a FN cutoff (pass/fail) value, also demonstrated low accuracy, with the best model correctly identifying 67% and 71% of the samples, respectively, above and below a threshold value established as the median value of FN in a calibration set of several hundred samples.

Conclusions: Replacement of the FN test with one based on NIR spectroscopy on either whole grain or ground meal for making decisions on segregating wheat lots according to α-amylase activity is not recommended.

Significance and novelty: Because NIR spectroscopy is not sufficiently accurate to quantitatively model FN or differentiate low from high FN grain, viscometry procedures for starch integrity, such as FN, will continue their use in grain commerce.

KEYWORDS
falling number, near-infrared spectroscopy, wheat

1 | INTRODUCTION

The Hagberg falling number (FN) procedure, by means of one reported value, measures the viscosity of a wheat flour, wheat meal, or barley meal in water mixture that is agitated and heated under a controlled regiment. It is used to characterize the activity level of endogenous enzymes, and in particular, α-amylase, that catalyze starch hydrolysis. Generally, the food and malt industries prefer that ingredient grain has low enzyme levels, such that they may control these levels during processing by addition of malt or by their own controlled malting regiments. FN provides the basis for characterizing the amylase level of raw grain, and because of this it is used in world commerce as either a

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.

component on official grade, as in the EU, or on contracts of sale, as in the United States. The procedure entails weighing a fixed amount of meal, combining this with a precise volume of distilled water in a dimensioned test tube, and placing a stirrer of precise geometry and weight into the tube. The tube is immersed in a boiling water bath whereupon the stirrer is moved up and down repeatedly for 1 min, during which time the starch gelatinizes and undergoes enzymatic hydrolysis. At the end of the agitation period, the stirrer is released at its highest position and the force of gravity moves the stirrer downward. The elapsed time, in seconds, from start of agitation until the bottoming out of the stirrer inside the test tube, is designated as the FN. The rate of heat transfer (boiling water bath to meal mixture within tube), starch gelatinization, and enzyme kinetics through activation and deactivation, all contribute to the FN time. Typically, sound U.S.-grown wheat will have FNs of 300 and higher, while wheat that has elevated α-amylase activity brought on by preharvest sprouting (PHS) from rain right before harvest or by late maturity amylase (LMA) from large daily temperature changes during grain fill, in extreme cases may have FNs <100 (Farrell & Kettlewell, 2008; Mares & Mrva, 2014). Successive harvests of low FN soft white wheat in spotted locations of the U.S. Pacific Northwest have renewed attention on the FN procedure. While it is relatively simple to perform, equipment necessary for this procedure is expensive (>20,000 USD) and not perfectly adapted to the first point of sale, the country elevator. Thus, FN values are not determined soon enough to allow segregation of low and high FN wheat lots at the elevator. This leads to financial losses when low FN grain is mixed with high FN grain. Near-infrared (NIR) analyzers have enjoyed widespread use at such locations for over two decades as tools for rapid analysis of moisture and protein. Hence, the question surfaces periodically as to whether the NIR technique may be used in lieu of FN to prevent mixing of low and high FN grain at the elevator. Several years have passed since the first research efforts on developing a proxy for FN by NIR methodology, during which time advances have occurred in both NIR instrumentation and chemometric analysis.

Among other bread-making quality parameters (hardness, SDS sedimentation volume, loaf volume), Starr, Morgan, and Smith (1981) reported on NIR reflectance calibrations for FN using prevailing technology at the time, an NIR spectrometer with 19 fixed interference filters and chemometric analysis consisting of multiple linear regression (MLR). Using five filters (1,778, 1,818, 1,982, 1,940, and 2,100 nm), the MLR model produced standard error for a small calibration set (n = 45) of 26.8 s. However, when the model was applied to a separate validation set (n = 43), the standard error rose to 62.3 s. Such a difference in error led to Starr's admonishment that NIR model evaluation should not be performed on calibration samples alone. Osborne (1984) performed similar analyses using new spectrometer technology at the time, a scanning monochromator, which allowed for a continuum of absorbance values collected over a range of 1,200–2,400 nm. With the same type of chemometric analysis as Starr et al. (1981), Osborne's overall findings on FN modeling (n = 52 UK wheat meal samples), with a high standard error of 73 s on the calibration set, precluded any pursuit of model validation. More recent investigations have involved on-combine spectral collection during harvesting operations for field mapping of FN (Risius, Hahn, Huth, Tölle, & Korte, 2015) and hyperspectral image analysis of kernels for α-amylase (Xing, Symons, Hatcher, & Shahin, 2011) or FN (Caporaso, Whitworth, & Fisk, 2017), which have led to a resurgence of interest in developing an NIR calibration for FN. Further, urgency in resolving an issue of a combination of PHS and LMA contributing to low FN wheat in the U.S. Pacific Northwest in 2016 has prompted the authors to return to the NIR feasibility question. Therefore, the objective of this study was to develop and evaluate NIR regression calibrations for FN and, alternatively, determine the feasibility of using NIR technology for a pass/fail evaluation of wheat samples about a cutoff value. While the pass/fail model would not likely serve as a replacement to a 300 s FN contract specification, if successful, the model could allow segregation of low and high FN grain at the elevator.

