

An independent validation reveals the potential to predict Hagberg–Perten falling number using spectrometers

Chun-Peng James Chen¹  | Yang Hu²  | Xianran Li³  | Craig F. Morris⁴ | Stephen Delwiche⁵ | Arron H. Carter²  | Camille Steber^{2,3}  | Zhiwu Zhang² 

¹School of Animal Sciences, Virginia Tech, Blacksburg, Virginia, USA

²Department of Crop and Soil Sciences, Washington State University, Pullman, Washington, USA

³USDA-ARS, Wheat Health, Genetics, and Quality Research Unit, Pullman, Washington, USA

⁴USDA-ARS, Western Wheat Quality Laboratory, Pullman, Washington, USA

⁵USDA-ARS, Beltsville Agricultural Research Center, Food Quality Laboratory, Beltsville, Maryland, USA

Correspondence

Chun-Peng James Chen, School of Animal Sciences, Virginia Tech, Blacksburg, VA 24061, USA.

Email: niche@vt.edu

Assigned to Associate Editor Valerio Hoyos-Villegas.

Funding information

Washington Grain Commission, Grant/Award Numbers: 126593, 134574; National Institute of Food and Agriculture, Grant/Award Numbers: 2019-67013-29171, 2020-67021-32460, 1014919

Abstract

The Hagberg–Perten falling number (HFN) method is the international standard used to evaluate the damage to wheat (*Triticum aestivum*) grain quality due to preharvest sprouting (PHS) and late maturity alpha-amylase (LMA). However, the HFN test requires specialized laboratory facilities and is time consuming. Spectrometers were known as a potential tool for quick HFN assessment, but none of the studies have validated the assessment results across different datasets. In this study, an independent validation was conducted using independent samples and spectral instruments. The calibration set had 462 grain samples of 92 varieties grown at 24 locations in 2019 and examined using a near-infrared spectrometer. In the validation set, 19 varieties collected from 10 locations in 2 years that experienced either PHS or LMA were scanned with a hyperspectral camera. The association between spectra and HFN was modeled by partial least square regression. As a result, the independent validation correlation accuracy was $r = 0.72$ and a mean absolute error of 56 s. Furthermore, this study showed a cost-effective alternative using only 10 spectral bands to predict HFN, and it achieved better performance than the full spectrum of the hyperspectral system. In conclusion, this is the first study that showed the potential that wheat HFN could be predicted on an independent dataset measured by a different instrument. The result suggested that spectrometers can potentially serve as a faster alternative for plant breeders to develop varieties resistant to PHS and LMA, and for growers to screen damaged grains in transportation processes.

Abbreviations: BLINK, Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway; GLM, generalized linear model; HFN, Hagberg–Perten falling number; HSI, hyperspectral image; LMA, late maturity alpha-amylase; NIR, near-infrared; PHS, pre-harvest sprouting; PLS, partial least square; PLSR, partial least square regression; WL, wavelength; WN, wavenumber.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *The Plant Phenome Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy and Crop Science Society of America.

1 | INTRODUCTION

The Hagberg–Perten falling number (HFN) test was developed to assess α -amylase activity in breaking down wheat (*Triticum aestivum*) starch for the wheat industry (Hagberg, 1960, 1961). This breakdown, attributed to sticky flour, can cause problems for manufacturing pipelines and poor quality of end products (reviewed by Bettge, 2018; Ross & Bettge,

2009). As a result, wheat flour with a low HFN can lead to large economic losses in an industry that is already structured on a low-profit margin. In Australia, low HFN led to a loss of up to \$40 per ton for Australian wheat harvested in 1983 and 1984 (Bettge, 2018; Ross & Bettge, 2009). In the western United States, the wheat grain industry lost \$140 million due to low HFNs in 2016 (Campbell, 2017).

Pre-harvest sprouting (PHS) and late-maturity α -amylase (LMA) are two major causes of low HFN. Pre-harvest sprouting is induced by rain or high humidity before harvest. Once the required moisture level is achieved, the germination process will be triggered, producing hydrolytic enzymes to break down starch and provide energy to support the germination (Mares & Mrva, 2014). Late-maturity α -amylase is a response to temperature shock during the maturation phase of grain development without visible sprout damage (Farrell & Kettlewell, 2008; Mares & Mrva, 2008). Both cold shock (Gale et al., 1987; Major, 1999) and heat shock (Major, 1999; Randall & Moss, 1990) have been reported to trigger LMA.

Due to these significant economic losses, the HFN test has recently come under increased scrutiny. The time-consuming nature of the traditional HFN test (typically 10 min between successive analyses) only allows a small portion of grain to be tested; such limitation can misrepresent the damage of a whole field or a load. One alternative to the HFN test is an enzyme-linked immunosorbent assay (ELISA) that detects α -amylase in wheat grain extracts using antibodies specific to α -amylase isozymes (Skerritt & Heywood, 2000; Verity et al., 1999). The ELISA approach can be faster than the HFN test, but it still requires a laboratory setting and is not suitable for field use. Besides, the ELISA test was reported to have batch-to-batch errors and is not reproducible in a new batch of samples (Neoh et al., 2021).

Given the difficulty in measuring HFN, developing a non-destructive, rapid, and accurate method to measure FN is highly desirable. Spectroscopy has been proposed to be the promising alternative that meets these requirements. Specific near-infrared (NIR) wavelength ranges, which cover from 850 to 1700 nm, have been reported to be able to characterize different levels of sprouted wheat kernels (Chen et al., 2014). Though there is no direct discussion of the connection between the spectral characteristics and low-HFN conditions (i.e., LMA or PHS) in this study, this work indicates the potential of NIR spectroscopy to assess the sprouting damage of kernels, which is an important factor in reducing HFN. In addition to in-lab studies, one study has investigated the in-field performance of a NIR sensor. The results suggested a correlation (r) accuracy of 0.84 and a standard error of 37 s in fitting a model predicting HFN (Risius, 2014). Besides NIR spectroscopy, imaging techniques, which have 2D spatial information, have also been studied to identify kernels with low HFN. X-ray imaging (Neethirajan et al., 2017) and thermal imaging (Vadivambal et al., 2011) systems were shown

Core Ideas

- This study is the first to validate a calibrated spectroscopy model of Hagberg–Perten falling number on an independent dataset.
- Spectral selection is a cost-effective strategy for replacing hyper spectrometers in predicting falling numbers.
- The unsupervised selection strategy presented in this study performed better than the supervised approaches.
- The presented model has a validation accuracy of $r = 0.72$, but a high bias of 56 seconds.
- Application of the spectroscopy model to a hyperspectral image of wheat kernels revealed biological insights.

to identify sprouted kernels yielding a correlation accuracy of 0.9 on kernel density and surface temperature, respectively. However, these imaging techniques are not suitable for field use due to the high cost and the need for a trained operator. Thanks to the development of sensor technology, hyperspectral imaging (HSI) systems, where every image pixel has a full spectral signature consisting of hundreds or thousands of spectral bands, have become more affordable and have been considered as an approach to extract more complex information. Many recent studies have demonstrated HSI systems' advantages in predicting FN-related traits. Xing et al. (2011) used a short-wavelength infrared HSI system that covers from 1000 to 2500 nm to predict α -amylase enzyme activity with a high correlation accuracy of 0.94. The classification of sprouting damage in wheat kernels was also reported with good accuracy using the HSI system (Armstrong et al., 2016; Zhang et al., 2020). Additionally, the direct association between the HSI-derived characteristics and Hagberg FN was investigated (Barbedo et al., 2018; Caporaso et al., 2017; Delwiche et al., 2018). Barbedo et al. divided the collected wheat samples into five groups with HFN ranging from 0 to 70 s, 70 to 150 s, 150 to 250 s, 250 to 350 s, and 350 s and above, respectively. The HFN groups were distinguishable under the HSI system covering from 528 to 1785 nm (Barbedo et al., 2018). Caporaso et al. (2017) also reported a promising calibration accuracy (r) of 0.77 using an HSI system with a more comprehensive spectral range (900–2500 nm). However, Delwiche et al. (2018) pointed out that good prediction results can only occur in single or multiple homogenous environments. The results were hardly validated across environments with different climatic conditions.

Therefore, this is the first study that validated the potential of using spectrometers to predict HFN on an independent

dataset. The independent samples for the validation were collected by a different spectral instrument that measures hyperspectral images. We specifically addressed three objectives in this study: (1) the transferability of an HFN calibrated model in a new environment, new variety, and a different spectral instrument; (2) the model performances using a subset of bands; (3) the spectral patterns of kernels with different low-HFN conditions in hyperspectral images. The results of this study will provide a better understanding of the potential of leveraging spectroscopy techniques to predict HFN in wheat kernels.

2 | MATERIALS AND METHODS

2.1 | Grain samples

The calibration and validation samples were collected from the multi-environment wheat variety trials, which were conducted by Washington State University Extension Cereal Variety Testing and USDA-ARS (www.smallgrains.wsu.edu, www.steberlab.org). The calibration dataset included 462 samples of unbalanced combinations from 92 varieties and 24 locations grown in 2019. Each sample had three NIR-scanned replicates, and each replicate contained 2 g of wheat kernels.

To independently validate the model predictability, 39 samples were collected from different locations and years to form a validation dataset. The validation samples comprised 19 varieties and 10 locations grown in 2018 and 2019. Half-grain α -amylase assays with the PhadebasTM Amylase Assay Kit (Pharmacia) were used to determine the cause of reduced HFN in validation samples. The assays are microspheres chemically bound to a blue dye, which will be released when the microspheres are hydrolyzed by α -amylase. Measuring the absorbance of the blue solution is a measure of the α -amylase enzymatic activity. Based on past studies, PHS-affected grain had much higher α -amylase activity at the embryo end of the grain, and LMA-affected grain had similar α -amylase levels at both ends (Mares et al., 1994; Mrva et al., 2006). The presence of visible sprouting was also used to identify PHS-affected grain. The details of the locations, years, and varieties of both datasets are shown in Tables S1 and S2.

