



WE'RE WITH YOU **EVERY STEP** OF THE WAY

Stop pathogens in their tracks! Sanitizing footwear helps reduce cross-contamination in production and other critical control points. Combine the HACCP SmartStep2™

Dual Footwear Sanitizing Unit with Alpet® D2 or Alpet® Quat-Free Surface Sanitizer to create a powerful tool to reduce cross-contamination from footwear.

CONTACT US AT 888-225-3267 OR VISIT WWW.BESTSANITIZERS.COM
FOR A COMPLIMENTARY REVIEW OF YOUR FOOTWEAR SANITATION PROGRAM

BEST SANITIZERS INC.
ALPET
Surface Sanitizer **D2**

BEST SANITIZERS INC.
ALPET D2
Quat-Free
Surface Sanitizer

HACCP
**Smart
Step2**
Walk-Through Dual Footwear
Sanitizing Unit

Food Safety is Our Priority

Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review

Zhenjiao Du¹ | Wenfei Tian² | Michael Tilley³ | Donghai Wang⁴ |
Guorong Zhang⁵ | Yonghui Li¹ 

¹Department of Grain Science and Industry, Kansas State University, Manhattan, Kansas, USA

²National Wheat Improvement Centre, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

³USDA, Agricultural Research Service, Center for Grain and Animal Health Research, Manhattan, Kansas, USA

⁴Department of Biological and Agricultural Engineering, Kansas State University, Manhattan, Kansas, USA

⁵Agricultural Research Center–Hays, Kansas State University, Hays, Kansas, USA

Correspondence

Yonghui Li, Department of Grain Science and Industry, Kansas State University, 201 Shellenberger Hall, Manhattan, KS 66506, USA.

Email: yonghui@ksu.edu

Abstract

Wheat is one of the most widely cultivated crops throughout the world. A great need exists for wheat quality assessment for breeding, processing, and products production purposes. Near-infrared spectroscopy (NIRS) is a rapid, low-cost, simple, and nondestructive assessment method. Many advanced studies associated with NIRS for wheat quality assessment have been published recently, either introducing new chemometrics or attempting new assessment parameters to improve model robustness and accuracy. This review provides a comprehensive overview of NIRS methodology including its principle, spectra pretreatments, spectral wavelength selection, outlier disposal, dataset division, regression methods, and model evaluation. More importantly, the applications of NIRS in the determination of analytical parameters, rheological parameters, and end product quality of wheat are summarized. Although NIRS showed great potential in the quantitative determination of analytical parameters, there are still challenges in model robustness and accuracy in determining rheological parameters and end product quality for wheat products. Future model development needs to incorporate larger databases, integrate different spectroscopic techniques, and introduce cutting-edge chemometrics methods. In addition, calibration based on external factors should be considered to improve the predicted results of the model. The NIRS application in micronutrients needs to be extended. Last, the idea of

Nomenclature: ACO, ant colony optimization; Bi-PLS, backward interval partial least square; BPNN, back propagation neural network; CARS, competitive adaptive reweighted sampling; CCA, canonical correlation analysis; CNN, convolutional neural network; CSMWPLSs, changeable size moving window partial least squares; CT, X-ray computed tomography; DBN, Deep belief network; Der, derivation; DOSC, direct orthogonal signal correction; DT, de-trending correction; ELM, extreme learning machine; EMSC, extended MSC; FD, first derivative; FS, fluorescence spectroscopy; HSI, hyperspectral imaging; i-PLS, interval partial least square; iRF, interval random frog; MC, mean centering; MIRS, mid-infrared spectroscopy; MLR, multiple linear regression; MSC, multiplicative scatter correction; MWPLS, moving window partial least square; NIRS, near-infrared spectroscopy; NL, normalization; NMR, nuclear magnetic resonance; NWD, Norris–Williams derivatives; OC, offset correction; OR, outliers removal; OSC, orthogonal signal correction; PCA, principal component analysis; PCC, Pearson product-moment correlation coefficients; PCR, principal component regression; PLSR, partial least squares regression; PMSC, piece-wise multiplicative scatter correction; PSO, particle swarm optimization; RBF, radial basic function; RBFNN, radial basic function neural network; RCA, regression coefficient analysis; RFR, random forest regression; ROI, region of interest; SA, simulated annealing; SCMWPLSs, searching combination moving window partial least square; SD, second derivative; SGA, Savitzky–Golay algorithm; SGD, Savitzky–Golay derivatives; Si-GA, synergy interval genetic algorithm; Si-PLS, synergy interval partial least square; SLS, straight line subtraction; SNV, standard normal variate; SPA, successive projections algorithm; SVMR, support vector machine regression; VCPA, variable combination population analysis; WT, wavelet transform.

combining standard product sensory attributes and spectra for model development deserves further study.

KEYWORDS

artificial intelligence, chemometrics, machine learning, near-infrared spectroscopy, wheat quality assessment

1 | INTRODUCTION

Wheat is one of the most widely cultivated crops worldwide, and its cultivated area and production reached 222.63 million hectares and 772.38 million tons in 2020, respectively (USDA, 2020). Approximately 65% of wheat is used for human consumption, though it is also widely utilized as feed, energy production, and nonfood industries (Gabriel et al., 2017). Depending on different end uses, quality classification based on wheat compositions or functionalities has been developed in many countries (Gabriel et al., 2017; Salimi Khorshidi et al., 2018). However, these classification methods are general and do not allow precise quantitative quality assessment. Breeders, millers, bakers, and other end users demand precise quantitative assessments of wheat, because quality has a great impact on breeding, processing, and marketing, and precise quantitative assessment allows maximum utilization of kernel, flour, and dough (Cevoli et al., 2015; Dowell et al., 2006; Pojić & Mastilović, 2013). Quality indicators of wheat include analytical parameters (e.g., protein content, falling number), rheological parameters (e.g., mixing resistance, viscoelasticity), and end product quality parameters (e.g., bread loaf volume and texture) (Dowell et al., 2006; Mutlu et al., 2011; Pojić & Mastilović, 2013). Conventional assessment methods that use wet chemistry are laborious, time-consuming, expensive, not environmentally friendly, and require experienced technicians (Manley, 2014; Shi et al., 2019). Given the disadvantages of traditional assessment methods and the heavy use of wheat as well as wheat products, there is a great demand for rapid, low-cost, nondestructive, simple, and environmentally friendly assessment methods to evaluate the quality of wheat kernels, ground flours, dough, and end products (Dowell et al., 2006; Salimi Khorshidi et al., 2018).

To meet this demand, researchers have studied and applied spectroscopic approaches such as nuclear magnetic resonance (NMR), fluorescence spectroscopy (FS), X-ray computed tomography (CT), hyperspectral imaging (HSI), ultraviolet spectroscopy, visible spectroscopy, and mid-infrared spectroscopy (MIRS) (Ahmad et al., 2016; Besançon et al., 2020; Botosoa et al., 2013; Caporaso et al., 2018a; Ezeanaka et al., 2019; Fox, & Manley, 2014; Sadat

et al., 2019; Salimi Khorshidi et al., 2018; Sendin et al., 2018). Compared to these methods, near-infrared spectroscopy (NIRS) is the most popular one given its advantages in throughput, portability, versatility, simplicity, and cost (Cen & He, 2007; Manley, 2014; Sadat et al., 2019).

NIRS was first studied in the early 1970s by Karl Norris and Phil Williams, who combined NIRS with chemometrics methods to measure protein and moisture contents (Manley, 2014; Norris & Williams, 1984; Williams et al., 2019). Subsequently, studies and reviews of NIRS applications significantly increased (Caporaso et al., 2018a; Lasztity & Abonyi, 2009; Liang et al., 2022; Mutlu et al., 2011; Pojić & Mastilović, 2013). While there are many studies combining NIRS with advanced chemometrics for wheat quality assessment, to our knowledge, there is no comprehensive and in-depth review for readers who have expertise in physicochemical experiments but lack knowledge in spectral data analysis, or for companies to improve the applications of NIRS in wheat quality assessment. To close this gap, this review presents the principles, spectra acquisition, pretreatments, wavelength selection, outlier treatments, dataset division, and model development in NIRS methods. Especially, the application of NIRS in quantitative assessments of compositional parameters, physical parameters, rheological parameters, and end product quality was also reviewed. Furthermore, it discusses future development trends of the NIRS technology in the wheat quality assessment (Figure 1).