2 | MATERIALS AND METHODS

2.1 | Samples

All wheat (soft white winter and soft white spring) samples were from a subset of the 2016 Washington State University cereal variety trials as described in Higginbotham, Jitkov, and Horton (2016), summarized in Table 1, and shown by geographical location in Figure 1. This subset consisted of commercial varieties or advanced breeders lines grown in research field plots at 16 locations in central and eastern Washington. Wheat was planted in plots of 5.0–7.4 m^2 arranged in an alpha lattice design with three replicated plots per entry and harvested with a small plot combine. Trials were maintained under the cooperating grower's management conditions specific to the trial location including fertilizer application, herbicide weed control, and fungicide treatment for stripe rust control. The samples were cleaned and weighed at the USDA-ARS laboratory in Pullman, WA and sent to the USDA-ARS Beltsville, MD laboratory in three shipments (groups) starting in January 2017 and lasting through late April 2017.

TABLE 1 Structure of soft wheat samples used in near-infrared (NIR) modeling of falling number by quantitative (partial least squares) and qualitative (linear discriminant analysis) approaches

	Calibration set			Validation set			
Group	*n*	Location	Number of varieties	*n*	Location[a]	Number of varieties	Number of varieties unique to validation set
1	145	Connell, Harrington, St. Andrews, Anatone, Creston, Lamont, Reardan, St. John, Fairfield, Moses Lake, Walla Walla	61	47	Farmington	35	9
2	101	Pullman	53	54	Farmington	24	23
3	153	Almira, Mayview	68	44	Lamont, Plaza, Walla Walla	22	14

[a]Farmington samples from group 1 are of different varieties compared with Farmington samples from group 2. Similarly, Lamont samples are of different varieties between these groups.

2.1.1 | Group 1

One hundred ninety-two pure line 10- to 15-g cleaned field plot samples were delivered to Beltsville in whole kernel (bulk) format. Samples from 11 of the 12 growing locations were assigned to calibration development and, as such, these were drawn from a much larger (super) set with the purpose of producing a uniform distribution of FN over a large range (<100–>450 s), as measured by the Pullman laboratory. The reserved growing location constituted samples that were selected at random without prior knowledge of their FN values. At Beltsville, all samples were scanned in NIR diffuse reflection, first in whole kernel format. After grinding in a cyclone mill (Udy, Ft. Collins, CO, USA), they were scanned again, this time as ground meal. Details of the NIR equipment and methodology appear in a separate section. FN measurement at Beltsville followed the collection of the meal spectra. Depending on the amount of meal available, FN measurements were collected on left and right tube aliquots or, for samples <14 g, on just one tube.

2.1.2 | Group 2

One hundred fifty-five 10-g samples (ground meal only) were sent from the Pullman laboratory to the Beltsville laboratory 3 months after the first group. During the period between shipments, the group 2 samples were stored as intact seed at −15°C until being ground into meal immediately before shipment. The samples were from two growing locations, with the location containing the larger number, at 101, relegated to calibration development, and the other location, with 54 samples, reserved for model validation. As with some of the samples from group 1, FN measurements at Beltsville were performed on a single tube per sample due to the limited mass of material.

2.1.3 | Group 3

One hundred ninety-seven samples in ground meal format were delivered to Beltsville 3 weeks after the group 2 samples. These samples were handled and ground at Pullman

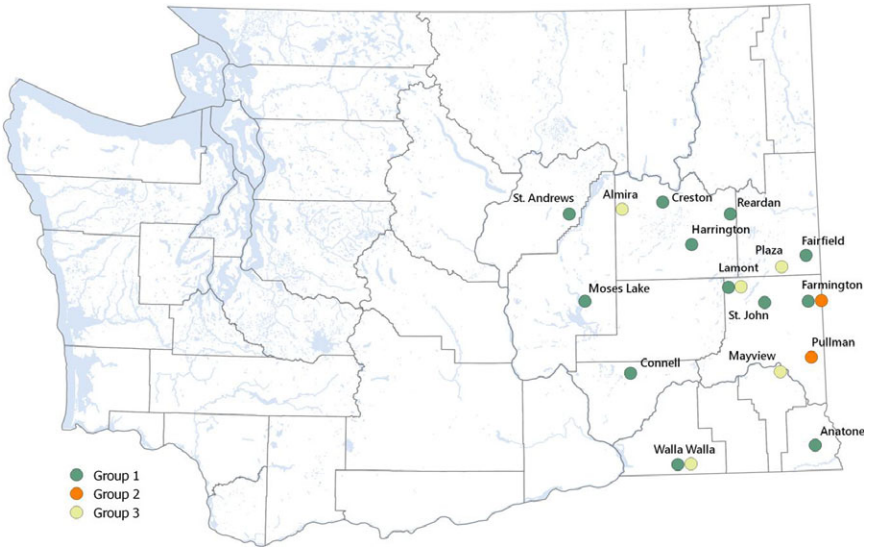

FIGURE 1 Growing locations in Washington State of the samples used in study, identified by analysis group (1–3)

in a similar fashion as the group 2 samples. Samples from the two growing locations that contributed the largest number of samples, at 153, were assigned to calibration development, while 44 samples from three other locations were reserved for model validation.