2.2 | Acquisition of spectral information

Hyperspectral information for the calibration and validation datasets was collected using whole kernels (Figure 1). In the calibration set, each sample consisted of 2 g of kernels and was scanned in a glass jar (Figure 1a). Three jars of kernels were scanned as replicates. The scans were conducted by a Fourier Transform Near-Infrared spectroscopy instrument

(MATRIX-I, Bruker optics), which covered a wavenumber range of 12,500 cm^{-1} (800 nm) to 3600 cm^{-1} (2778 nm) and had an average spectral resolution of 1.71 nm resulting in 1154 spectral spectra. The measurement results were exported by Spectral Acquisition and Processing Software (OPUS 7.2, Bruker optics), and the data dimension was 1386 (i.e., three replicates of 462 samples) data points by 1154 spectra.

In the validation set, spectral information was collected using an HSI system assembled in the USDA Beltsville laboratory and described in (Delwiche et al., 2019). The hyperspectral imaging system consisted of an imaging spectrograph (SWIR Hyperspec, Headwall Photonics), an InGaAS focal-plane array camera (320 \times 256 pixels, 14-bit A/D, Model Xeva-1.7-320, Xenics) with a 25-mm zoom lens (Optec, Model OB-SWIR25/2), two glass fiber (low-OH) optic bundles for directing light from separate DC-regulated 150 W quartz tungsten halogen light sources (Dolan Jenner, Model DC-950) to the imaging enclosure, and a stepper motor movable stage (Velmex, Model XN10-0180-M02-21) that moved the seeds across the field of view. At an average wavelength spacing of 4.8 nm, 150 spectral wavelengths were recorded and spanned a wavelength range of 936–1654 nm. The spectral data are stored in a hyperspectral image per sample with a dimension of 320-pixel width \times 540-pixel height \times 150 spectra. Each sample image contains 4 g of \sim 200 kernels (Figure 1b). A contour-searching algorithm (Bradski, 2000) was used to identify kernels in the images. And a 150-band vector representation of each kernel was obtained by averaging all pixel values in the kernel-contour region.

As the two spectrometers had different spectral resolutions and ranges, three adjustments were made to make the two datasets comparable. First, the two datasets were unified by truncating the spectra to the spectral range of 1000–1654 nm shared by both datasets. The spectra between 936 and 1000 nm, which show fluctuating signals over another spectral region, were excluded to avoid noise interference. Second, the calibration and validation dataset measures signals in uniform wavenumber (WN) increments (cm^{-1}) and uniform wavelength (WL) increments (nm), respectively. To unify the units, the WN reads were converted from the calibration dataset to WN by a simple conversion formula:

$$\text{WL in nm} = 10^7 / \text{WN in cm}^{-1}$$

Last, to deal with different spectral resolutions, every 2 nm wavelength was binned to one spectrum. Because the spectrometer in the calibration dataset has a higher resolution than the validation dataset, which missed some wavelength spectra that only exist in the spectral region of the calibration dataset. The missing spectra were represented by the nearest wavelength spectra in the validation dataset. With the three adjustments, the two datasets can share the same number of spectral features (327 spectra) and range (1000–1654 nm).

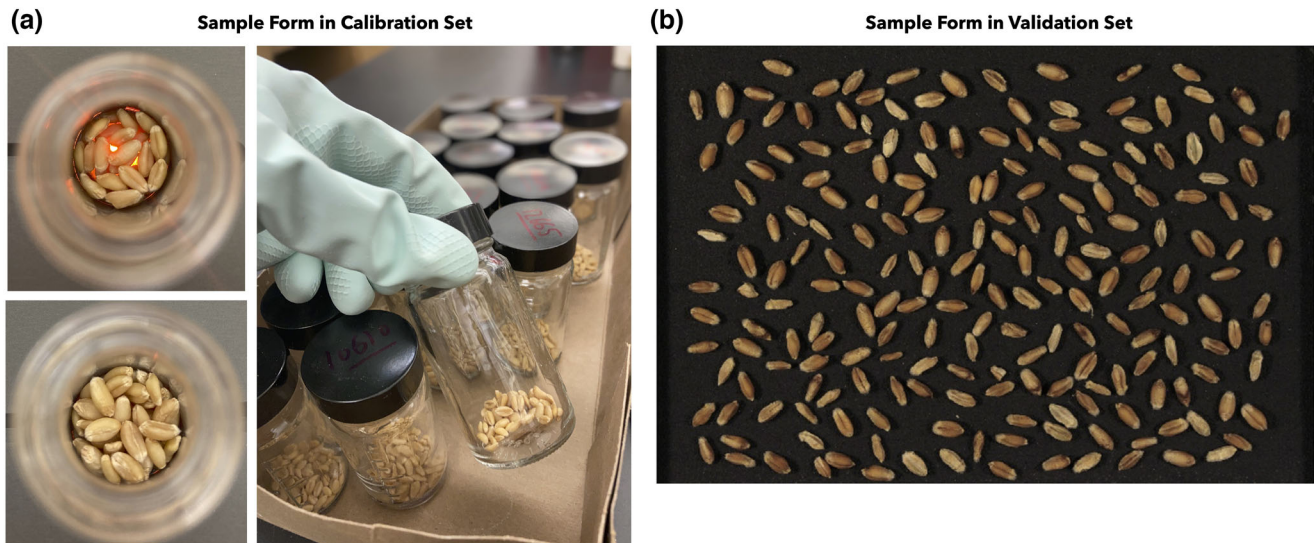


FIGURE 1 Wheat samples in the datasets. (a) Samples in the calibration set were stored in a glass jar, and each jar was considered a sample replica. Each replicate was scanned to acquire hyperspectral values and saved as a data point. (b) The validation set had kernels placed on emery cloth and imaged by hyperspectral cameras.

2.3 | Independent validation

The independent validation was conducted by first calibrating the prediction model on the calibration set, and the model was then used to predict the validation set where the validated samples were collected in different environments or varieties. This validation was meant to evaluate the transferability of the calibrated HFN model. The validation result was evaluated by Pearson's correlation coefficient (r) and mean absolute error (MAE). In addition to validating the model performance using the entire dataset, the evaluation was also conducted on the subgroups of the datasets. The subgroups were defined by the low-HFN conditions (i.e., LMA or PHS), environments (i.e., location-year pairs), and varieties. The subgroup validation was able to evaluate if the model performance was based on the HFN or the potential confounding factors, such as the environment or variety.

2.4 | Feature selection and permutation test

Except for the full model using all the available bands, we also validate the performance of reduced models, where only 10 key bands were selected to be predictors of HFN. We conducted supervised and unsupervised strategies to select the key bands based on the calibration set. With the selected bands, we derived a reduced model and evaluated its performance on the validation set. There were two supervised strategies in this study: The first strategy was to use a generalized linear model (GLM) to fit each band in a least square regression with HFN as the response variable. The regression coefficients of the fitted band have tested the

hypothesis that it was zero, which indicates that the fitted variable has no significant association with HFN. This strategy only considered a single-band effect on HFN without taking collinearity between spectral bands into account; hence the selected bands may be correlated with each other. The second supervised strategy was Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK; Huang et al., 2019). Linkage-disequilibrium Iteratively Nested Keyway is a stepwise regression that tests a single band similarly to GLM. However, BLINK effectively resolved the collinearity problem by fitting the selected band from the first iteration as linear covariates. It retested the association between the band and HFN in the following iterations. With this BLINK strategy, the selected bands were not only associated with HFN but also had reduced dependency among predictors.

On the other hand, three unsupervised selection strategies were conducted. The first strategy was to split the full spectrum into 10 bins with equal wavelength ranges, and the band located in the middle of the bin was selected. The strategy simulated the bands measured by a spectrometer with a lower spectral resolution. We named this strategy Low-Res hereafter. The second strategy used Ward's agglomerative clustering algorithm (Ward, 1963) to derive 10 groups from the spectrum. Similar to the first strategy, the band located in the middle wavelength position of each group was selected. This clustering algorithm clustered bands based on the Euclidean distance among samples. Because it is a hierarchical clustering algorithm, where only adjacent spectra are clustered into the same group, the clustering algorithm was implemented using the function *AgglomerativeClustering* in the Python library *scikit-learn* (Pedregosa et al., 2011). The

third unsupervised strategy derived the second derivatives of reflectance with respect to wavelength, quantifying the signal change over the spectrum. The top 10 spectral bands with local maxima in the second derivative were selected. Local maxima were identified by a peak searching algorithm provided by the R package, *pracma* (Borchers, 2021). This approach was noted as second derivative hereafter.

The five selection strategies were examined by a permutation test. This is a statistical test to examine whether the proposed approach is significantly better than random selection. We derived a reduced model that used 10 randomly selected bands as predictors. The model was then used to predict the validation set. The process was repeated 10,000 times, and the 10,000 prediction accuracies formed an empirical null distribution. The prediction accuracies of the full model and the five different reduced models were tested against the null distribution. If the proposed model performs better than 9500 out of 10,000 random selections (p value < 0.05), the selection strategy was considered significantly better than random selection. The permutation test was implemented using the function *permutation_test_score* in the Python library *scikit-learn* (Pedregosa et al., 2011).