2 | NIRS METHODOLOGY

2.1 | Principle and spectra acquisition

NIRS works by using the near-infrared (NIR) radiation absorption in the 750–2500 nm range (Manley, 2014). Absorptions in the NIR region are generated from fundamental vibrations by overtones and combinations. The NIR spectrum contains information about the X–H chemical bonds (e.g., C–H, N–H, and O–H). The absorption of these bonds indicates unique compositions of the analytes, and the widely abundant hydrogen in organic materials enables NIRS to quantitatively determine various

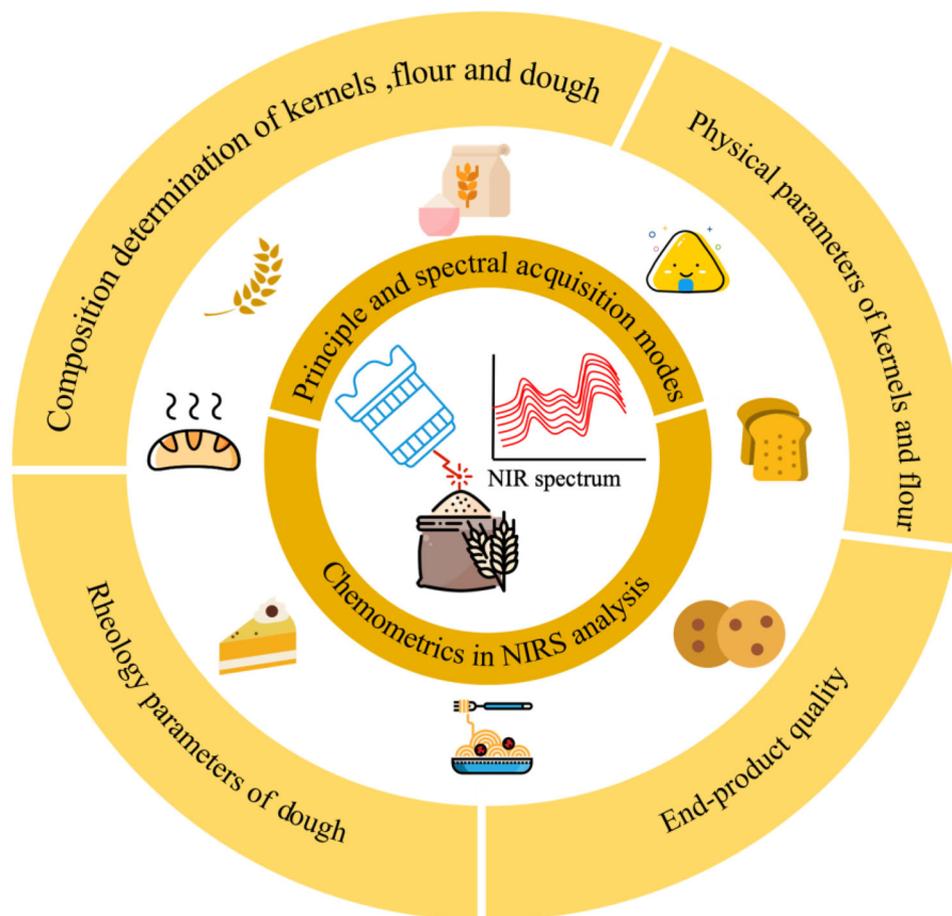


FIGURE 1 Application of near-infrared spectroscopy for wheat quality assessment

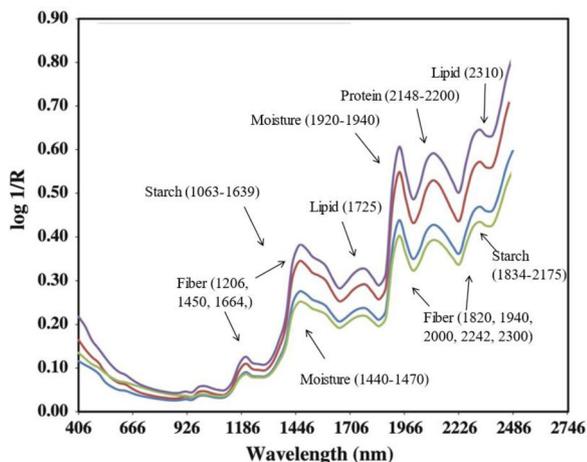


FIGURE 2 Spectra of several wheat flour samples and the approximate absorption wavelength ranges for different compositions (Badaró et al., 2019; Manley, 2014; Mutlu et al., 2011)

organic materials (Agelet & Hurburgh, 2010; Manley, 2014). Figure 2 shows the typical spectral profiles of dif-

ferent wheat flour samples and the approximate absorption wavelength ranges for different compounds (Badaró et al., 2019; Manley, 2014; Mutlu et al., 2011). Due to the overtone and combination bands, it is difficult to directly use the spectrum for quantitative assessments of wheat and its end products (Manley, 2014). Modern chemometrics methods had been developed to extract information from the spectral data and establish calibration models that allow the practical application of NIRS (Agelet & Hurburgh, 2010). The main procedures for establishing calibration models are shown in Figure 3. The spectral datum is a mean spectrum of a sample's measured positions (average measurement), and there are several spectral data acquisition modes (transmittance, reflectance, transfectance, and interactance) (Manley, 2014; Pojić & Mastilović, 2013). While the reflectance mode is the most popular one among wheat quality assessments, the suitable data acquisition mode is highly dependent on sample types and practical needs (Albanell et al., 2021; Armstrong, 2014; Owens et al., 2009; Sato et al., 2001).

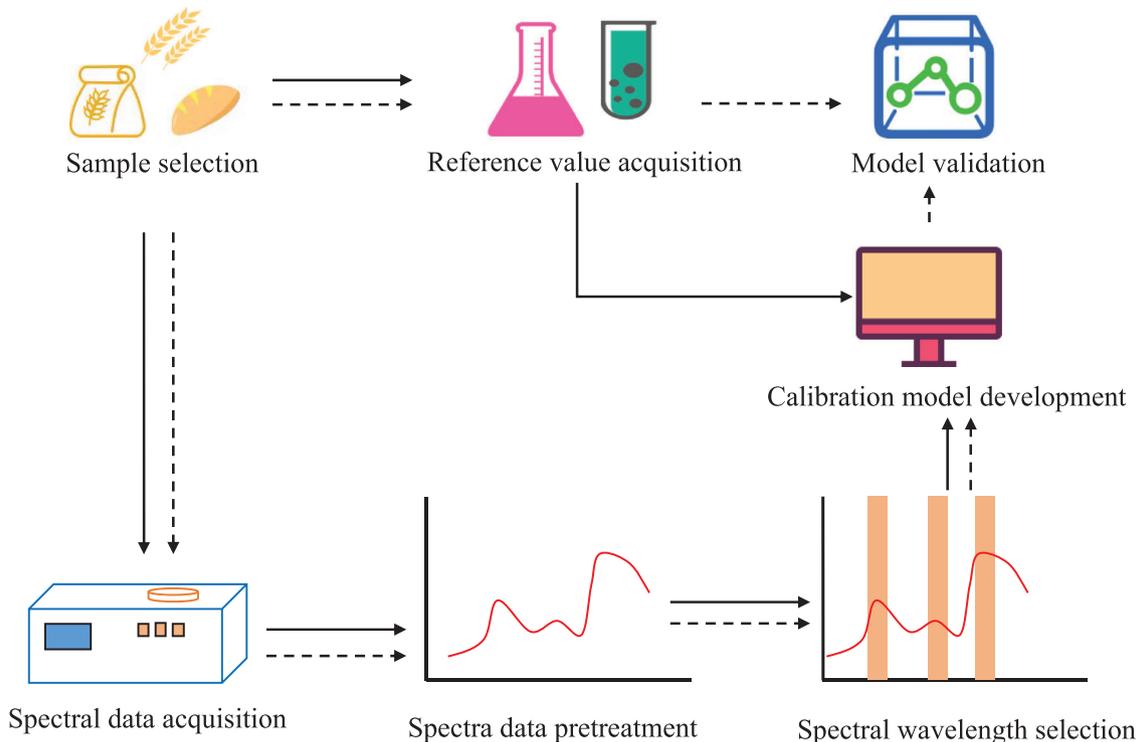


FIGURE 3 The main procedures for establishing calibration models for wheat quality assessment using near-infrared spectroscopy

2.2 | Chemometrics in NIRS analysis

2.2.1 | Spectral pretreatment

The objective of pretreatment is to reduce systematic noise and increase or enhance the signals from the chemical information associated with the samples to improve calibration model building (Agelet & Hurburgh, 2010; Manley, 2014; Pojić & Mastilović, 2013; Rinnan et al., 2009). An illustration of frequently used spectra pretreatment methods is shown in Figure 4. Raw spectral data changed when different methods are employed, and each pretreatment method has its own advantage, so there is no general rule for pretreatment selection. In practice, optimal pretreatment selections require trial and error guided by the cross-validation/test set results and experience (Agelet & Hurburgh, 2010; Cui & Fearn, 2018). The commonly used pretreatment methods for biological materials and their principles, functions, and characteristics are shown in Table 1. It is worth mentioning that all of the pretreatment methods may lead to information loss from the raw spectrum, although they can improve the signal to noise ratio (Bian et al., 2018; Dong & Sun, 2013; Shi & Yu, 2017; Williams, 2020). In addition, multiple studies have shown that with the adoption of optimal datasets and the application of advanced algorithms, the effect of pretreatment on the performance of the final model was no longer significant (Shi & Yu, 2017; Dong & Sun, 2013; Bian et al., 2018;

Williams, 2020; Zhang et al., 2019; Harrington, 2018; Mao et al., 2014; Mutlu et al., 2011; Chen et al., 2017; Peiris et al., 2017; Gabriel et al., 2017).