2.2 | Near-infrared data collection

Group 1 samples were scanned in whole kernel format and again in ground meal format. Samples from groups 2 and 3 were scanned in ground meal format only. In all cases, material was inserted in a 38-mm diameter × 10 mm height standard ring cell possessing a quartz window on the face oriented toward the spectrometer, and a flat rubber stopper on the opposite face to hold the material in place. An analytical bench scanning monochromator (Foss NIRSystems model 6500) with rotating drawer front attachment was used to collect the NIR spectral data. Diffuse reflection scans (1,100–2,498 nm, every 2 nm) were collected using the default settings of 32 repetitions per scan. Before each sample scan, a similar scan was made of the internal reference ceramic material. Sample spectra, referenced to the ceramic, were stored in log(1/*R*) format. Two fills of the ring cell were scanned back-to-back, with new material used in the second fill. Spectra, stored to disk in the manufacturer's native software (Vision) binary format, were later imported into a user-written SAS program that invoked PROC PLS for initial analysis to optimize spectral preprocessing (Reeves & Delwiche, 2003). Later, the spectra were imported into the software package Unscrambler (v. 10.4 Camo, Oslo, Norway) for quantitative and qualitative modeling of FN using the favorable preprocessing conditions.

2.3 | Falling number measurement

Falling number measurements were performed within 2 weeks after the spectral data collection of each group. A Perten model 1000 instrument with cooling tower was used. Following the USDA Federal Grain Inspection Service directive (FGIS, 2013), meal and water quantities were 7.00 g and 25.0 ml, respectively, for all samples. Meal and water were combined in a standard FN test tube using the layering procedure described in Delwiche, Vinyard, and Bettge (2015) to ensure the complete wetting of meal. Because the moisture contents of all samples were consistently close to 10%, FN values were recorded on an as is basis. Approximately half of the group 1 samples possessed sufficient material for left and right tube assays. Samples with insufficient material for two tubes were run in tandem with a dummy sample. Seventy of the dummy sample loadings consisted of material from the same grind. The FN values of these loadings were used to calculate a standard deviation for the FN procedure and hence establish the lower limit for standard error of any NIR regression model for FN.

All samples in group 2 were <14 g, so two samples were analyzed per FN run. For group 3, nearly all samples were >14 g, and consequently each sample was analyzed in left and right tubes, with their average FN used in the modeling analysis.

2.4 | Spectral analysis

The three groups were first analyzed separately, starting with group 1 and continuing with the second and third groups as each became available. Later, similar analysis was performed on pooled samples from all groups. The user-written SAS macro program for partial least squares (PLS) regression, a slight modification of the one described in an earlier publication (Reeves & Delwiche, 2003), was used to evaluate the effect of spectral preprocessing and to compare whole kernel format with ground meal format. Specifically, 136 trials were run in a loop structure to examine effects of two common spectral normalization functions (standard normal variate transformation, with and without spectral detrending, and multiplicative scatter correction), three derivative orders (zeroth or smooth, first derivative, and second derivative, performed according to Savitzky and Golay's (1964) convolutions), and 11 convolution window widths for the derivatives (5–25 points, odd). Program output consisted of a listing of calibration and validation statistics for each trial, whereupon trials were ranked by one figure of merit, the root mean square of differences (RMSD; reference FN minus NIR-predicted FN) of a leave one out cross-validation. This procedure was performed on each group, whereupon the rankings were inspected for any trends in spectral preprocessing performance. Preprocessing conditions that consistently produced relative favorable model results then became the settings for in depth analysis in Unscrambler. Once imported into Unscrambler, the spectra were transformed according to the optimal preprocessing conditions. Additionally, a category (class) variable was created, this being a high–low indicator for when a sample was above or below a cutoff value. Although 300 s value is often used by the wheat industry in the U.S. Pacific Northwest as the point at which wheat lots falling below this value become discounted, 355 s was selected because it was the median of the calibration samples pooled from all three groups. Analysis at the higher value permitted a more balanced set to perform linear discriminant analysis (LDA) and PLS discriminant analysis (PLSDA), with the assumption being that if proof of concept could be established at this value, then the concept would also apply to a lower value, such as 300 s. Predictor variables for the LDA models were

principal component scores from the decomposition of the spectra after the selected optimal preprocessing.

3 | RESULTS AND DISCUSSION

Histograms of FN are shown in Figure 2, with the calibration set and validation set samples from all three groups contained in the upper (2a) and lower (2b) graphs, respectively. Also included in these graphs are histograms of group 1 samples alone as the calibration samples from this group, by design, had the largest range (62–606 s). The vertical dashed bars in each graph are situated at 300 and 355 s, where the lower value corresponds to the cutoff commonly used in U.S. commerce at which a lot may be discounted or refused when FN falls below the cutoff. The upper value corresponds to the median value of the all groups calibration set and, while well within the acceptable range for a traded lot, is better suited in this study to establish feasibility of using NIR spectroscopy in qualitative (categorical) modeling of FN. Because of the deliberate structuring of group 1 calibration samples, the range is wider in the upper graph (calibration) than the lower graph (validation). Raw spectra of the 145 group 1 calibration samples in bulk and ground meal formats are shown in Figure 3a,b, respectively. Typical of NIR spectra, a systemic (wavelength region-wide) variation occurs among the samples, primarily caused by the physical effects of kernel size in the bulk format and particle size in meal format, with the latter format producing less variation.