2.5 | Partial least square regression

Partial least square regression (PLSR; Wold, 1980) is a method that transforms data into latent dimensions where the covariance of the transformed regressors (i.e., spectral bands) and the transformed response variables (i.e., HFN) are maximized. The advantage of this method is that it can reduce the size of feature space to avoid large- p and small- n problems (Marimont & Shapiro, 1979), and it can remove the redundancy among feature space. It is especially useful with hyperspectral data that have strong collinearity. We used the same notations from Abdi (2003) to describe the PLSR model:

$$\hat{Y} = X (P^{T+}) BC^T$$

where $\hat{Y} \in \mathbb{R}^{n \times 1}$ is the estimated response matrix and n is the number of HFN observations; $X \in \mathbb{R}^{n \times p}$ is the regressor matrix, where p is the number of fitted spectral bands; $P \in \mathbb{R}^{p \times q}$ is the loading matrix of X , where q is the number of latent dimensions; P^{T+} is Moore-Penrose pseudo-inverse of P^T ; $B \in \mathbb{R}^{q \times q}$ is the diagonal matrix of the regression coefficients; and $C \in \mathbb{R}^{1 \times q}$ is the loading matrix of Y , which is the original response matrix. In this study, p was 327 in the full model and 10 in the reduced models. We transformed the original regressor matrix X to the first 2 latent dimensions, so the $q = 2$ in our study. In the calibration stage, the regressors $X_{cali} \in \mathbb{R}^{1386 \times p}$ from the calibration set were first standardized to have zero mean and unit variance. The standardization scaler was stored and used to standardize the

regressors from the validation set. Then, we fitted the X_{cali} and $Y_{cali} \in \mathbb{R}^{1386 \times 1}$ to the PLSR model to find the optimal parameters of B , C , and P . In the validation stage, we used the fitted parameters to estimate the $Y_{valid} \in \mathbb{R}^{39 \times 1}$ from the $X_{valid} \in \mathbb{R}^{39 \times p}$. The PLSR model was implemented using the function *PLSRRegression* in the Python library *scikit-learn* (Pedregosa et al., 2011).

Additionally, we applied the calibrated PLSR model to map the HFN in the hyperspectral images in the validation set. Since every pixel is a spectrum of 327 bands (i.e., $\mathbb{R}^{1 \times 327}$), we used the same calibrated PLSR model to estimate the HFN for every pixel. The pixel-level estimation of HFN allowed us to visualize the spatial pattern of HFN on the kernel surface.

3 | RESULTS

3.1 | Data characteristics

The hyperspectral bands and HFN of both datasets are presented in Figure 2. We grouped the spectra based on whether the sample had a measured HFN greater than 300 seconds, and the spectra were found to be mostly overlapping between the two groups (Figure 2a). The distribution of HFN in the calibration set was skewed to the left, with a mean of 306.4 seconds and a standard deviation of 53.78 seconds. And it covers a range from 63 to 460 seconds. On the other hand, the validation set showed a smaller range of HFN from 107 to 445 seconds, and the mean and standard deviation were 256.6 seconds and 81.31 seconds, respectively (Figure 2b). By examining the low-HFN conditions in the validation set, it was found that 8 of 39 samples were affected by PHS, 21 were affected by LMA, and 10 were not affected by either PHS or LMA (noted as “sound” thereafter) (Table S2).

Additionally, the validation set was geographically mapped in the map of Washington State, USA (Figure 3). Besides the geography information, the validation samples were listed for their varieties, observed HFN, errors of predicted HFN, and low-HFN conditions. Most PHS samples were found in Creston in 2019, and the rest two PHS samples were collected from Ritzville in 2018. Unlike other locations where only one low-HFN condition, either PHS or LMA, was observed, both types of low-HFN samples existed in the location of Ritzville. In Ritzville, different varieties were found to have different low-HFN conditions. For example, the variety *Jasper* had two samples with PHS, and the variety *KWS 147* was affected by LMA. This result suggested a possibility that the low-HFN conditions were not only affected by the environment but also by the variety.

Combining both datasets, we applied principal component analysis to obtain the first two principal components (PCs) explaining 98.76% of the total variance in the hyperspectral data. The first two PCs were used to represent the data in a

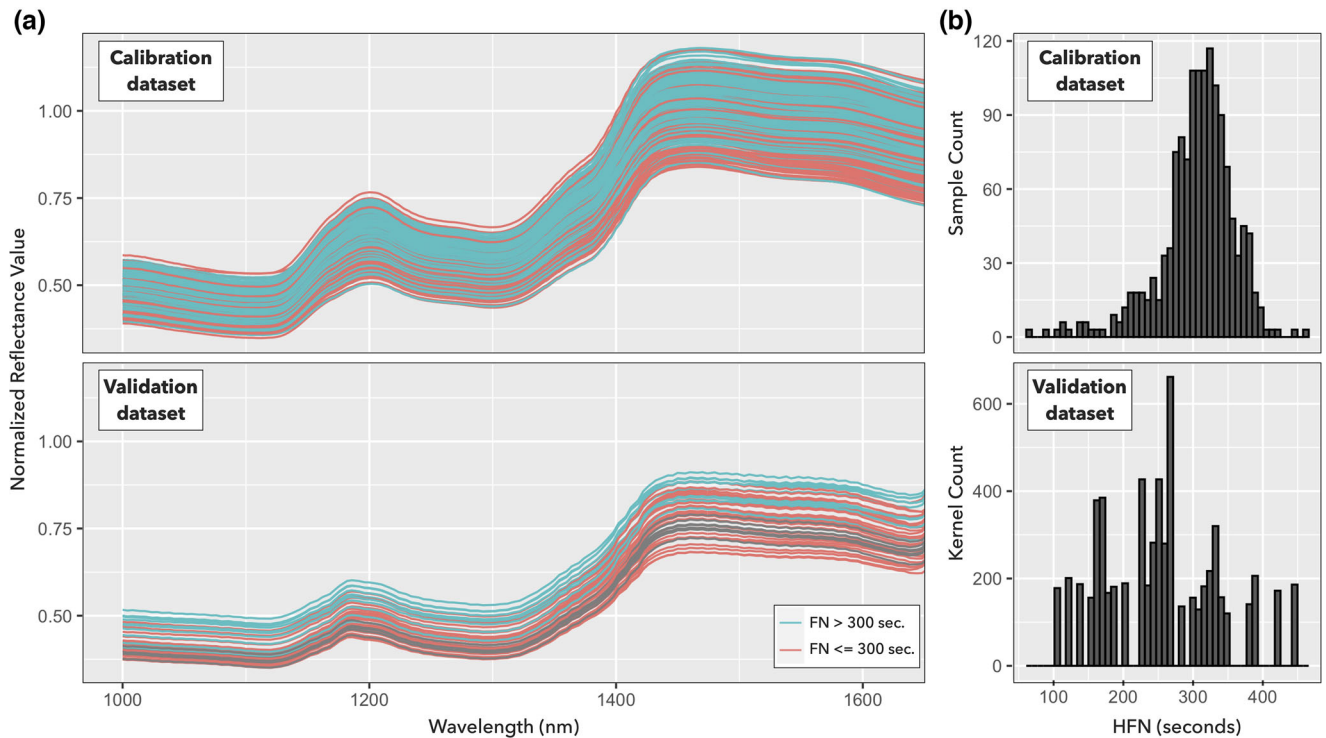


FIGURE 2 Overview of the dataset. (a) The line chart displayed the averaged spectrum of each sample represented by a line. There were 462 lines in the calibration set (top) and 41 lines in the validation set (bottom). Lines were colored in blue if the measured Hagberg–Perten falling number (HFN) was greater than 300 seconds. Otherwise, the line was colored red. The wavelength coverage was from 1000 to 1652 nm. (b) The histogram showed the HFN distribution among calibration (top) and validation (bottom) sets.

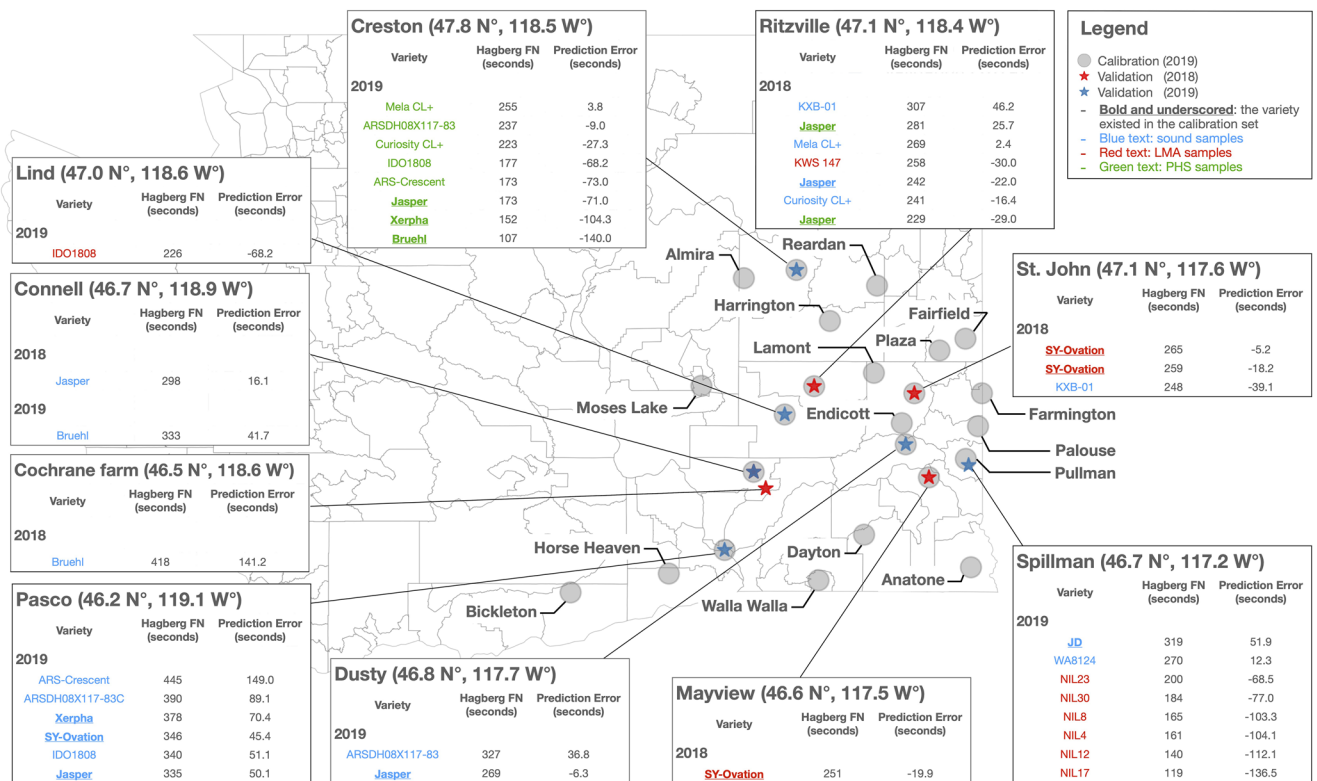


FIGURE 3 Geographic distribution of the validation set in Washington State.