2.2.2 | Spectral wavelength selection

The objective of spectral wavelength selection, also known as data dimension reduction, is to reduce redundant information in the raw spectral data, since different chemical compositions usually have specific absorption bands. The commonly used spectral wavelength selection methods for biological materials and their principles, advantages, and disadvantages are shown in Table 2. Although some researchers believe that it is better to choose the full spectrum for the calibration model, theoretical and experimental evidence showed that spectral wavelength selection could also significantly improve the performance of models with a lower computation consumption (Chen et al., 2008; Seema et al., 2020; Tian et al., 2021; Williams, 2020). There are two technical routes to spectral wavelength selection. The first is to select a specific range of wavelengths to represent full wavelength range in the raw spectral data. With this route, researchers had gradually increased their reliance on data analysis for spectral range selection and decreased their reliance on experience (Chen, Guo, et al., 2012; Yang et al., 2017). This could be seen in the change from manual spectral range

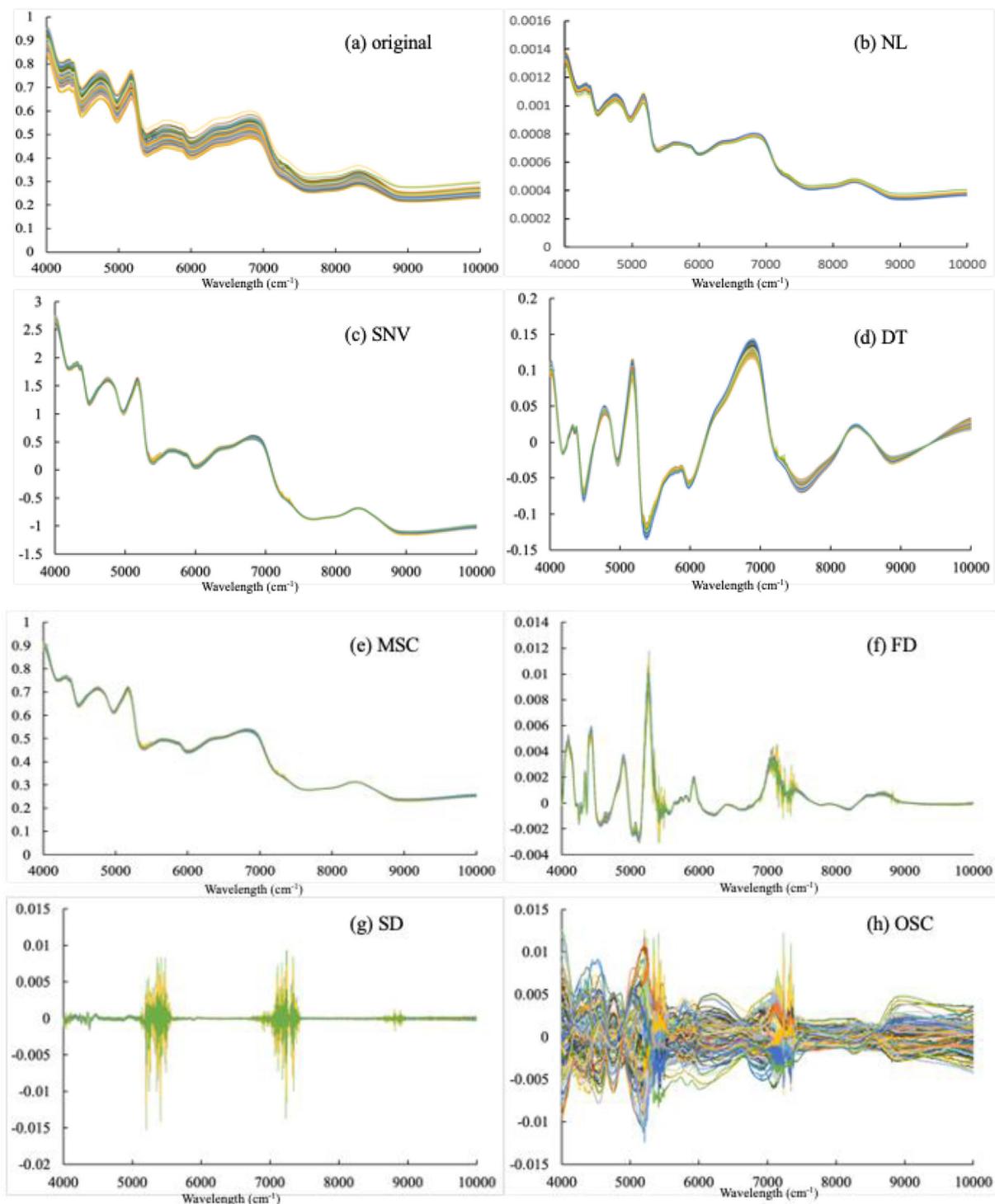


FIGURE 4 Frequently used pretreatment methods: (a) original absorption spectra, (b) normalization (NL), (c) standard normal variate (SNV), (d) detrending correction (DT), (e) multiplicative scatter correction (MSC), (f) first derivative (FD), (g) second derivative (SD), and (h) orthogonal signal correction (OSC). The original spectra were collected using 107 hard red winter wheat samples at the authors' laboratory

selection to automatic optimal spectral range combination search (e.g., interval partial least square [i-PLS]) (Chen et al., 2017; Dong & Sun, 2013). The other route is to select specific wavelengths instead of spectral ranges to represent raw data (e.g., variable combination population anal-

ysis [VCPA]). This second route could reduce more redundant information than the first route, but there were only a few applications of this method in wheat NIR spectral wavelength selection (Luo et al., 2015; Yang et al., 2017).

TABLE 1 An overview of the principles, functions, and characteristics of different pretreatment methods

| Pretreatment methods | Principles | Functions | Characteristics | References |
|----------------------|---|--|---|---|
| MSC | Average the full wavelength range and let each individual spectrum be regressed by partial least squares to the total average | Spectral distortion correction | Removal of undesirable and multiplicative scatter effects | Dhanoa et al., 1994 |
| PMSC | Full spectra are divided into nonoverlapping subranges of equal width and adopt MSC to each subrange | Correct for nonlinear additive scatter and multiplicative effects | Suitable for nonlinear models based on local linear approaches; easily remove significant differences between spectra | Isaksson & Kowalski, 1993 |
| SNV | The mean of each spectrum is subtracted from the full wavelength range, and centered values are divided by the standard deviation of each spectrum | Spectral distortion correction | More sensitive to noise than MSC; no need for the reference spectrum compared to MSC | Dhanoa et al., 1994 |
| NL | The whole spectrum is divided by different vector norms (city blocks, Taxicab norms, and Euclidian norms) | Correct for spectral distortions due to the interference of scatter and particle size | More sensitive to noise than MSC; no need for the reference spectrum compared to MSC | Rinnan et al., 2009 |
| Der | Carry out derivatives to a spectrum | Correct the effect of overlapping peaks and remove the spectral base line offset and baseline slope | Enhancement of noise | Manley, 2014 |
| NWD | Choose k point window for smoothing and then calculate the derivative between the two smoothed values with a given gap size | Avoid noise inflation in finite differences | Difficult data interpretability; reduction in signal-to-noise ratio | Rinnan et al., 2009 |
| SGA | A polynomial function of a selected degree is fit to a window of spectral points by least squares | Smooth spectra and decrease detrimental effects on the signal-to-noise ratio | Hard to find a proper window size | Rinnan et al., 2009 |
| SGD | Smoothing method is the same as that of SGA. Calculate derivation of the fitted polynomial function as the derivative of the spectrum. | Correct the effect of overlapping peaks and remove the spectral base line offset and baseline slope | Hard to choose the proper polynomial order and window size | Agelet & Hurburgh, 2010; Rinnan et al., 2009 |
| MC | Each spectrum is subtracted by the average spectrum from the entire dataset | Reduce complexity of calibration models | Commonly used for PCA-based calibration methods | Agelet & Hurburgh, 2010; Manley, 2014 |
| OC | Correction is performed by subtracting the mean of a few arbitrary variables from each spectrum | Baseline correction | Correct for parallel baseline shifts | Luypaert et al., 2004 |
| DT | The baseline is fitted by a second-degree polynomial and subsequently subtracted from the spectra | Baseline correction | Remove offset and curvilinearity | Luypaert et al., 2004 |
| OSC | Calculate and remove the variables in the spectrum that are orthogonal to the reference values based on a constrained bilinear model and the inverse PLSR model | Decrease the number of variations in the spectrum, especially redundant variation, and reduce model complexity | Hard to find and remove the optimal number of orthogonal variables from the spectrum | Westerhuis et al., 2001 |
| DOSC | Calculate and remove the variables in the spectrum that are orthogonal to the reference values by using only least squares steps | Remove information not related to the chemical content of the sample | Suitable to handle spectra obtained from different external conditions; tend to overfit | Teye et al., 2014; Luypaert et al., 2007; Westerhuis et al., 2001 |