3.1 | Quantitative models

Preprocessing effect results are presented in Figure 4. Clear from this plot is that the format of the material, intact kernels (bulk) or ground meal, had a direct bearing on PLS model error, with meal demonstrating typically about half the error as intact kernels. This analysis was conducted on the group 1 samples alone, and because of the clear inferiority of the bulk format, the two subsequent groups were analyzed as ground meal only. Also evident in this figure is that once the 25 worst preprocess conditions were removed, the remaining ones showed only slight

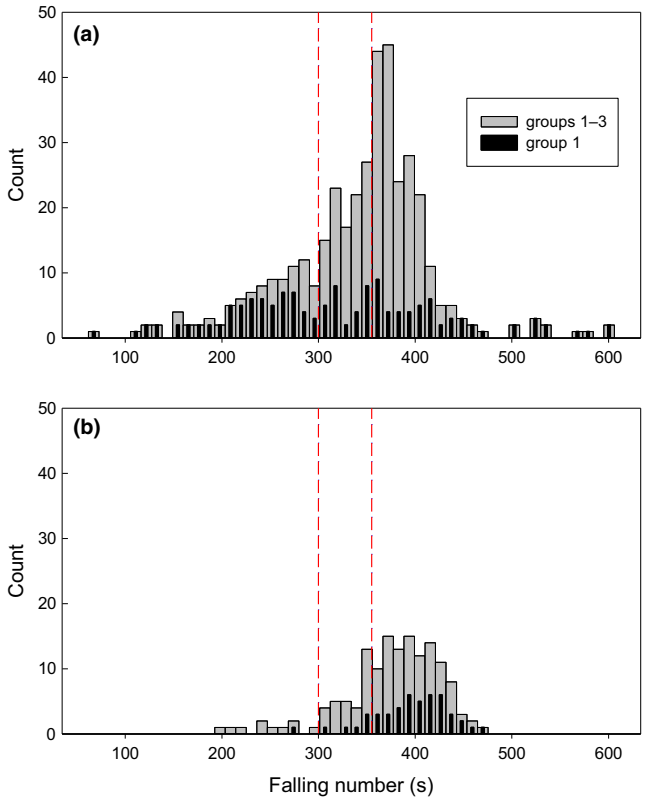

FIGURE 2 Histograms of falling number (FN) of all three analysis groups pooled ($n_{calibration}$ = 399, $n_{validation}$ = 145) and of group 1 alone ($n_{calibration}$ = 145, $n_{validation}$ = 47). (a) calibration samples, (b) validation samples. The dashed vertical line at 300 s corresponds to a typical threshold for which wheat lots having FNs below this value are discounted. The dashed vertical line at 355 s corresponds to the median of pooled calibration samples

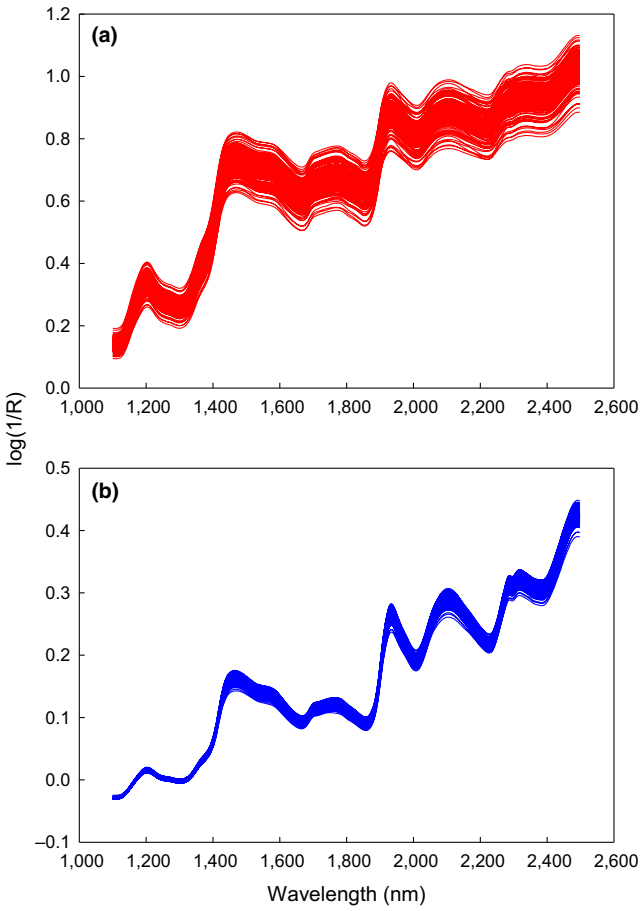

FIGURE 3 Raw spectra of group 1 calibration set samples (n = 145). (a) bulk kernels, (b) ground meal