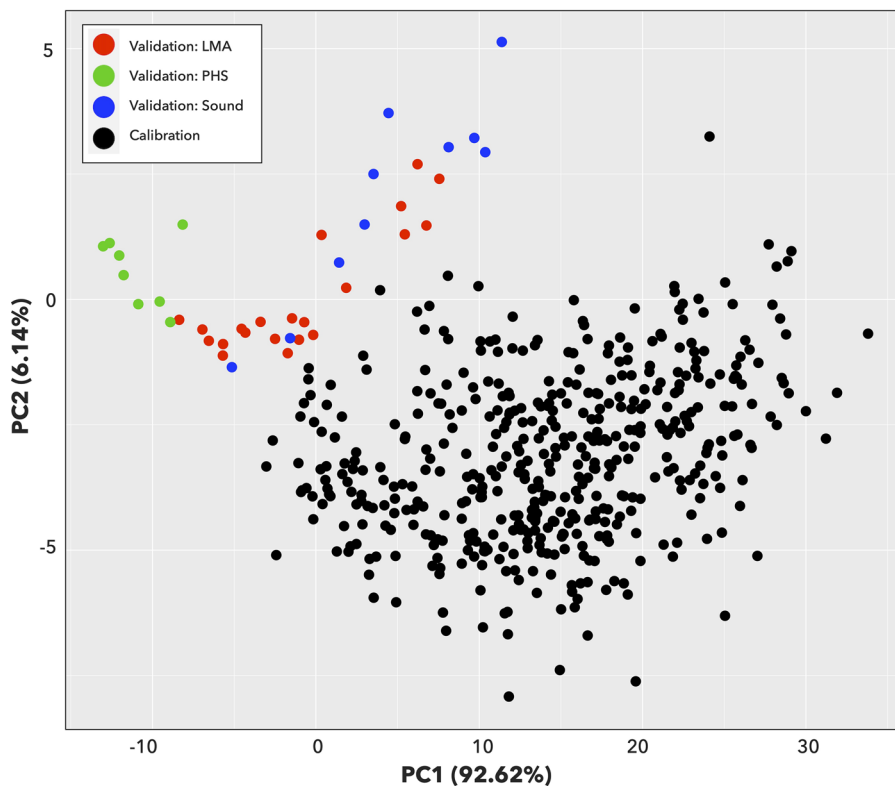


FIGURE 4 Principal component analysis (PCA) to visualize the spectral similarity among samples. Spectral values of each sample were mapped as a dot on the x axis of the first principal component and the y axis of the second component. A sample was defined for kernels collected from the same variety, same location, and same year. Variance explained by the component was shown in the parenthesis of each axis. Black dots represented the samples from the calibration dataset. The remaining dots were from the validation set. They were colored red, green, and blue to show the observed conditions of late maturity alpha-amylase (LMA), pre-harvest sprouting (PHS), and sound condition, respectively.

2D space, where the hyperspectral band of each sample was represented by a point (Figure 4). In the space of the two PCs, two datasets were found to be linearly separable, which suggested that the different instruments would affect the spectral pattern. In the validation set, the LMA samples were clustered within the sound samples but separated from the PHS samples, which could suggest that the spectral characteristics of the LMA samples were similar to the sound samples. We also examined the autocorrelation of the spectrum in both datasets. Both spectrums showed a two-cluster autocorrelation pattern: the first cluster covered from 1000 to 1400 nm, and the second cluster covered from 1400 to 1654 nm. Both the clusters showed a strong autocorrelation higher than 0.95 (r), and the lowest correlation between the clusters was 0.82 (r) (Figure S1).

3.2 | Independent validation

The full model calibrated on all 327 bands by a two-component PLSR model achieved the validation accuracy (r) of 0.72. To avoid overestimating the accuracies, we further examined the validation in different group settings; the val-

idation samples were grouped by low-HFN conditions (i.e., LMA, PHS, and sound samples), environments, and varieties. The accuracies in each group indicated that the prediction model could explain the FN variation which was not derived from either low-HFN conditions, environments, or varieties. For example, if the correlation accuracies were both low in the groups of PHS and LMA, it would suggest that the prediction model was only capable of differentiating the conditions but not the HFN itself. As a result, the full model performed similarly on LMA ($r = 0.81$) and sound samples ($r = 0.73$). But the accuracy was lower in the PHS samples, which was 0.39 (Figure 5). When the samples were grouped by the environment, except for the groups with less than three observations, the accuracies were below 0.5 (Figure S2). Last, most variety groups show high accuracies ($r > 0.5$). However, there were only 5 of 19 variety groups with more than two observations; this group setting needs more samples to be validated (Figure S3).

Besides the correlation accuracy, the prediction error was also examined. The MAE was 56 s; we performed an Analysis of Variance (ANOVA) table to investigate further which experimental factors contributed significantly to a higher error. As a result, the error was found to be significantly

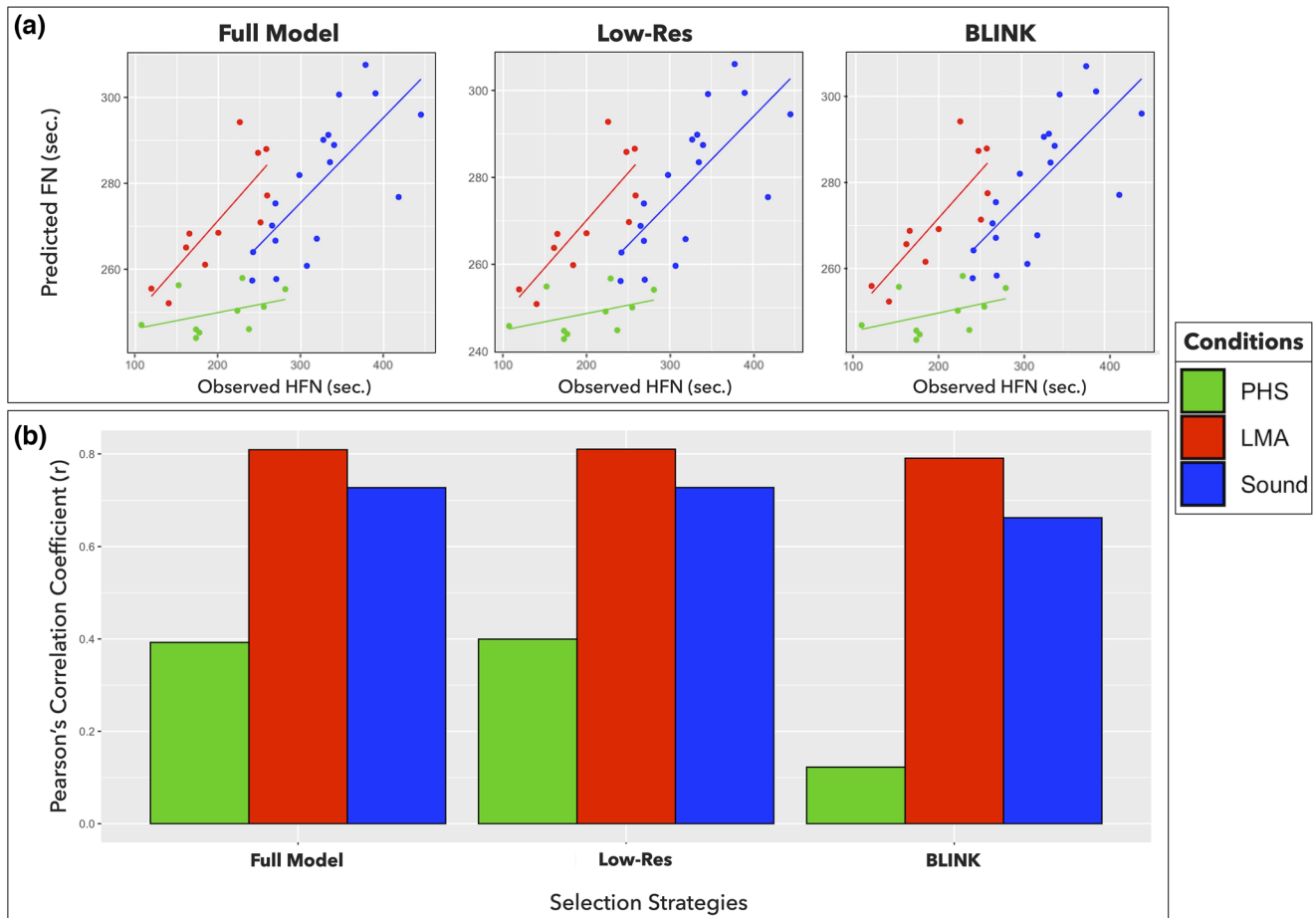


FIGURE 5 The prediction accuracies of the selected models in the independent validation (color-coded by Hagberg–Perten falling number (HFN)). The validation samples were colored green, red, and blue representing the condition preharvest sprouting (PHS), late maturity alpha-amylase (LMA), and sound kernels, respectively. The results of the full model, the best-unsupervised model, Low-Res, and the best-supervised model, Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK), were selected to demonstrate. (a) The predicted and observed HFN were plotted in scatter plots. The regression lines were fitted by linear least squares. (b) The bar charts showed the correlation accuracy within each condition.

affected by the observed HFN (p value < 0.001); as the HFN increased, the error also increased (Figure 6). We also explored two other potential factors that may affect the error: (1) Whether the validated sample existed in the same year and as the same variety of the calibration set. (2) The latitude and longitude of the validated location. Surprisingly, the result showed that there was no significant difference in the error between the samples with the same year or variety with the calibration set and those with different years and variety. But the error was associated with the latitude (p value < 0.005 ; Table 1).