TABLE 2 An overview of the principles, advantages, and disadvantages of frequently used spectral wavelength selection methods

| Spectral wavelength selection methods | Principles | Advantages | Disadvantages | References |
|--|---|--|--|--|
| Full range | Directly use the full spectrum | Contain all information in the spectrum | Too much redundant information | Luo et al., 2015 |
| Manual selection | Selected manually based on the targeted composition absorption band | Greatly reduce the variables in the spectrum | Some unknown bands beneficial to calibration are eliminated | Chen, Guo, et al., 2012 |
| PCC | Calculate the Pearson product-moment correlation coefficient between the absorption and the reference values in each spectral wavelength | Remove some noisy regions | Cannot evaluate the collinear interaction between spectral variables | Dong & Sun, 2013 |
| RCA | Calculate the regression coefficient between the absorption and the reference values in each spectral wavelength | Select important wavelengths for calibration | Cannot evaluate the collinear interaction between spectral variables | Shi et al., 2019 |
| SPA | Calculate the optimal subset of the spectral matrix by a sequence of projection operation | Select subsets of variables with minimal redundancy | Hard to choose the initial spectral range | Ye et al., 2018 |
| SA | Introduce a random factor and natural mechanism in the physical annealing process to discriminate the new variables | Globally optimize the spectra | Cannot be repeated | Ye et al., 2018 |
| PSO | Initialize a group of variables to represent potential optimal wavelengths and calculate the fitness values to compare with the Pbest value and Gbest value for the optimal wavelengths | High converging speed and wide applicability | Converge slowly and diverge easily in its later evolutionary stage | Mao et al., 2014 |
| VCPA | Use the binary matrix sampling method to resample the sample space, and adopt the exponential decay function to eliminate variables that have nothing to do with sample characteristics | Fully consider the possible interactions between randomly combined wavelengths | Have some randomness since sampling is based on the BMS method | Jiang et al., 2020 |
| i-PLS | Divide the spectrum into a certain number of equidistant subintervals and optimize using PLS to evaluate the best intervals with the optimal width and position | Reduce redundant information | Need to conduct a large number of interval shifts and width change; only one interval chosen | Chen et al., 2008 |
| SI-PLS | Divide the full spectrum into several subintervals and use PLS model to fit the combination of two, three, or four intervals and find the globally optimal combination of several intervals | All possible interval combinations can be searched; use optimized combinations for the calibration model | Only test a series of adjacent but nonoverlapping intervals | Chen, Ding, et al., 2012; Yang et al., 2017 |

(Continues)

TABLE 2 (Continued)

| Spectral wavelength selection methods | Principles | Advantages | Disadvantages | References |
|---------------------------------------|---|--|---|-------------------------------------|
| Bi-PLS | Divide the spectrum into a given number of intervals and calculate the PLS model with respect to RMSECV value, eliminating the interval with lowest RMSECV at every turn till the last interval | Spectral intervals with high degrees of interference will be removed; fully automated | Only test a series of adjacent but nonoverlapping intervals | Leardl & Norgaard, 2004 |
| Si-GA | The spectral intervals are selected by Si-PLS and further optimized by genetic algorithm | Suitable for selecting the best variables from less than 200 variables in the extracted spectral intervals | Performance varies greatly depending on the original variables and the size of test dataset; easy to fall into local optimal problems | Yang et al., 2017 |
| MWPLS | Build a series in a window that moves through the full wavelength range and then choose the informative intervals with low model complexity and low value of the sum of residuals | Select all possible informative intervals but not the optimized ones | Fixed window size and no combination of the selected intervals | Yang et al., 2017 |
| CSMWPLSs | Compared with MWPLS, the difference is that the algorithm adopts a changeable size moving windows for the informative spectra selection | Collect all possible subwindows for every window size with useful information | Time-consuming; locally optimized algorithm | Du et al., 2004 |
| SCMWPLSs | Change the window size and move the window over a selected informative region with each window size | Look for an optimized combination of informative regions by performing CSMWPLS | Time-consuming; locally optimized algorithm | Du et al., 2004 |
| iRF | The spectrum is first divided into subintervals by moving the window of a fixed width over the full wavelength range. These optimal intervals are chosen by random frog coupled with PLS | Consider all possible continuous spectral intervals | Interval selection cannot be reproduced | Li et al., 2012 |
| ACO | The algorithm employs the concept of cooperative pheromone accumulation and optimizes spectral variables using a predefined number of variables; use a Monte Carlo approach to discard irrelevant variables | Greatly reduce the number of variables; suitable for global search | Some relevant variables may be omitted by the ACO algorithm | Zhang et al., 2017 |
| CARS | Conduct variable runs iteratively to obtain optimal subset variables with the lowest root mean square of cross-validation (RMSECV) | Able to locate an optimal combination of some key wavelengths | Only consider the value of regression coefficient for each variable | Luo et al., 2015; Yang et al., 2017 |

2.2.3 | Outliers and dataset division

Outliers and dataset division need to be taken into consideration before model development. Outliers exist both in reference values and spectral data. To remove outliers in reference values, test methods like the Dixon's test, Tukey's rule, or Grubbs' studentized mean deviation can be useful tools (Dixon, 1950; Grubbs, 1950). When it comes to spectral data, there are two different approaches toward outliers. One is that outliers in spectra should be detected and removed since they may influence the performance of calibration models (Jiang et al., 2007; Shi et al., 2019; Tian et al., 2021; Viljoen et al., 2005). The most popular method adopted for spectral data outlier elimination is principal component analysis (PCA), which relies on either the F-residuals or Hotelling's T^2 -values exceeding the corresponding limit of 5% (Kucheryavskiy, 2013; Shi et al., 2019; Tian et al., 2021). Alternative methods include the Mahalanobis distance, Monte Carlo cross-validation, Chauvenet's test, and hierarchical clustering algorithms in connection with PCA (Jiang et al., 2007; Ye et al., 2018). Another approach is that outliers contain spectral information of raw materials, and they are needed to ensure a robust model, so they should not be removed. In some studies, there were no significant improvements in model performance with outlier removal (Williams, 2020; Ye et al., 2018). Overall, to a large extent, outlier removal mainly depends on experience instead of a general standard.

As for dataset division, collected data are commonly divided into the calibration set (training set), the validation set, and the prediction set (test set). In the most commonly used cross-validation methods, there is no independent validation set, only a calibration set and a prediction set with the common ratio being approximately 3:1 (Jirsa et al., 2008; Li et al., 2017; Shi et al., 2019; Teye et al., 2014; Zeng et al., 2019). The principle of the dataset division is that the distribution of the measured parameter variations in each dataset is even, which means that the standard deviation values and ranges in both datasets are similar (Jiang et al., 2020; Jirsa et al., 2008; Shi et al., 2019; Zeng et al., 2019).

2.2.4 | Calibration model development and performance evaluation

The methods for calibration model development can be divided into two categories: linear regression methods and nonlinear regression methods. The principle, advantages, and disadvantages of commonly used regression methods are described in Table 3. While linear regression meth-

ods, especially partial least squares regression (PLSR), are popular, they have many practical drawbacks and require specific conditions such as linearity between sample spectra and reference values, which limit their application (Agelet & Hurburgh, 2010; Chen, Ding, et al., 2012; Manley, 2014). Indeed, the nonlinearity between sample spectra and reference values seemed unavoidable because of the violations of the Beer-Lambert law due to sample homogeneity, stray light, errors in the reference methods, and the nonlinearity between the NIR spectra and the property of interest (Centner et al., 1998; Chen, Ding, et al., 2012; Pojić & Mastilović, 2013). Therefore, it is important to evaluate how strong nonlinearity is before using linear regression methods. Useful tools for nonlinearity detection include Run Tests and the Durbin-Watson statistic approach of augmented partial residual plot (Centner et al., 1998; Mark & Workman, 2005; Zareef et al., 2020). When nonlinearity is not noticeable, linear calibration models are reliable. When nonlinearity is significant, several measures, including deleting wavelengths, adding extra principal components/latent variables, and splitting data, should be considered before proceeding with the linear calibration models (Mark & Workman, 2005). These measures were tested in some studies and improved performance. For example, the introduction of the modified restricted Boltzmann machine (MRBM) and VCPA for wavelength selection gained the best results in PLSR models (Jiang et al., 2020; Harrington, 2018).