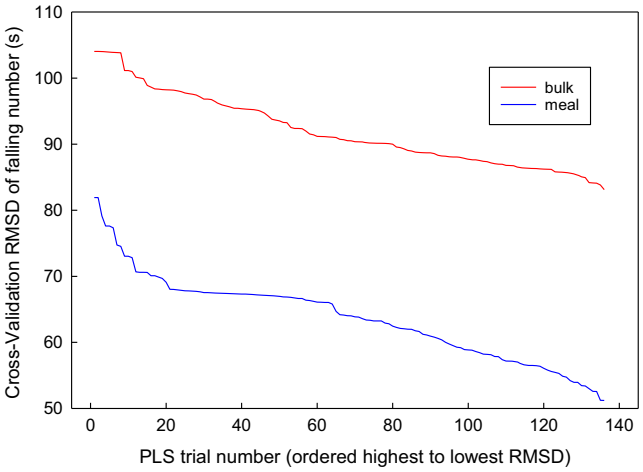

FIGURE 4 Partial least squares (PLS) model performance as defined by the root mean square of differences (RMSD) of near-infrared (NIR) modeled (on ground meal and bulk kernel formats) and actual falling number values in a one-sample-out cross-validation of the group 1 calibration set at various trials of spectral preprocessing (normalization, detrending, smooth, first derivative, second derivative, and convolution window size used in the smooth and derivative operations)

differences in error. For each preprocessing condition, the ensuing PLS regression equation was based on no more than 15 factors (and often, fewer than 15), for which the cross-validation was at a minimum or the difference between the error at the minimum factor number and the error at a smaller factor number was not statistically significant (Reeves & Delwiche, 2003). Although this figure does not indicate the preprocessing condition with each trial number, a general trend was observed when this analysis was applied to the other two groups. In general, normalization either by SNV (with or without a post application of spectral detrending) or MSC was beneficial. Secondly, a wider convolution width for the Savitzky–Golay convolution window was beneficial regardless of derivative order. Lastly, the second derivative preprocessing tended to produce models with lower error than either the smoothing or first derivative preprocessing. Taking these observations into consideration, one preprocessing condition was selected for all further analysis, this being SNV with no detrending, followed by a 15-point second derivative. The number of factors was set at seven for individual group analyses and both seven and 11 for the pooled group analyses.

The PLS regression results are summarized in Table 2. Included in this table are the ranges and standard deviations for the reference FN measurements, separated by calibration and validation sets. Judging from the calibration set alone, the best model, based on goodness of fit, occurred with the group 1 samples, in which the R^2 was .708. This value was noticeably greater than those of the other groups (R^2 of .533 and .375), most likely because the samples for the group 1 calibration were purposely selected to span a very wide FN range (62–606 s). However, even with this wider range, the ratio of *SD* to RMSD, a dimensionless indicator of the goodness of a model, at 1.56, was low for NIR modeling purposes (Williams & Sobering, 1993). The corresponding ratios for groups 2 and 3, at 1.12 and 1.06, were even lower. The expansion of the calibration set to 399 spectra, accomplished by pooling all groups, resulted in very poor modeling statistics as well, with R^2 values of .371 and .541 for the 7- and 11-factor models, respectively, and *SD* to RMSD ratios on par with group 2. Overall, the calibration set RMSD values ranged from 29 to 68 s, with the magnitude depending on the distribution of FN used in the calibration. Group 1, having the broadest distribution (*SD* = 107 s), produced the largest RMSD. By contrast, group 2, with the narrowest distribution in FN, produced the smallest RMSD (28.7 s). However, at 29 s, even this value is very large compared with the standard deviation of the repeated runs (n = 70) of the dummy sample in group 1 (the best estimate of the lowest possible

TABLE 2 Summary of partial least squares regression modeling of falling number

	Calibration set						Validation set				
Group	*n*	Range (s)	*SD*[a] (s)	Factors	R^{2b}	RMSD[c] (s)	*n*	Range (s)	*SD*[a] (s)	SEP[d] (s)	Bias[e] (s)
1	145	62–606	107.1	7	.708	68.5	47	276–472	39.5	42.8	−142.6
2	101	264–435	32.3	7	.533	28.7	54	238–455	48.0	39.9	−6.2
3	153	151–434	51.5	7	.375	48.8	44	194–433	54.3	55.9	−9.9
1–3	399	62–606	75.3	7	.371	65.7	145	194–472	50.8	77.4	−64.3
1–3	399	62–606	75.3	11	.541	66.0	145	194–472	50.8	58.3	−55.4

[a]Standard deviation of falling number measurements of respective set (calibration or validation) as measured by falling number procedure.
[b]Coefficient of determination of regression equation developed on calibration set.
[c]Root mean square of differences (RMSD) (actual—modeled) from a one-sample out cross-validation of the calibration set.
[d]Standard error of performance (= standard deviation of differences) from application of regression equation onto the validation set.
[e]Difference in means of modeled and actual values of validation set.

value for NIR model error), which was 7.1 s. Hence, not even the best model reported in Table 2 would be satisfactory for use as a replacement for the FN procedure.