3.3 | Selection of key bands and the reduced models

The validation was also evaluated by the reduced model, which is a PLSR model using only 10 selected bands as pre-

dictors. The five selection strategies reported different sets of key bands which were visualized in Figure 7. The Low-Res strategy selected bands every 66 nm starting at 1030 nm (Figure 7a). The clustering strategy indicated which wavelength regions had variation among observed samples; there were four clusters in the small range from 1347 to 1409 nm. Two clusters were found from ~ 1140 to 1230 nm (Figure 7b). This result implies that these two regions may be distinct in each sample but not necessarily associated with FN variation. The 2nd derivatives (Figure 7c) were designed to find curvature in spectral signals with respect to wavelength. There were four peaks found to have optimum derivatives in the regions from 1134 to 1240 nm and from 1386 to 1424 nm. Generalized linear model, the fourth strategy, focused on the association between the linear effects of a single spectral band and the FNs. The tested score of each band was plotted as a curve line (Figure 7d). Using this strategy, all significant hits were in the range of 1344–1365 nm. Due to the strong

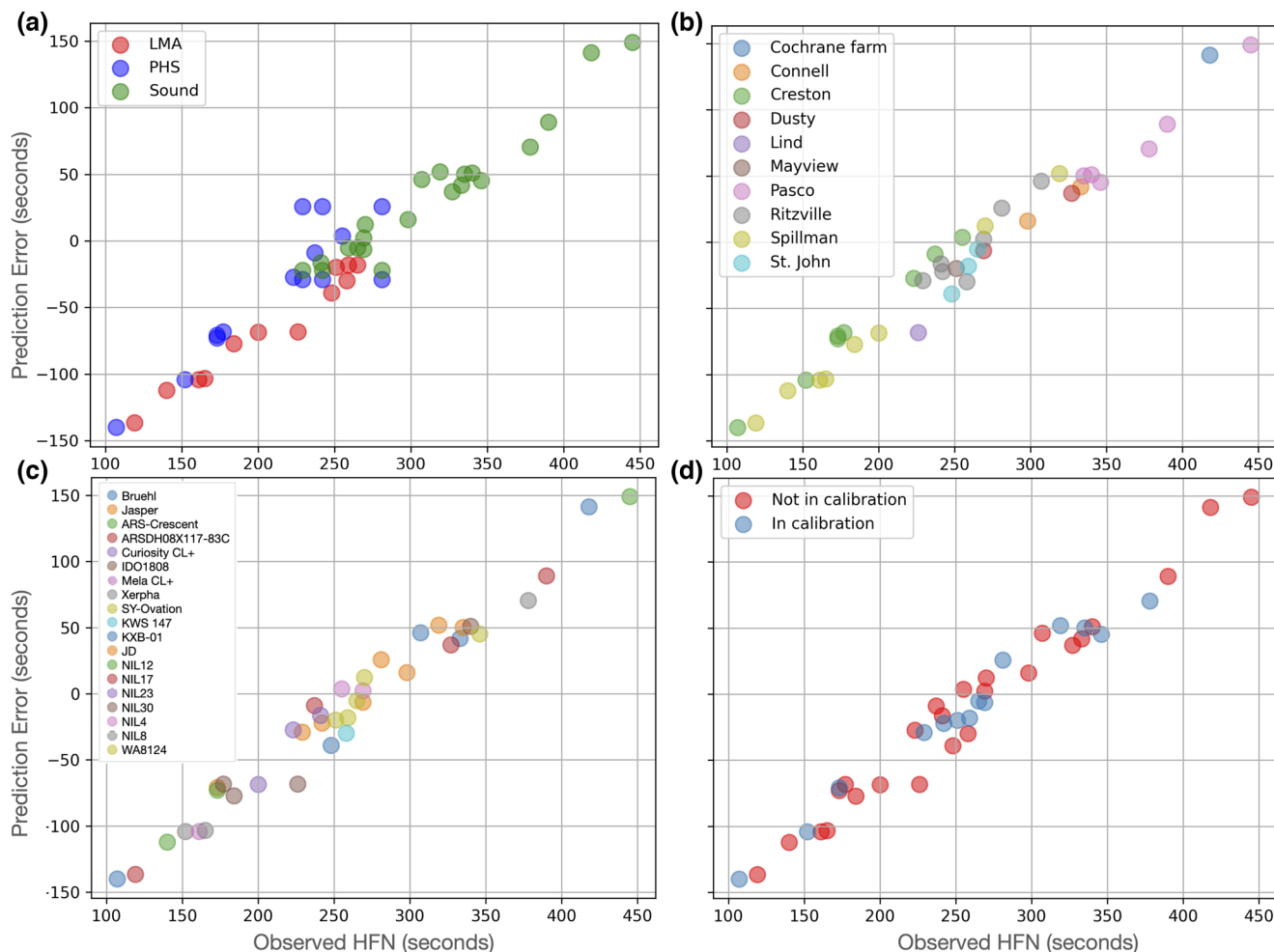


FIGURE 6 The prediction errors of the full model. The samples were colored by (a) the low-Hagberg–Perten falling number (HFN) condition, (b) location, (c) variety, and (d) whether the sample variety was from the calibration set.

TABLE 1 Analysis of variance (ANOVA) of the validation error.

	Sum of squares	Degrees of freedom	<i>F</i>	<i>p</i> value (> <i>F</i>)
in_calibration	1657.335792	1	1.091189	0.303578
FN	12,552.612528	1	8.264632	0.006926
Lat	12,808.818375	1	8.433318	0.006430
Lon	5758.498572	1	3.791392	0.059815
Residual	51,640.389822	34	–	–

Note: The error variance was decomposed into a binary factor of whether the sample was the same variety from the calibration set (in_calibration), the observed falling numbers (FN), latitude (lat), and longitude (lon).

collinearity of spectral data, the results of the GLM strategy were easily trapped in the region with the highest testing scores. In contrast to GLM, BLINK selects bands using single-band information and collinearity. Bands with similar information were included in the BLINK model and diversified the selection pool (Figure 7e). For example, the 1100 nm region was shared with the 2nd derivatives, and the 1408 nm region was also identified by the clustering strategy.

Overall, all the reduced models showed similar HFN prediction performance to the full model. There was no difference in the accuracies in the precision of two decimal places. The best-supervised selection strategy was BLINK ($r = 0.720$), and the worse was GLM ($r = 0.718$). On the other hand, without considering the HFN, the unsupervised strategy Low-Res performed slightly better ($r = 0.721$) than both the supervised strategies and the full model. A permutation test was con-