The application of nonlinear algorithms coupled with NIRS is superior both in accuracy and robustness for nonlinear data (Zareef et al., 2020; Bian et al., 2018; Zhang et al., 2019). This approach usually requires two preconditions. The first one is a large number of samples collected from diverse sources (Williams, 2020; Cui & Fearn, 2018; Bian et al., 2018; Zhang et al., 2019; Dowell et al., 2006). This can be seen in the study of Zhang et al. (2019), which contained 775 wheat samples from 1998 to 2005. Second, some nonlinear regression methods might need to be modified to fit the wheat NIR spectra. For example, an improved deep belief network (DBN) model, a modified convolutional neural network (CNN) model, and a modified radial basic function neural network (RBFNN) model have shown feasibility and potential in model performance improvements (Li et al., 2017; Zhang et al., 2019; Mao et al., 2014). After calibration model establishment, models are optimized by the validation dataset to find the optimal model, which will then be tested by the prediction dataset to evaluate robustness. The most frequently used validation method is cross-validation, particularly leave-one-out cross-validation, in which only one spectrum and its associated reference values in the calibration set are left as the validation set and

TABLE 3 An overview of the principles, advantages, and disadvantages of frequently used regression methods

| Regression methods | Principles | Advantages | Disadvantages | References |
|--------------------|---|--|--|---|
| MLR | An extension of bivariate regression for more than two variables | Suitable for few weak or noncorrelated wavelengths | Cannot solve wavelength multicollinearity or variable codependency | Manley, 2014; Agelet & Hurburgh, 2010 |
| PCR | Spectral data are projected to new orthogonal noncorrelated dimensional axes to get the principal components (PCs) by PCA and then PCs use least squares to fit the reference values | Successfully solve multicollinearity or variable codependency | Cannot deal with nonlinear data | Mahesh et al., 2014; Zareef et al., 2020 |
| PLSR | A predictive two-block regression method based on estimated latent variables and is used to simultaneously analyze two datasets (e.g., comparison of CCA and PLS to explore and model wheat kernel samples) | Fewer latent variables; faster than PCR; models have higher precision than PCR | Overfit when reference data are noisy and model has higher complexity; model needs to rebuild with all the data when a new set of data is added | Mahesh et al., 2014 |
| CCA | Define the directions of maximum correlation between two sets of variables as latent variables to ensure the maximum explained variance in each factor | Factors arise in the order of the explained variance | Instability in regression coefficient for one reference variable | Gatius et al., 2017 |
| BPNN | Consists of an input layer, one or more hidden layers, and an output layer of neurons. Each neuron has an activation function (e.g., Tanh and Sigmoid) to forward propagate variables, and the errors are back propagated to optimize weights and biases layer by layer | Solve complex problems and automatically find out the rule of evaluation | Low speed in convergence; has the risk of overtraining; difficult to find a proper learning rate; local minima | Che et al., 2017; Liang et al., 2020 |
| RBFNN | The biggest difference of RBFNN from BPNN is the activation function in RBFNN's hidden layer has radial basis function | Faster learning speed and more suitable for regression analysis than BP-ANN | Not suitable for a large number of neurons in the input layer; difficult to choose the appropriate number of cluster centers in the hidden layer, which easily leads to "dimensionality curse" | Mao et al., 2014; Che et al., 2017 |
| ELM | Has one hidden layer; the weights and biases between the input layer and the hidden layer are randomly generated without adjustment | Faster learning speed (10–100 times) and better generalization than BP-ANN | Cannot model high-dimension data | Chen, Ding, et al., 2012; Zareef et al., 2020 |

(Continues)

TABLE 3 (Continued)

| Regression methods | Principles | Advantages | Disadvantages | References |
|--------------------|---|---|--|--|
| SVMR | Map data into a higher dimension space by a kernel function (e.g., polynomial, radial basic function, and sigmoid) and then solve the nonlinear problem by a linear tool | Obtain a great model based only on a subset of the calibration set | Nonprobabilistic sampling; various key parameters need to be set and optimized for the kernel function | Chen, Guo, et al., 2012; Teye & Huang, 2015 |
| CNN | Has convolutional layers, pooling layers, and a fully-connected layer between the input layer and output layer. The convolutional layer is used to extract features by activation functions, the pooling layer is used to descend the dimensionality of feature variables, and the fully connected layer completes the regression model | Convolutional layer can replace classical pretreatment and automatically find the best pretreatment; lower risk of overfitting than ANN | Need a large training dataset | Cui & Fearn, 2018; Zhang et al., 2019 |
| DBN | Has a plurality of restricted Boltzmann machine (RBM) and a layer of BPNN between the input layer and output layer. ***The training of RBM is only based on the spectral data. After all of the RBM get trained, the BPNN is trained with the initial values from RBM and reference values | Feature extraction can be performed by the RBM | Need a large number of training dataset | Li et al., 2017 |
| RFR | Many regression tree models are built from bootstrap samples of the calibration dataset. At each node, only a subset of reference values is randomly selected to split and fit spectral data | High tolerance to outliers and noise; low computational cost; easy to explain; resistance to overfitting; strong generalizability | Weight on each variable is to a great extent randomly influenced | Breiman, 2001 |

the rest are used for calibration set, and that is then done repeatedly until all spectra have been left out once and used for validation (Che et al., 2017; Chen, Guo, et al., 2012).

The evaluation parameters of model performance include the following: coefficient of determination for calibration (R_C^2); correlation coefficient for prediction (r_C); root mean square error of calibration (RMSEC); standard error of calibration (SEC); coefficient of determination for cross-validation (R_{CV}^2); correlation coefficient for prediction (r_{CV}); root mean square error of cross-validation (RMSECV); standard error of cross-validation (SECV); coefficient of determination for prediction (R_P^2); correlation coefficient for prediction (r_P); root mean square error of prediction (RMSEP); standard error of prediction (SEP); ratio of prediction to deviation (RPD); and the range error ratio (RER) (Shi et al., 2019; Gabriel et al., 2017).

3 | APPLICATION OF NIRS IN WHEAT QUALITY EVALUATION

3.1 | Types of wheat product samples suitable for NIRS

The methodologies of NIRS application, including spectra acquisition, pretreatment, wavelength selection, and model development, are generally the same for kernels, whole meals, and flours. Specific methods are universal to some extent, but the optimal model is usually obtained by trial and error to combine these methods.

The goal of NIRS assessment is different for kernels and for flour (Pojić & Mastilović, 2013). Kernel assessment, especially for single kernels, is more likely to be used for breeding purposes and is primarily concerned with protein content and its related parameters such as gluten content and gluten index (Armstrong, 2014; Başlar & Ertugay, 2011; Bian et al., 2018; Chen et al., 2021; Harrington, 2018; Jirsa et al., 2008; Mao et al., 2014; Shi & Yu, 2017; Sinelli et al., 2011). For flour, assessments usually focus on processing, storage, dough making, and end product quality properties. Therefore, the related parameters usually include starch content, ash content, protein content, lipid content, moisture, rheological properties, bioactive compounds, and baking performance (Dong & Sun, 2013; Ferreira et al., 2015; Finney et al., 1988; Garnsworthy et al., 2000; Li et al., 2017; Liu et al., 2015; Miralbé, 2004; Morgan & Williams, 1995; Mutlu et al., 2011; Owens et al., 2009; Shi et al., 2019). Recently, researchers explored spectra from single kernel or bulk samples, and some advances were also made in parameter prediction that was previously studied for early screening (Arazuri et al., 2012; Gabriel et al., 2017; Mishra et al., 2018; Mutlu et al., 2011; Peiris et al., 2017; Shi & Yu, 2017; Sinelli et al., 2011).

Bran, the major by-product in the milling industry, was first studied by NIRS for its fiber content in 1984 (Horvath et al., 1984). Recently, as bran's nutritional values were further recognized, studies of wheat bran have extended to other compositions such as protein content, moisture, starch, and lipids (Hell et al., 2016; Sujka et al., 2017). In addition, dough, bread, cake, and pasta are also considered samples of spectral acquisition, though NIRS application in them is not as popular as that in kernels and flour. In these products, studies mainly focused on physical attributes (e.g., loaf volume, moisture, and hardness) and nutritional parameters.

3.2 | Quantitative determination of analytical quality parameters

Analytical quality attributes can be divided into two categories. The first category is the contents of various components, such as protein, moisture, starch, lipid, and fiber. Each of these compositions has specific chemical structures and therefore specific classical absorptions in NIR spectra. The approximate wavelength absorption ranges of different components used in NIRS assessment are shown in Figure 2. The absorption wavelengths of the specific chemical structures can be used to interpret the performance of linear regression models (e.g., the PLSR model), but this will not work for nonlinear regression models (Chen et al., 2008; Liang et al., 2022; Du et al.,). The second category is physical parameters other than composition. These include physical parameters of wheat (e.g., hardness and test weight) or physical parameters correlating to compositions (e.g., water absorption, gluten index) (Başlar & Ertugay, 2011). An overview of NIRS applications in the quantitative determination of compositions and other physical parameters of wheat kernel and flour in the past 10 years is summarized in Table 4.

3.2.1 | Composition determination

Protein

Protein content is the most important quality indicator of wheat and flour and has attracted the most research attention (Table 4). Protein content influences the quality of end products, especially baked products (Ahmad et al., 2016). Its determination using NIRS is a proven technique according to model performances shown in Table 4, and the NIRS has been a standard method since 1985 (Cereals & Grains Association, 1999a, 1999b).