For visual representation purposes, plots of reference vs. NIR FN values are shown for a 7-factor PLS model developed on group 1 samples in the left column of Figure 5 (graphs a–c). In general, for good models, the cross-validation scatter plot will not be substantially different from the calibration plot. In Figure 5, the amount of dispersion from the line of perfect fit (the 45° line) is slightly greater in the cross-validation graph (b) than in the calibration graph (a). Although the clustering of points about the 45° line in the calibration graph (and to a lesser extent in the cross-validation graph) may signify some power of the NIR model, this, to a certain extent, is arising from a random fitting of the model to the prevailing spectral and reference values. To illustrate this point, a second calibration graph (d) is included in Figure 5. This graph was produced by a 12-factor PLS model in which the FNs in group 1 were no longer affixed to their true spectra, but instead were randomly assigned to the spectra. The clustering of points along the 45° line demonstrates the trap of presenting model findings without validation. In this case of fictitious data (graphs d–f), it is seen that the trending of samples along the 45° line disappears with cross-validation (graph e) and is completely absent with the validation set (graph f). Lastly, the plot of true reference vs. NIR values of the validation set, as seen in graph c, confirms the lack of NIR modeling ability for FN. The large bias (−143 s) is also suggestive of a lack of model robustness. Thus, despite the

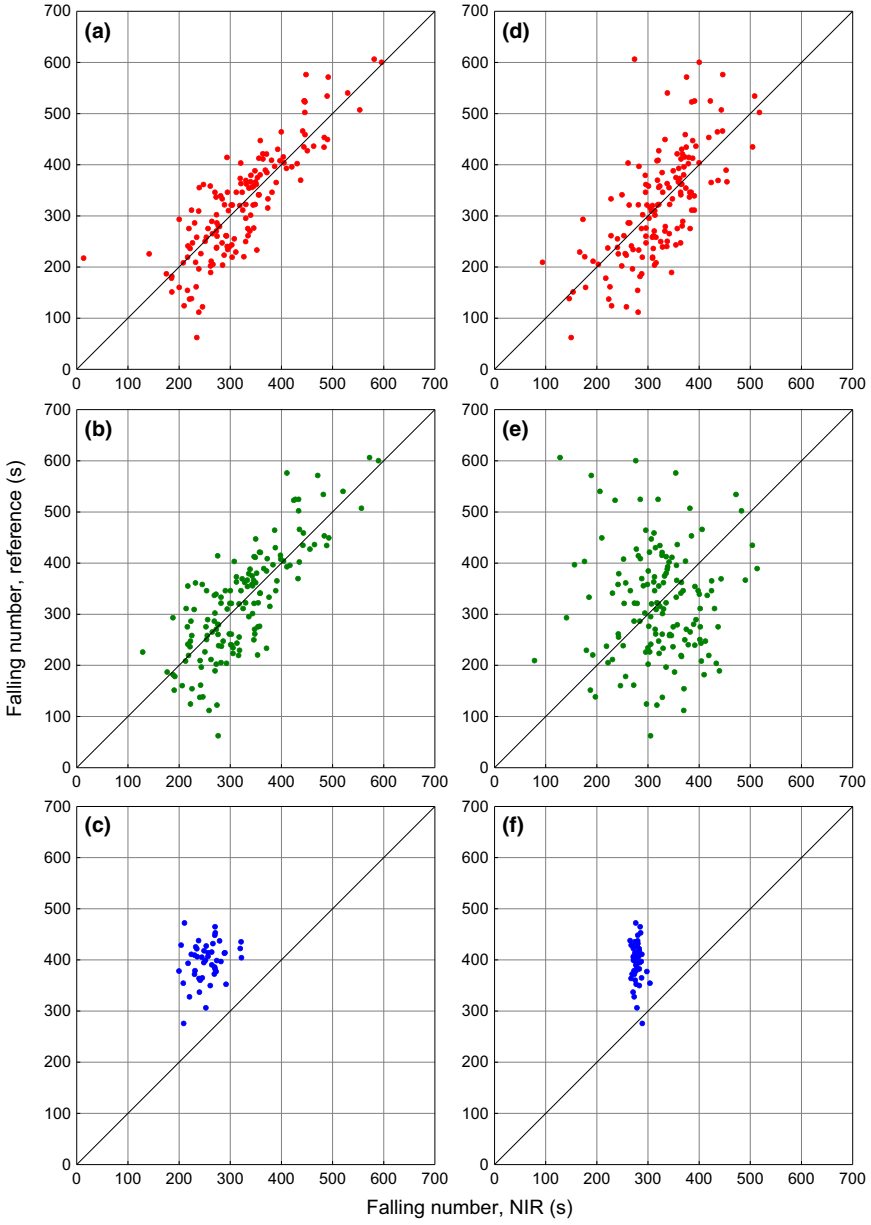

FIGURE 5 Reference vs. NIR plots of falling number (FN) by partial least squares (PLS) regression developed on group 1 samples, with actual and fictitious values for FN. Left column (a–c) actual values (model statistics summarized in Table 2); right column (d–f) fictitious values formed by shuffling the calibration set FN values but leaving validation set FN values in correct order. Top row (a & d) calibration (resubstitution of spectral data used to develop regression equation); middle row (b & e) one-sample-out cross-validation of calibration set, bottom row (c & f) validation set

use of regression algorithms that were generally not available in the early 1980s, the current quantitative model results are no better than those of Starr et al. (1981) or Osborne (1984). Limiting the wavelength region in the PLS regressions to only the lower half (1,100–1,798 nm), or by removal of the 1,900–2,000 nm water absorption region, did not improve the model results.