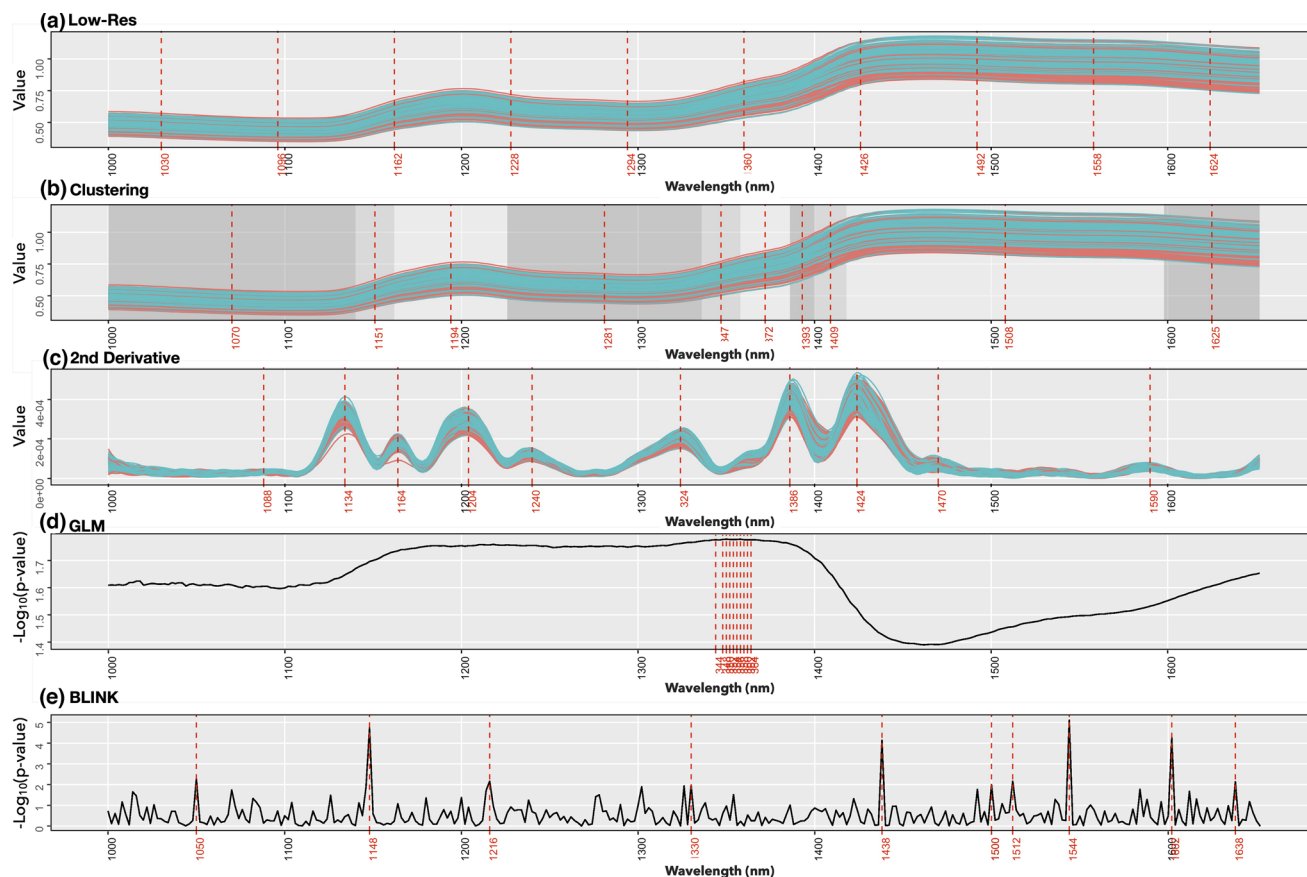


FIGURE 7 Feature selection strategies. From the spectra covering 1000–1654 nm, 10 bands were selected and marked by red vertical lines. (a) Low-Res: 10 bands were selected with an equal window. (b) Clustering: Hierarchical clustering algorithm was carried out to cluster 327 bands into 10 groups. The middle bands of each group were selected. (c) 2nd derivative: Second-order derivatives were computed. Bands with the highest absolute value of derivatives were selected. (d) Generalized linear model (GLM): Bands were tested for an association with Hagberg–Perten falling number (HFN) by a generalized linear model. (e) Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK): Bands were tested for an association with HFN by iteratively including significant bands into the testing model. Meanwhile, bands with a high correlation ($r > 0.7$) with other selected bands were excluded.

ducted where the reduced models were further examined with a random model, which was a PLSR model using 10 randomly selected bands as the predictors. Neither the reduced models nor the full model showed a significant difference in the accuracies compared to the random model (p value > 0.05). The best model was the reduced model Low-Res (p value = 0.178), and the worse was the GLM (p value = 0.749; Figure 8). Different group settings were also used to examine the reduced models. We compared the accuracies of the full model and two reduced models. The reduced models were the Low-Res and BLINK, which were the best unsupervised and supervised models in our permutation, respectively. The Low-Res model performed similarly to the full model in all group settings (i.e., low-HFN conditions, environments, and varieties). And the BLINK model showed similar performance except for the group setting of low-HFN conditions; the BLINK model showed a lower accuracy than the other two models in the PHS samples ($r = 0.12$), LMA samples ($r = 0.79$), and the sound samples ($r = 0.66$; Figure 5).

3.4 | Spatial pattern in hyperspectral images

The spatial pattern distinguished samples with different low-HFN conditions. The pixel-wise predictions of HFN were plotted in a heatmap, and the gradient from high HFN (>250 s) to low HFN (≤ 250 s) was color-coded using a yellow-to-red scale (Figure 9). We used the varieties *Jasper* and *IDO 1808* as examples to illustrate the spatial pattern of the predicted HFN. And the rest of the varieties were also displayed in the supplementary figure (Figure S4). As a result, PHS-affected kernels were mostly covered by red pixels with no obvious spatial tendency pattern. However, the distribution of red pixels had stronger tendencies in LMA and sound kernels; as HFN increases, red pixels are prone to gather on one end of the kernel. The tendency is stronger in *IDO 1808* over *Jasper* when the kernels are affected by LMA environments. In *IDO 1808*, most LMA-affected and sound kernels showed an accumulation of red pixels on only one side of the kernel. It is also noted that high predicted HFN (yellow

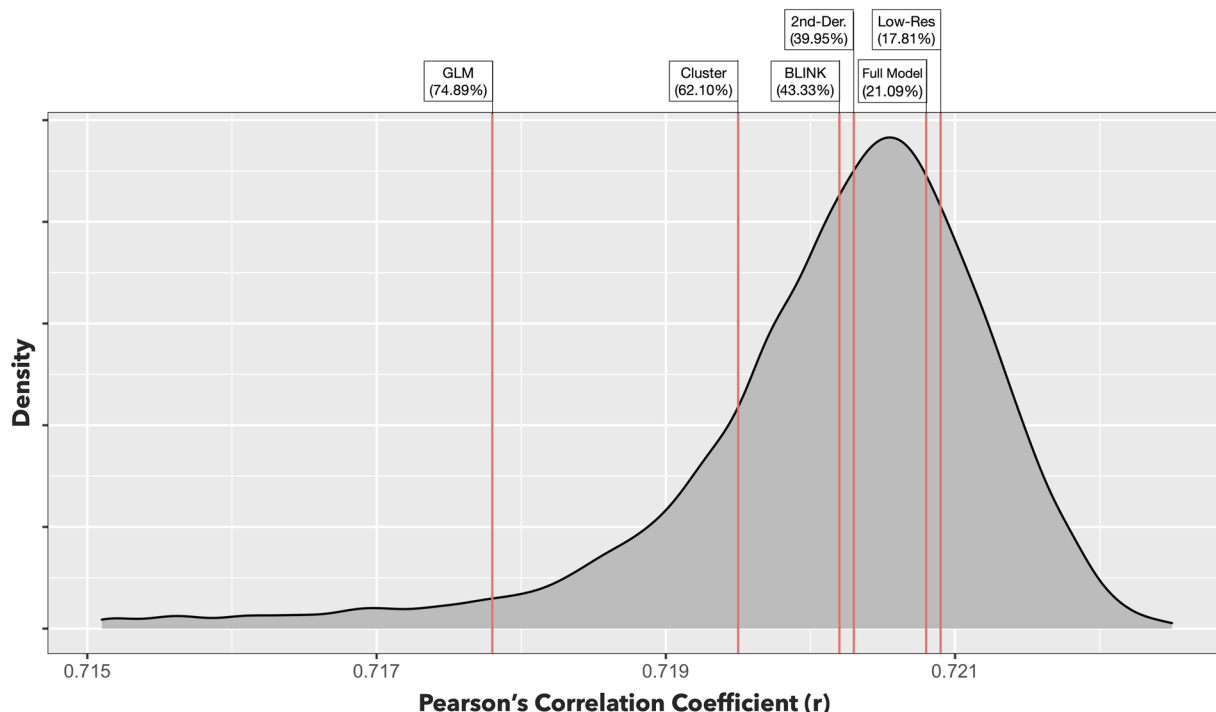


FIGURE 8 A permutation test to validate the performance of the feature selection strategies in the validation dataset. The different strategies were tested for the chance to perform worse than the models trained by 10 randomly drawn features (spectra). Red vertical lines mark the correlation accuracies (r) on x axis and had tested p value (H_0 : Presented strategies have the same prediction results made by randomly drawn bands) labeled in the parenthesis. The density function represents the null distribution of the permutation test.

pixels) was always observed on the kernel edges regardless of the sample conditions or the variety. The kernel edges reflected less light captured by the sensor, which was a consequence of the Lambertian response for the diffuse reflection (Delwiche et al., 2021).

4 | DISCUSSION

4.1 | Limitation and transferability

The presented model in this study has demonstrated a promising prediction accuracy ($r = 0.72$) of HFN on a distinct dataset, which was characterized by a different spectral instrument. However, the model had a mean absolute error of 56 s, which was higher than the measurement error (20 s) in HFN tests. This model only showed a low bias when the inspected samples were in the HFN range of around 250 s (Figure 6). Two potential sources contributed to this high-bias prediction. As the two datasets were collected by different spectrometers, the first potential source was the difference in the spectral response (i.e., absorbance) of the two datasets. For example, the peak absorbance near the region 1200 nm in the calibration set was 0.75, while the peak absorbance in the same region was 0.625 in the validation set (Figure 2a). This difference may introduce a bias when transferring the estimated

linear coefficients from one dataset to another. Although the bias can be avoided by applying a standardization method to the spectra, this approach may violate the motivation of this study, where the validation set, or any new dataset, was not available to be considered for the standardization. Hence, we should expect such error as we applied the same standardizing scaler, which was derived from the calibration set to the validation set. The second potential source of error is the difference between the HFN distributions. The median HFN in the calibration set and the validation set are 312 s and 255 s, respectively (Figure 2b). This 57-s difference cannot be explained by linear combinations alone without transforming both datasets to share the same distributions, which can be characterized by descriptive statistics such as means, standard deviation, medians, or ranges. This becomes a problem when there is a noticeable difference in the distributions between the calibration and validation samples. This problem was defined as “domain adaptation” and was well-discussed by Ben-David et al. (2010). The solution to this problem is identifying the conversion functions between datasets for both labels and features, which refer to HFN and spectra in this study, respectively. Therefore, to alleviate the concern of domain adaptation, samples from the validated dataset must be collected and labeled if the validated trial is known to have a strong deviation from the calibration system. In conclusion, given that the 20-s measurement error is a key benchmark for