PLSR as a traditional linear regression method still accounted for a large proportion of model development for protein content determination (Shi & Yu, 2017; Mishra

TABLE 4 An overview of NIR application in quantitative determination of wheat compositions and physical parameters

| Application | Chemometrics | | | | | | | | | |
|-------------|------------------|-----------------------------------|-----------------------------|----------------|--------------------------------------|--|---------------|-------|---|------------------------|
| | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Protein | Reflectance | Si-Ware NIR Systems | 1298–2606 | Flour | 150 (112/38) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .9488$; RMSEP = 0.22; $R_{CV}^2 = .9518$; RMSECV = 0.22; $R_C^2 = .9624$; RMSEC = 0.19 | Chen et al., 2021 |
| Protein | Reflectance | Viavi MicroNIR 1700 | 908–1676 | Flour | 150 (112/38) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .9422$; RMSEP = 0.24; $R_{CV}^2 = .9400$; RMSECV = 0.25; $R_C^2 = .9531$; RMSEC = 0.21 | Chen et al., 2021 |
| Protein | Reflectance | Spectral Engines NIR 2.0-W | 1550–1950 | Flour | 150 (112/38) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .8370$; RMSEP = 0.42; $R_{CV}^2 = .8491$; RMSECV = 0.39; $R_C^2 = .8843$; RMSEC = 0.34 | Chen et al., 2021 |
| Protein | Reflectance | NIR MPA Spectrometer | 800.7–1838.9; 2171.6–2361.3 | Kernel | 128 (108/20) | Single variety | FD | PLSR | $R_p^2 = .938938$; RMSEP = 0.248; RPD = 3.971 | Mishra et al., 2018 |
| Protein | Reflectance | CDI 256L-1.7T1 spectrometer | 906–1683 | Kernel | 336 (240/96) | Five varieties from 1998 to 2005 | SMC + SNV | PLSR | $R_p^2 = .93$; RMSEP = 0.50–0.72; RPD = 3.14–4.86 | Armstrong, 2014 |
| Protein | Reflectance | Bruker MPA FT-NIR spectrometer | 940–2700 | Kernel | 547 (289/157) | 30 varieties from eight different agro-climate areas | MSC + FD + OR | PLSR | $r_p = .98$; RMSEP = 0.20; RPD = 4.9; $r_{CV} = .97$; RMSECV = 0.24; RPD = 4.3 | Sinelli et al., 2011 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 120 (80/40) | Collected from different regions | SD | PLSR | SEP = 0.377 | Başlar & Ertugay, 2011 |

(Continues)

TABLE 4 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|-------------------------------------|---------------------------------|----------------|--------------------------------------|--|----------------|-------|---|---------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 722 (644/100) | Collected from a commercial flour company | None | PLSR | RMSEP = 0.24; RMSECV = 0.23 | Gatius et al., 2017 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 722 (644/100) | Collected from a commercial flour company | None | CCA | RMSEP = 0.35; RMSECV = 0.34 | Gatius et al., 2017 |
| Protein | Reflectance | SpectraStar 2500XL-R Spectrometer | 1400–2500 selected by RCA | Flour | 48 (32/16) | Collected from 2 years and grounded by the same mills | FD + SNV + OR | PLSR | $R_p^2 = .98$; RMSEP = 0.22; SEP = 0.22; RPD = 8.0; $R_{cv}^2 = .98$; RMSECV = 0.31; SECV = 0.32 | Shi et al., 2019 |
| Protein | Reflectance | SpectraStar 2500XL-R Spectrometer | 1400–2500 selected by RCA | Flour | 71 (48/23) | Collected randomly during May and June 2016 | SNV + FD + OR | PLSR | $R_p^2 = .98$; RMSEP = 0.31; SEP = 0.32; RPD = 6.09; $R_{cv}^2 = .97$; RMSECV = 0.33; SECV = 0.34 | Shi & Yu, 2017 |
| Protein | Reflectance | SpectraStar 2500XL-R Spectrometer | 1980–2250 selected by RCA | Flour | 71 (48/23) | Collected randomly during May and June 2016 | OR | PLSR | $R_p^2 = .97$; RMSEP = 0.35; SEP = 0.36; RPD = 5.37; $R_{cv}^2 = .96$; RMSECV = 0.39; SECV = 0.39 | Shi & Yu, 2017 |
| Protein | Reflectance | FOSS InfraTec 1241 NIR spectrometer | 13 wave-lengths selected by SPA | Kernel | 112 (102/10) | Collected from six regions along Yangtze river in 2013 | SGA + SD + SNV | PLSR | $R_p^2 = .9980$; RPD = 22.2275; SEP = 0.0639 | Ye et al., 2018 |

(Continues)

TABLE 4 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|---|---------------------------------------|----------------|--------------------------------------|---|----------------|-------|---|---------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Protein | Reflectance | FOSS InfraTec 1241 NIR spectrometer | 10 wave-lengths selected by SPA | Kernel | 112 (102/10) | Collected from six regions along Yangtze river in 2013 | SGA + SD + SNV | PLSR | $R_p^2 = .9978$; RPD = 21.3482; SEP = 0.0665 | Ye et al., 2018 |
| Protein | Reflectance | FOSS InfraTec 1241 NIR spectrometer | 11 wave-lengths selected by ISA + SPA | Kernel | 112 (102/10) | Collected from six regions along Yangtze river in 2013 | SGA + SD + SNV | PLSR | $R_p^2 = .9986$; RPD = 26.9075; SEP = 0.0528 | Ye et al., 2018 |
| Protein | Reflectance | Instruments from three companies (Bruins FOSS and Perten) | Selected by MRBM | Kernel | 2529 (2190/339) | Collected around USA | None | PLSR | $R_p^2 = .991$; SEP = 0.194; | Harrington, 2018 |
| Protein | Reflectance | Bruker MPA FT-NIR spectrometer | Full range | Kernel | 100 (80/20) | Clean samples and insect-infested samples | NL | PLSR | $R_{CV}^2 = .9179$ | Pandey et al., 2018 |
| Protein | Reflectance | Bruker Multi-Purpose NIR Analyzer | Three intervals selected by Si-PLS | Flour | 96 (72/24) | Collected from supermarkets and small-scale mills | SNV | PLSR | $R_p^2 = .885$; RMSEP = 0.459; SEP = 0.469 | Chen et al., 2017 |
| Protein | Reflectance | Bruker MPA Multi-Purpose NIR Analyzer | Seven intervals selected by Si-PLS | Flour | 96 (72/24) | Collected from supermarkets and small-scale mills | SNV + SD | SVMIR | $R_p^2 = .906$; RMSEP = 0.425; SEP = 0.434 | Chen et al., 2017 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | RFR | $R_p^2 = .988$; SEP = 0.153; RPD = 10.1 | Williams, 2020 |

(Continues)

TABLE 4 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|-----------------------------------|--------------------|----------------|--------------------------------------|---|------------|--------------|---|--------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | SVMR | $R_p^2 = .99$; SEP = 0.18; RPD = 8.73 | Williams, 2020 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 720–1700 | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | SD | BNNN | $R_p^2 = .988$; SEP = 0.179; RPD = 8.78 | Williams, 2020 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 600–2200 | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | FD | PLSR | $R_p^2 = .98$; SEP = 0.177; RPD = 8.88 | Williams, 2020 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2498 | Kernel | 882 (775/107) | Five varieties collected from 1998 to 2005 | OR | PLSR | $R_p^2 = .9213$; RMSEP = 0.6986; RPD = 2.547 | Bian et al., 2018 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2498 | Kernel | 882 (775/107) | Five varieties collected from 1998 to 2005 | OR | NNRW | $R_p^2 = .9526$; RMSEP = 0.4776; RPD = 3.281 | Bian et al., 2018 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2498 | Kernel | 882 (775/107) | Five varieties collected from 1998 to 2005 | OR | BNNRW | $R_p^2 = .9239$; RMSEP = 0.6505; RPD = 2.634 | Bian et al., 2018 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2498 | Kernel | 882 (775/107) | Five varieties collected from 1998 to 2005 | OR | RBNNRW | $R_p^2 = .9767$; RMSEP = 0.3703; RPD = 4.669 | Bian et al., 2018 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2498 | Kernel | 882 (775/107) | Five varieties collected from 1998 to 2005 | SNV + SGA | Modified CNN | $R_p^2 = .98 \pm .002$; MSEP = 0.2 ± 0.01 | Zhang et al., 2019 |

(Continues)

TABLE 4 (Continued)

| | | Chemometrics | | | | | | | | |
|-------------|------------------|--|------------------------------|----------------|--------------------------------------|--|---------------|----------------|--|------------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 850–1650 | Flour | 1597 (1000/597) | Different varieties and origins | NL | CNN | RMSEP = 0.23; SEP = 0.229 | Cui & Fearn, 2018 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 11 wave-lengths by PSO | Kernel | 140 (100/40) | Collected from a dozen of major wheat-producing areas in 2010 | None | Improved RBFNN | $R_p^2 = .975$; RMSEP = 0.2657 | Mao et al., 2014 |
| Protein | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | ANN | $R_p^2 = .991$; SEP = 0.952 | Mutlu et al., 2011 |
| Protein | Reflectance | SupNIR-1520 near infrared spectrometer | 1000–2500 | Flour | 400 (300/100) | Unknown | WT | BPNN | $r_p = .9212$; RMSEP = 0.5236 | Li et al., 2017 |
| Protein | Reflectance | SupNIR-1520 near infrared spectrometer | 1000–2500 | Flour | 400 (300/100) | Unknown | WT | IDBN | $r_p = .9928$; RMSEP = 0.0628 | Li et al., 2017 |
| Protein | Reflectance | Bruker MPA FT-NIR spectrometer | 1063.5–1333.7; 1648.8–2175.3 | Bran | 88 (44/44) | Collected from four different sources | FD + NL + OR | PLSR | $R_p^2 = .66$; RMSEP = 0.96; $R_{CV}^2 = .87$; RMSECV = 0.70 | Hell et al., 2016 |
| Dry gluten | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 120 (80/40) | Collected from different regions | SD | PLSR | SEP = 0.635 | Başlar & Ertugay, 2011 |
| Dry gluten | Reflectance | Bruker FT-NIR spectrometer | 940–2700 | Kernel | 547 (289/157) | 30 varieties collected from eight different agro-climate areas | SNV + FD + OR | PLSR | $r_p = .94$; RMSEP = 0.29; RPD = 3.0; $r_{CV} = .98$; RMSECV = 0.20; RPD = 2.3 | Sinelli et al., 2011 |