3.2 | Qualitative models

Using the same preprocessing conditions as chosen for the quantitative models (i.e., SNV, 15-point second derivative), the results of the classification trials for the upper cutoff (355 s) are presented in Table 3. Of the three LDA classifier metrics examined, Mahalanobis distance was best, with 61 of the 104 high samples (58.7%) correctly classified and 32 of the 41 low samples (78.0%) correctly classified. Expressed as total correct/total number, the correctness ratio was 64.1%. This ratio is slightly less than that for an 11-factor PLSDA model (total = 68.3%, with high = 67.3% and low = 70.7%). A graphical representation of the challenging nature of low vs. high classification is shown in a factor 2 vs. factor 1 scores plot of the PLSDA model (Figure 6), in which there is no clear boundary that separates low (<355 s) from high (>355 s) samples. In fact, the only clustering evident in this figure arises from an environmental effect, sample grouping, with all group 1 samples situated on the left region of the graph, as identified by a dashed boundary. Score plots involving all other combinations of factors (not shown) were also incapable of separating low and high FNs.

As mentioned earlier, it is assumed that comparable performance would have been obtained if the median value in the calibration set was lower, for example at 300 s. Because all correctness rates fall very short of an acceptable value for commerce, for example, 95% or higher, the qualitative modeling results of this study, similar to quantitative modeling results, are not encouraging. Thus, it does not appear that NIR spectroscopy can be used to differentiate low from high FN grain at the elevator for making decisions on binning.

TABLE 3 Summary of linear discriminant analysis and partial least squares discriminant analysis (PLSDA) models of falling number applied to the validation samples from all three groups (n = 145)

Classifier function	High samples (≥355 s) correctly classified (%)	Low samples (<355 s) correctly classified (%)
Linear	45.2	75.6
Quadratic	55.8	78.0
Mahalanobis	58.7	78.0
PLSDA	67.3	70.7

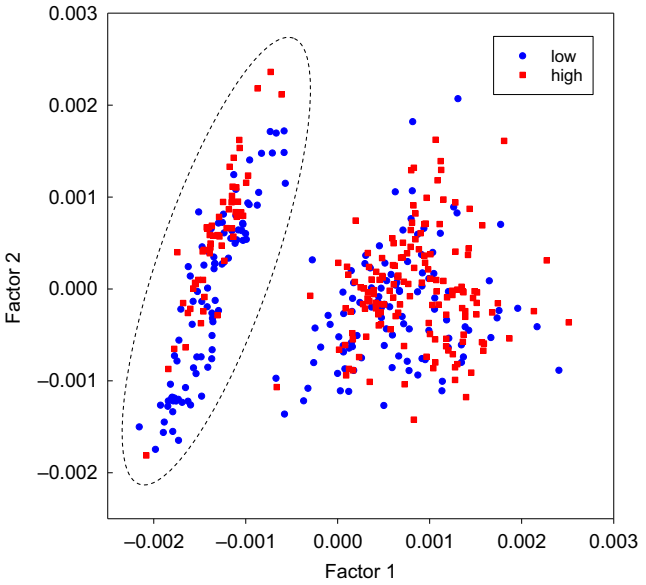

FIGURE 6 Scores plot of the pooled groups 1–3 calibration set partial least squares discriminant analysis of a two-class model (spectral preprocessing: SNV, Savitsky-Golay 2nd derivative, 15-point window). Classes: low-FN <355 s, high-FN >355 s. Dashed ellipse encapsulates group 1 samples. FN, falling number

3.3 | Why the limitation

Preharvest sprouting is the most common cause of elevated α-amylase levels in wheat grain. Rainy conditions before harvest can lead to the onset of seed germination, whereupon gibberellic acid is released by the embryo which then initiates the de novo synthesis of α-amylase (Singh & Kayastha, 2014). A possible reason why the bulk spectra produced poorer NIR models was the lack of homogeneity of the scanned seeds compared to that of the ground meal. Under mildly unfavorable weather conditions that cause a very small fraction of seeds to break dormancy, these seeds are randomly distributed throughout a bulk sample during the NIR scanning operation. The resulting spectrum will reflect the kernel-to-kernel spatial variability of the chemical constituents, including α-amylase. Conversely, the grind process produces a more homogeneous material that is presented to the spectrometer, such that the α-amylase that was localized to the germinated kernels is now spread throughout the meal. However, the underlying assumptions to this postulate are that (i) α-amylase is at a sufficiently high concentration to be detectable by NIR spectroscopy and (ii) the molecular structure of α-amylase is spectroscopically distinguishable from that of the endosperm storage proteins, the overwhelming source of seed protein, as well as that from proteins contained in the embryo and aleurone layer. The isoform of the cereal α-amylase molecule synthesized during germination, and best studied in barley, consists of 403 amino acid residues folded into three domains, with the