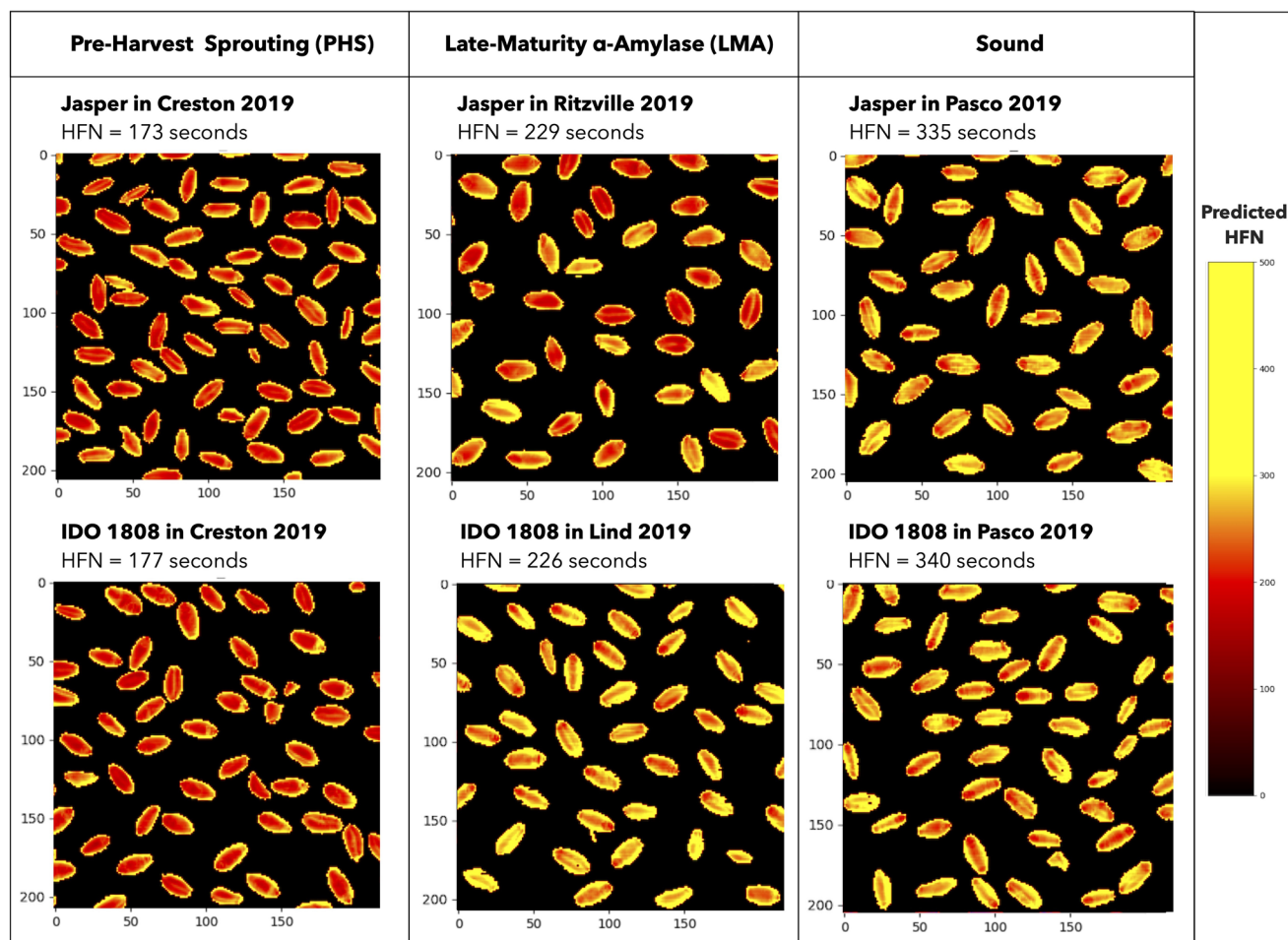


FIGURE 9 Heatmap of pixel-wise predictions of falling numbers in the validation dataset. The varieties Jasper and IDO 1808 were visualized for the spatial pattern of different conditions (Preharvest sprouting [PHS], Late maturity alpha-amylase [LMA], and sound). The color gradient, which shows the predicted Hagberg–Perten falling number (HFN), is scaled from 0 to 500 and is colored from black, red, to yellow.

the community to consider if it is worth deploying this model instead of conducting the traditional HFN tests, the presented model can only be used as a screening tool to rank the samples based on their predicted HFN values. The final HFN value still needs to be confirmed by the traditional HFN test.

4.2 | Spectra as predictors of HFN

Spectroscopy was used to predict the presence of molecular substances which emit spectral signals from the vibrations of chemical bonds. Near-infrared spectra, ranging from 800 to 2500 nm, were broadly used to identify objects such as moisture and lipid contents in coffee beans (Caporaso et al., 2018), or diagnose moldy peanuts (Jiang et al., 2016). However, compared to mid-range infrared (spectra in the range from 2500 nm to 25 μ m), the range studied in our study covering from 1000 to 1652 nm would have molecular excitation caused by overtones (Manley, 2014). For example, ignoring anharmonicity effects, signals detected around a wavelength

of 1100 nm could be overtones of its multiples, such as 2200 and 3300 nm. This fact implies that our studies in the short-wavelength range, where spectral absorption is relatively weak, may originate from wavelengths beyond our instrument's range. Including a wider and longer wavelength range will allow us to validate the selected bands by their wavelength multiples.

Our studies focused on the absolute reflectance of spectra in predicting HFN, while there is another approach, which is known as vegetation indices, to model the difference between multiple spectra in the predictions. The approaches were proposed in many different forms, such as Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974) and Difference Vegetation Index (DVI) (Tucker, 1979), and can be commonly seen in crop breeding studies (Tattaris et al., 2016). In our studies, two spectrometers show different spectral patterns across the spectrum (Figure 2a). The monitored changes over different bands cannot be consistent from one dataset to another. The hypothesis was validated by incorporating the indices into the prediction models, and there was no

improvement observed compared to the one using absolute band reflectance alone. Hence, the model's performance with indices information was not demonstrated in this study.

4.3 | PLSR in past works and hyperspectral images

As PLSR is a linear model that can effectively remove the autocorrelation among the predictors (i.e., spectra in our case), the recent studies that leveraged spectroscopy to predict HFN were mostly based on PLSR (Caporaso et al., 2017; Delwiche et al., 2018). This merit was also demonstrated in our full model using two latent variables to fit 327 spectral bands on 462 samples. With the two latent variables, the model achieved a correlation accuracy of 0.72 and an MAE of 56 s in our validation. This result was similar to Delwiche et al. (2018) (MAE = 55.4 s) and slightly better than Caporaso et al. (2017) ($r = 0.65$; MAE = 63 s). It is worth noting that the major difference between our validation and previous studies was that it was conducted on a dataset collected by a different spectrometer. This setting can reduce a potential overestimation of the model performance. We also examined the performance of the model in the calibration set. The performance only had an accuracy of $r = 0.38$, which was poorer than Caporaso et al. (2017) ($r = 0.77$) and Delwiche et al. (2018) ($r = 0.73$). Our interpretation was that we only used two latent variables that were fewer than the previous studies which both used at least seven latent variables. More variables used in a regression model can surely increase the calibration accuracy, but it may also lead to overfitting problems and poor predictability on a new dataset. In our dataset, using two variables was a good balance between the two extremes. The model was able to explain the variation of HFN in the calibration set, but it was also capable of generalizing the prediction on the validation set. To extend this idea, we also examined the performance of an ordinary least square regression, where all the 327 predictors were fitted to the calibration set. As a result, the calibration accuracy was increased to $r = 0.73$, and the validation accuracy was dramatically decreased to $r = -0.06$ as expected (Figure S5).

It is interesting to see a spatial pattern in hyperspectral images using the calibrated PLSR model. In Figure 9, the red pixels represented a lower predicted HFN. As α -amylase was known to be the major cause of low FN, it was hypothesized that the red pixels might represent the indicators of α -amylase distribution on the kernel surface. The synthesis of α -amylase is known to be triggered by gibberellin accumulated in the embryo during germination; when α -amylase is present throughout the kernel, the samples will likely have sprout damage and lower FN (Mrva & Mares, 1996; Mrva et al., 2006). The synthesis process aligned with the observation in our hyperspectral images: The HFN of a kernel started

to decrease when red pixels were observed accumulating near the embryo end of the kernel.

5 | CONCLUSION

This is the first study that validated the potential of transferring a calibrated spectroscopy model of HFN to an independent dataset measured by a different spectral instrument. The independent validation shows a promising correlation accuracy r of 0.72, which was close to the previous studies that used the same instrument in the validation. This study also showed a cost-effective alternative that used only 10 bands to predict HFN; the alternative achieved similar or better performance than using the full spectrum of the current HSI system. However, the presented model has a high bias of 56 s, which is still not acceptable for the industry to replace the conventional HFN test where the measurement error is 20 s. Considering the validation was conducted on a small sample batch, the bias is anticipated to be reduced by using a larger and more balanced dataset that covers a wider range of environments and varieties. In conclusion, this study suggested that using spectrometers has the potential to serve as a faster alternative for plant breeders to develop varieties resistant to PHS and LMA, and for growers to screen damaged grains in harvesting and transportation processes.