(Continues)

TABLE 4 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|-----------------------------------|------------------------------------|----------------|--------------------------------------|---|------------|-------|---|------------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Dry gluten | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 722 (644/100) | Collected from a commercial flour company | None | PLSR | RMSEP = 0.23; RMSECV = 0.21 | Gatius et al., 2017 |
| Dry gluten | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 722 (644/100) | Collected from a commercial flour company | None | CCA | RMSEP = 0.15; RMSECV = 0.14 | Gatius et al., 2017 |
| Wet gluten | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 120 (80/40) | Collected from different regions | SD | PLSR | SEP = 1.36 | Başlar & Ertugay, 2011 |
| Wet gluten | Reflectance | Bruker Multi-Purpose NIR Analyzer | 800–3300 | Flour | 96 (72/24) | Collected from supermarkets and small-scale mills | SNV + FD | SVMR | $R_p^2 = .781$; RMSEP = 1.303; SEP = 1.331 | Chen et al., 2017 |
| Wet gluten | Reflectance | Bruker Multi-Purpose NIR Analyzer | 800–3300 | Flour | 96 (72/24) | Collected from supermarkets and small-scale mills | SNV + FD | PLSR | $R_p^2 = .675$; RMSEP = 1.527; SEP = 1.56 | Chen et al., 2017 |
| Wet gluten | Reflectance | Bruker Multi-Purpose NIR Analyzer | Three intervals selected by Si-PLS | Flour | 96 (72/24) | Collected from supermarkets and small-scale mills | SNV + FD | SVMR | $R_p^2 = .850$; RMSEP = 1.024; SEP = 1.046 | Chen et al., 2017 |
| Wet gluten | Reflectance | Bruker Multi-Purpose NIR Analyzer | Four intervals selected by Si-PLS | Flour | 96 (72/24) | Collected from supermarkets and small-scale mills | None | PLSR | $R_p^2 = .779$; RMSEP = 1.283; SEP = 1.31 | Chen et al., 2017 |

(Continues)

TABLE 4 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|-----------------------------------|--------------------------|----------------|--------------------------------------|--|---------------|-------|---|-------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Wet gluten | Reflectance | Si-Ware NIR Systems | 1298–2606 | Flour | 192 (144/48) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .7874$; RMSEP = 0.66; $R_{CV}^2 = .836$; RMSECV = 0.73; $R_C^2 = .7675$; RMSEC = 0.72 | Chen et al., 2021 |
| Wet gluten | Reflectance | Viavi MicroNIR 1700 | 908–1676 | Flour | 192 (144/48) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .8585$; RMSEP = 0.57; $R_{CV}^2 = .8515$; RMSECV = 0.65; $R_C^2 = .7675$; RMSEC = 0.66 | Chen et al., 2021 |
| Wet gluten | Reflectance | Spectral Engines NIR 2.0-W | 1550–1950 | Flour | 192 (144/48) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .6117$; RMSEP = 0.83; $R_{CV}^2 = .6701$; RMSECV = 0.91; $R_C^2 = .7675$; RMSEC = 0.89 | Chen et al., 2021 |
| Moisture | Reflectance | Thermo Antaris II FT-NIR | Two intervals by PCC | Flour | 169 (127/42) | Collected from different factories | OR | PLSR | $R_p^2 = .895$; RMSEP = 0.020; $R_{CV}^2 = .890$; RMSECV = 0.022 | Dong & Sun, 2013 |
| Moisture | Reflectance | Thermo Antaris II FT-NIR | Two intervals by i-PLS | Flour | 169 (127/42) | Collected from different factories | OR | PLSR | $R_p^2 = .911$; RMSEP = 0.019; $R_{CV}^2 = .923$; RMSECV = 0.019 | Dong & Sun, 2013 |
| Moisture | Reflectance | CDI 256L-1.7T1 spectrometer | 906–1683 | kernel | 300 (180/120) | Five different varieties | SMC + SNV | PLSR | $R_p^2 = .978$; RMSEP = 0.49–0.5; RPD = 4.4–4.8 | Armstrong, 2014 |
| Moisture | Reflectance | SpectraStar 2500XL-R Spectrometer | 680–2500 selected by RCA | Flour | 54 (31/23) | Collected randomly during May and June 2016 | DT + OR | PLSR | $R_p^2 = .83$; RMSEP = 0.21; SEP = 0.22; RPD = 2.39; $R_{CV}^2 = .88$; RMSECV = 0.22; SECY = 0.23 | Shi & Yu, 2017 |

(Continues)

TABLE 4 (Continued)

| | | Chemometrics | | | | | | | | | |
|-------------|------------------|-----------------------------------|-----------------------------|----------------|--------------------------------------|---|---------------|-------|---|---------------------|--|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References | |
| Moisture | Reflectance | SpectraStar 2500XL-R Spectrometer | 1850–2200 selected by RCA | Flour | 54 (31/23) | Collected randomly during May and June 2016 | FD + SNV + OR | PLSR | $R_p^2 = .86$; RMSEP = 0.20; SEP = 0.20; RPD = 2.56; $R_{CV}^2 = .88$; RMSECV = 0.22; SECV = 0.22 | Shi & Yu, 2017 | |
| Moisture | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 722 (644/100) | Collected from a commercial flour company | None | PLSR | RMSEP = 0.10; RMSECV = 0.10 | Gatius et al., 2017 | |
| Moisture | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 722 (644/100) | Collected from a commercial flour company | None | CCA | RMSEP = 0.10; RMSECV = 0.10 | Gatius et al., 2017 | |
| Moisture | Reflectance | Perkin 400 FT-NIR spectrometer | 1000–2500 | Kernel | 396 (40/356) | 27 varieties collected in 4 years | None | PLSR | $R_p^2 = .98$; RMSEP = 0.19; $R_{CV}^2 = .99$; RMSECV = 0.10 | Peiris et al., 2017 | |
| Moisture | Reflectance | NIR MPA™ Spectrometer | 800.7–1335.1; 1831.2–2361.3 | Kernel | 128 (108/20) | Single variety | FD | PLSR | $R_p^2 = .901$; RMSEP = 0.485; RPD = 3.108 | Mishra et al., 2018 | |
| Moisture | Reflectance | Bruker MPA FT-NIR spectrometer | Full range | Kernel | 100 (80/20) | Clean samples and insect-infested samples | NL | PLSR | $R_{CV}^2 = .9526$ | Pandey et al., 2018 | |

(Continues)

TABLE 4 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|--|--------------------|----------------|--------------------------------------|--|---------------|-------|---|-------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Moisture | Reflectance | Si-Ware NIR Systems | 1298–2606 | Flour | 239 (179/60) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .7079$; RMSEP = 0.42; $R_{cv}^2 = .7406$; RMSECV = 0.39; $R_c^2 = .778$; RMSEC = 0.38 | Chen et al., 2021 |
| Moisture | Reflectance | Viavi MicroNIR 1700 | 908–1676 | Flour | 239 (179/60) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .8175$; RMSEP = 0.32; $R_{cv}^2 = .8047$; RMSECV = 0.39; $R_c^2 = .8367$; RMSEC = 0.33 | Chen et al., 2021 |
| Moisture | Reflectance | Spectral Engines NIR 2.0-W | 1550–1950 | Flour | 239 (179/60) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .6361$; RMSEP = 0.46; $R_{cv}^2 = .6366$; RMSECV = 0.39; $R_c^2 = .6839$; RMSEC = 0.44 | Chen et al., 2021 |
| Moisture | Reflectance | SupNIR-1520 near infrared spectrometer | 1000–2500 | Flour | 400 (300/100) | Unknown | WT | BPNN | $r_p = .9322$; RMSEP = 0.3245 | Li et al., 2017 |
| Moisture | Reflectance | SupNIR-1520 near infrared spectrometer | 1000–2500 | Flour | 400 (300/100) | Unknown | WT | DBN | $r_p = .9327$; RMSEP = 0.2523 | Li et al., 2017 |

(Continues)