largest domain containing 286 residues in a (βα)8-barrel formation which houses the active-sites for starch hydrolysis (Kadziola, Abe, Svensson, & Haser, 1994). NIR spectroscopy is based on a measurement of the combination and overtone band vibrational frequencies arising from interatomic bonds of atoms of low mass such as C, N, O, and H. Although collectively sensitive to protein by a conglomeration of the N-H bond vibrations, the sensitivity of the NIR response to individual amino acids is quite challenging. As challenging is the feat of assigning spectral characteristics of amino acids to protein units, be they endosperm storage proteins, embryo proteins, or enzymes. With these conditions taken together, it is not surprising that FN inferred by NIR spectroscopy is tenuous at best.

4 | CONCLUSIONS

Based on the evaluation of a diverse set of soft white wheat, NIR spectroscopy models for FN, either in the form of quantitative prediction of the value or a qualitative prediction of a low–high decision around a cutoff, are not sufficiently accurate to act as replacements for the method itself.

ACKNOWLEDGMENTS

Authors would like to thank Tracy Harris and Steven Rausch for expert technical assistance. This work was funded by the USDA-ARS and by the Washington Grain Commission.

REFERENCES

Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2017). Application of calibrations to hyperspectral images of food grains: Example for wheat falling number. *Journal of Spectral Imaging*, *6*(a4), 1–15.

Delwiche, S. R., Vinyard, B. T., & Bettge, A. D. (2015). Repeatability precision of the falling number procedure under standard and modified methodologies. *Cereal Chemistry*, *92*, 177–184. https://doi.org/10.1094/CCHEM-07-14-0156-R

Farrell, A. D., & Kettlewell, P. S. (2008). The effect of temperature shock and grain morphology on alpha-amylase in developing wheat grain. *Annals of Botany*, *102*, 287–293. https://doi.org/10.1093/aob/mcn091

FGIS. (2013). *Falling number determination for wheat*. Directive 9180.38. Online publication. U.S. Department of Agriculture Grain Inspection, Packers and Stockyards Administration, Federal Grain Inspection Service. Retrieved from http://gipsa.usda.gov/laws/directives/9180-38.pdf.

Higginbotham, R., Jitkov, V., & Horton, A. (2016). *WSU extension variety testing program: Cereal variety performance trials 2016 annual report*. Retrieved from http://smallgrains.wsu.edu/wp-content/uploads/2015/10/2016-VT-Annual-Report.pdf.

Kadziola, A., Abe, J. I., Svensson, B., & Haser, R. (1994). Crystal and molecular structure of barley α-amylase. *Journal of Molecular Biology*, *239*, 104–121. https://doi.org/10.1006/jmbi.1994.1354

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

Osborne, B. G. (1984). Investigations into the use of near infrared reflectance spectroscopy for the quality assessment of wheat with respect to its potential for bread baking. *Journal of the Science of Food and Agriculture*, *35*, 106–110. https://doi.org/10.1002/(ISSN)1097-0010

Reeves, J. B. III, & Delwiche, S. R. (2003). SAS® partial least squares regression for analysis of spectroscopic data. *Journal of Near Infrared Spectroscopy*, *11*, 415–431. https://doi.org/10.1255/jnirs.393

Risius, H., Hahn, J., Huth, M., Tölle, R., & Korte, H. (2015). In-line estimation of falling number using near-infrared diffuse reflectance spectroscopy on a combine harvester. *Precision Agriculture*, *16*, 261–274. https://doi.org/10.1007/s11119-014-9374-5

Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*, 1627–1639. https://doi.org/10.1021/ac60214a047

Singh, K., & Kayastha, A. M. (2014). α-Amylase from wheat (*Triticum aestivum*) seeds: Its purification, biochemical attributes and active site studies. *Food Chemistry*, *162*, 1–9. https://doi.org/10.1016/j.foodchem.2014.04.043

Starr, C., Morgan, A. G., & Smith, D. B. (1981). An evaluation of near infra-red reflectance analysis in some plant breeding programmes. *The Journal of Agricultural Science*, *97*, 107–118. https://doi.org/10.1017/S0021859600035929

Williams, P. C., & Sobering, D. C. (1993). Comparison of commercial near-infrared transmittance and reflectance instruments RPD for analysis of whole grains and seeds. *Journal of Near Infrared Spectroscopy*, *1*, 25–32. https://doi.org/10.1255/jnirs.3

Xing, J., Symons, S., Hatcher, D., & Shahin, M. (2011). Comparison of short-wavelength infrared (SWIR) hyperspectral imaging system with an FT-NIR spectrophotometer for predicting alpha-amylase activities in individual Canadian Western Red Spring (CWRS) wheat kernels. *Biosystems Engineering*, *108*, 303–310. https://doi.org/10.1016/j.biosystemseng.2011.01.002

How to cite this article: Delwiche SR, Higginbotham R, Steber CM. Falling number of soft white wheat by near-infrared spectroscopy: A challenge revisited. *Cereal Chem*. 2018;00:1–9. https://doi.org/10.1002/cche.10049