AUTHOR CONTRIBUTIONS

Chun-Peng James Chen: Data curation; formal analysis; investigation; methodology; project administration; validation; visualization; writing—original draft. **Yang Hu:** Data curation; writing—review & editing. **Xianran Li:** Validation; writing—review & editing. **Craig F. Morris:** Validation; writing—review & editing. **Stephen Delwiche:** Data curation; methodology; validation; writing—review & editing. **Arron H. Carter:** Conceptualization; resources. **Camille Steber:** Conceptualization; data curation; investigation; methodology; project administration; resources; supervision; validation; writing—review & editing. **Zhiwu Zhang:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; visualization; writing—original draft; writing—review & editing.

ACKNOWLEDGMENTS

This project was partially supported by the USDA National Institute of Food and Agriculture (Hatch project 1014919, Award #s 2019-67013-29171, and 2020-67021-32460) and the Washington Grain Commission (Endowment and Award #s 126593 and 134574). Dr. Michael O. Pumphrey was acknowledged for his help in the field management of the study. The authors dedicate this article to the memory of our co-author Dr. Craig Morris.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Chun-Peng James Chen  <https://orcid.org/0000-0002-2018-0702>

Yang Hu  <https://orcid.org/0000-0002-7350-6147>

Xianran Li  <https://orcid.org/0000-0002-4252-6911>

Arron H. Carter  <https://orcid.org/0000-0002-8019-6554>

Camille Steber  <https://orcid.org/0000-0001-6255-7670>

Zhiwu Zhang  <https://orcid.org/0000-0002-5784-9684>

REFERENCES

- Abdi, H. (2003). Partial least square regression PLS-regression. In *Encyclopedia of social sciences research methods* (pp. 792–795). Sage.
- Armstrong, P., Maghirang, E., Yaptenco, K., & Pearson, T. C. (2016). Visible and near-infrared instruments for detection and quantification of individual sprouted wheat kernels. *Transactions of the ASABE (American Society of Agricultural and Biological Engineers)*, *59*, 1517–1527. <https://doi.org/10.13031/trans.59.11566>
- Barbedo, J. G. A., Guarienti, E. M., & Tibola, C. S. (2018). Detection of sprout damage in wheat kernels using NIR hyperspectral imaging. *Biosystems Engineering*, *175*, 124–132. <https://doi.org/10.1016/j.biosystemseng.2018.09.012>
- Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Vaughan, J. W. (2010). A theory of learning from different domains. *Machine Learning*, *79*, 151–175. <https://doi.org/10.1007/s10994-009-5152-4>
- Bettge, A. D. (2018). Low falling numbers in the pacific northwest wheat growing region: Preharvest sprouting, late maturity amylase, falling number instrument, or low protein? *Cereal Foods World*, *63*, 12–16. <https://doi.org/10.1094/CFW-63-1-0012>
- Borchers, H. W. (2021). *pracma: Practical numerical math functions*. <https://cran.r-project.org/web/packages/pracma/index.html>
- Bradski, G. (2000). The openCV library. *Dr. Dobb's Journal of Software Tools*, *120*, 122–125.
- Campbell, K. (2017). *White paper: Strategies to reduce economic losses due to low falling number in wheat*. Washington State University Extension.
- Caporaso, N., Whitworth, M., & Fisk, I. (2017). Application of calibrations to hyperspectral images of food grains: Example for wheat falling number. *Journal of Spectral Imaging*, *6*, Article a4. <https://doi.org/10.1255/jsi.2017.a4>
- Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2018). Near-infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Applied Spectroscopy Reviews*, *53*, 667–687. <https://doi.org/10.1080/05704928.2018.1425214>
- Chen, J., Chen, H., Wang, X., Yu, C., Wang, C., & Zhu, D. (2014). The characteristic of hyperspectral image of wheat seeds during sprouting. In D. Li & Y. Chen (Eds.), *IFIP advances in information and communication technology*. Springer.
- Delwiche, S. R., Baek, I., & Kim, M. S. (2021). Effect of curvature on hyperspectral reflectance images of cereal seed-sized objects. *Biosystems Engineering*, *202*, 55–65. <https://doi.org/10.1016/j.biosystemseng.2020.11.004>
- Delwiche, S. R., Higginbotham, R. W., & Steber, C. M. (2018). Falling number of soft white wheat by near-infrared spectroscopy: A challenge revisited. *Cereal Chemistry*, *95*, 469–477. <https://doi.org/10.1002/cche.10049>
- Delwiche, S. R., Rodriguez, I. T., Rausch, S. R., & Graybosch, R. A. (2019). Estimating percentages of fusarium-damaged kernels in hard wheat by near-infrared hyperspectral imaging. *Journal of Cereal Science*, *87*, 18–24. <https://doi.org/10.1016/j.jcs.2019.02.008>
- Farrell, A., & Kettlewell, P. (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>
- Gale, M. D., Salter, A. M., & Lenton, J. R. (1987). *The induction of germination alpha-amylase during wheat grain development in unfavourable weather conditions*. CRC Press.
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, *58*, 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)
- Hagberg, S. (1960). Note on a simplified rapid method for determining alpha-amylase activity. *Cereal Chemistry*, *38*, 202–203.
- Hagberg, S. (1961). A rapid method for determining alpha-amylase activity. *Cereal Chemistry*, *37*, 218–222.
- Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, *8*, giy154. <https://doi.org/10.1093/gigascience/giy154>
- Jiang, J., Qiao, X., & He, R. (2016). Use of near-infrared hyperspectral images to identify moldy peanuts. *Journal of Food Engineering*, *169*, 284–290. <https://doi.org/10.1016/j.jfoodeng.2015.09.013>
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews*, *43*, 8200–8214. <https://doi.org/10.1039/C4CS00062E>
- Major, B. J. (1999). Environmental factors affecting pre-maturity alpha-amylase activity in winter wheat (*Triticum aestivum*) [Ph. D. Thesis, The Open University].
- 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). Wheat grain preharvest sprouting and late maturity alpha-amylase. *Planta*, *240*, 1167–1178. <https://doi.org/10.1007/s00425-014-2172-5>
- Mares, D. J., Mrva, K., & Panozzo, J. F. (1994). Characterization of the high α -amylase levels in grain of the wheat cultivar BD 159. *Australian Journal of Agricultural Research*, *45*, 1003–1011. <https://doi.org/10.1071/ar9941002>
- Marimont, R. B., & Shapiro, M. B. (1979). Nearest neighbour searches and the curse of dimensionality. *IMA Journal of Applied Mathematics*, *24*, 59–70. <https://doi.org/10.1093/imamat/24.1.59>
- Mrva, K., & Mares, D. (1996). Expression of late maturity alpha-amylase in wheat containing gibberellic acid insensitivity genes. *Euphytica*, *88*, 69–76. <https://doi.org/10.1007/BF00029267>
- Mrva, K., Wallwork, M., & Mares, D. (2006). Amylase and programmed cell death in aleurone of ripening wheat grains. *Journal of Experimental Botany*, *57*, 877–885. <https://doi.org/10.1093/jxb/erj072>
- Neethirajan, S., Tuteja, S. K., Huang, S.-T., & Kelton, D. (2017). Recent advancement in biosensors technology for animal and livestock health management. *Biosensors and Bioelectronics*, *98*, 398–407. <https://doi.org/10.1016/j.bios.2017.07.015>
- Neoh, G. K. S., Tao, K., Dieters, M. J., Fox, G. P., & Gilbert, R. G. (2021). Late-maturity α -amylase (LMA) testing and its

- methodological challenges. *LWT*, *151*, 112232. <https://doi.org/10.1016/j.lwt.2021.112232>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.
- Randall, P. J., & Moss, H. J. (1990). Some effects of temperature regime during grain filling on wheat quality. *Australian Journal of Agricultural Research*, *41*, 603–617. <https://doi.org/10.1071/ar9900603>
- Risius, H., Hahn, J., Markus, H., Rainer, T., & Hubert, K. (2014). 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>
- Rouse, W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS-1 Symposium NASA, NASA SP-351*, Washington, DC.
- Ross, A. S., & Bettge, A. D. (2009). *Passing the test on wheat end-use quality*. John Wiley & Sons, Ltd.
- Skerritt, J. H., & Heywood, R. H. (2000). A Five-minute field test for on-farm detection of pre-harvest sprouting in wheat. *Crop Science*, *40*, 742–756. <https://doi.org/10.2135/cropsci2000.403742x>
- Tattaris, M., Reynolds, M. P., & Chapman, S. C. (2016). A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Frontiers in Plant Science*, *7*, 1–9. <https://doi.org/10.3389/fpls.2016.01131>
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Vadivambal, R., Chelladurai, V., Jayas, D. S., & White, N. D. G. (2011). Determination of sprout-damaged barley using thermal imaging. *Agricultural Engineering International: CIGR Journal*, *13*, 9.
- Verity, J. C. K., Hac, L., & Skerritt, J. (1999). Development of a field enzyme-linked immunosorbent assay (ELISA) for detection of α -amylase in preharvest-sprouted wheat. *Cereal Chemistry*, *76*(5), 673–681. <https://doi.org/10.1094/CCHEM.1999.76.5.673>
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Wold, H. (1980). *Model construction and evaluation when theoretical knowledge is scarce*. Elsevier.
- 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>
- Zhang, L., Sun, H., Rao, Z., & Ji, H. (2020). Non-destructive identification of slightly sprouted wheat kernels using hyperspectral data on both sides of wheat kernels. *Biosystems Engineering*, *200*, 188–199. <https://doi.org/10.1016/j.biosystemseng.2020.10.004>

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: Chen, C.-P. J., Hu, Y., Li, X., Morris, C. F., Delwiche, S., Carter, A. H., Steber, C., & Zhang, Z. (2023). An independent validation reveals the potential to predict Hagberg–Perten falling number using spectrometers. *The Plant Phenome Journal*, *6*, 1–15. <https://doi.org/10.1002/ppj2.20070>