TABLE 4 (Continued)

| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Chemometrics | | | References |
|--------------|------------------|--|---------------------------------|----------------|--------------------------------------|--------------------------------------|---------------|-------|---|--------------------|
| | | | | | | | Preprocess | Model | Performance | |
| Moisture | Reflectance | SupNIR-1520 near infrared spectrometer | 1000–2500 | Flour | 400 (300/100) | Unknown | WT | IDBN | $r_p = .997$; RMSEP = 0.0069 | Li et al., 2017 |
| Moisture | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .991$ | Mutlu et al., 2011 |
| Moisture | Reflectance | Bruker MPA FT-NIR spectrometer | 1063.5–1836.2; 2173.4–2355.2 | Bran | 88 (44/44) | Four different sources | LS + OR | PLSR | $R_p^2 = .78$; RMSEP = 1.1; $R_{CV}^2 = .87$; RMSECV = 0.95 | Hell et al., 2016 |
| Starch | Reflectance | Bruker MPA FT-NIR spectrometer | 1063.5–1639.9; 1834.9–2175.3 | Bran | 88 (44/44) | Four different sources | FD + OR | PLSR | $R_p^2 = .88$; RMSEP = 1.92; $R_{CV}^2 = .94$; RMSECV = 1.25 | Hell et al., 2016 |
| Soluble DF | Reflectance | Bruker MPA FT-NIR spectrometer | 932.9–1016.3; 1648.8–2175.3 | Bran | 88 (44/44) | Four different sources | FD + MSC + OR | PLSR | $R_p^2 = .77$; RMSEP = 0.55; $R_{CV}^2 = .84$; RMSECV = 0.52 | Hell et al., 2016 |
| Insoluble DF | Reflectance | Bruker MPA FT-NIR spectrometer | 1333.0–1639.9 | Bran | 88 (44/44) | Four different sources | FD + OR | PLSR | $R_p^2 = .93$; RMSEP = 1.79; $R_{CV}^2 = .98$; RMSECV = 1.15 | Hell et al., 2016 |

(Continues)

TABLE 4 (Continued)

| | | Chemometrics | | | | | | | | |
|-------------|------------------|---------------------------------------|---|----------------|--------------------------------------|-------------------------------|------------|-------|---|---------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| Lipids | Reflectance | Bruker MPA FT-NIR spectrometer | 1063.5–1639.9 | Bran | 88 (44/44) | Four different sources | LS + OR | PLSR | $R_p^2 = .87$; RMSEP = 0.26; $R_{CV}^2 = .91$; RMSECV = 0.25 | Hell et al., 2016 |
| Phenolics | Reflectance | Thermo Antaris II FT-NIR spectrometer | 1000–2500 | Flour | 107 (81/26) | Collected in 2018 and 2019 | MSC + FD | PLSR | $R_p^2 = .90$; RMSEP = 70.5; RPD = 3.4; $R_{CV}^2 = .92$ | Tian et al., 2020 |
| Phenolics | Reflectance | Thermo Antaris II FT-NIR spectrometer | 1333–2222 | Flour | 107 (81/26) | Collected in 2018 and 2019 | FD | PLSR | $R_p^2 = .88$; RMSEP = 78.5; RPD = 3; $R_{CV}^2 = .92$ | Tian et al., 2020 |
| Fatty acid | Reflectance | NIRQuest512 spectrometer | Five wave-lengths selected by VCPA | Flour | 120 (90/30) | Bought from supermarket | SNV | ELM | $R_p^2 = .951$ –.972; RMSEP = 0.8645–1.1531 | Jiang et al., 2020 |
| Fatty acid | Reflectance | NIRQuest512 spectrometer | Wavelengths selected by 50 runs of VCPA | Flour | 120 (90/30) | Bought from supermarket | SNV | PLSR | $R_p^2 = .967$ –.972; RMSEP = 0.868–0.939 | Jiang et al., 2020 |
| Uric acid | Reflectance | NIR MPA™ Spectrometer | 833.1–1429.2; 2019.1–2502.5 | Kernel | 128 (108/20) | Single variety | FD + VN | PLSR | $R_p^2 = .895$; RMSEP = 3.034; RPD = 3.034 | Mishra et al., 2018 |

(Continues)

TABLE 4 (Continued)

| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Chemometrics | | | |
|-------------|------------------|---------------------------------------|------------------------|----------------|--------------------------------------|--|---------------|-------|--|---------------------|
| | | | | | | | Preprocess | Model | Performance | References |
| Fat | Reflectance | Bruker MPA FT-NIR spectrometer | Full range | Kernel | 100 (80/20) | Clean samples and insect-infested samples | NL | PLSR | $R_{CV}^2 = .8275$ | Pandey et al., 2018 |
| Ash | Reflectance | Thermo Antaris II FT-NIR spectrometer | Two intervals by PCC | Flour | 170 (128/42) | Collected from different factories | OR | PLSR | $R_p^2 = .990$; RMSEP = 0.090; $R_{CV}^2 = .990$; RMSECV = 0.102 | Dong & Sun, 2013 |
| Ash | Reflectance | Thermo Antaris II FT-NIR spectrometer | Two intervals by i-PLS | Flour | 170 (128/42) | Collected from different factories | OR | PLSR | $R_p^2 = .990$; RMSEP = 0.090; $R_{CV}^2 = .990$; RMSECV = 0.102 | Dong & Sun, 2013 |
| Ash | Reflectance | Si-Ware NIR Systems | 1298–2606 | Flour | 124 (93/31) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .9225$; RMSEP = 0.07; $R_{CV}^2 = .9144$; RMSECV = 0.07; $R_C^2 = .9367$; RMSEC = 0.06 | Chen et al., 2021 |
| Ash | Reflectance | Viavi MicroNIR 1700 | 908–1676 | Flour | 124 (93/31) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .9055$; RMSEP = 0.06; $R_{CV}^2 = .9162$; RMSECV = 0.06; $R_C^2 = .9431$; RMSEC = 0.05 | Chen et al., 2021 |
| Ash | Reflectance | Spectral Engines NIR 2.0-W | 1550–1950 | Flour | 124 (93/31) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .8106$; RMSEP = 0.08; $R_{CV}^2 = .8739$; RMSECV = 0.08; $R_C^2 = .9081$; RMSEC = 0.07 | Chen et al., 2021 |
| Ash | Reflectance | Unknown | 850–1650 | Flour | 7605 (6987/618) | Different varieties and origins | DT + SNV | PLSR | RMSEP = 0.094; SEP = 0.088 | Cui & Fearn, 2018 |

(Continues)

TABLE 4 (Continued)

| | | | | | | | | | | Chemometrics | | |
|-------------|------------------|--|--------------------|----------------|--------------------------------------|--|---------------|-------|---|----------------------|--|--|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References | | |
| Ash | Reflectance | SupNIR-1520 near infrared spectrometer | 1000–2500 | Flour | 400 (300/100) | Unknown | WT | BPNN | $r_p = .8881$; RMSEP = 0.6525 | Li et al., 2017 | | |
| Ash | Reflectance | SupNIR-1520 near infrared spectrometer | 1000–2500 | Flour | 400 (300/100) | Unknown | WT | DBN | $r_p = .9121$; RMSEP = 0.3225 | Li et al., 2017 | | |
| Ash | Reflectance | SupNIR-1520 near infrared spectrometer | 1000–2500 | Flour | 400 (300/100) | Unknown | WT | IDBN | $r_p = .9920$; RMSEP = 0.0535 | Li et al., 2017 | | |
| Ash | Reflectance | Unknown | 850–1650 | Flour | 7605 (6987/618) | Different varieties and origins | NL | CNN | RMSEP = 0.085; SEP = 0.075 | Cui & Fearn, 2018 | | |
| Ash | Reflectance | Bruker MPA FT-NIR spectrometer | 1648.8–1836.2 | Bran | 88 (44/44) | Four different sources | FD + MSC + OR | PLSR | $R_p^2 = .91$; RMSEP = 0.33; $R_{cv}^2 = .96$; RMSECV = 0.25 | Hell et al., 2016 | | |
| Ash | Reflectance | Bruker MPA FT-NIR spectrometer | Full range | Kernel | 100 (80/20) | Clean samples and insect-infested samples | NL | PLSR | $R_{cv}^2 = .8318$ | Pandey et al., 2018 | | |
| GI | Reflectance | Bruker FT-NIR spectrometer | 830–2700 | Kernel | 547 (289/157) | 30 varieties collected from eight different agro-climate areas | SNV + FD + OR | PLSR | $r_p = .82$; RMSEP = 5.4; RPD = 1.8; $r_{cv} = .88$; RMSECV = 3.9; RPD = 2.0 | Sinelli et al., 2011 | | |

(Continues)

TABLE 4 (Continued)

| | | Chemometrics | | | | | | | | |
|-------------|------------------|-----------------------------------|--------------------|----------------|--------------------------------------|--|---------------|-------|---|------------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| GI | Reflectance | Si-Ware NIR Systems | 1298–2606 | Flour | 124 (93/31) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .8185$; RMSEP = 2.12; $R_{cv}^2 = .8084$; RMSECV = 2.3; $R_C^2 = .8789$; RMSEC = 1.80 | Chen et al., 2021 |
| GI | Reflectance | Viavi MicroNIR 1700 | 908–1676 | Flour | 124 (93/31) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .7217$; RMSEP = 2.61; $R_{cv}^2 = .7155$; RMSECV = 2.73; $R_C^2 = .816$; RMSEC = 2.16 | Chen et al., 2021 |
| GI | Reflectance | Spectral Engines NIR 2.0-W | 1550–1950 | Flour | 124 (93/31) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .6309$; RMSEP = 3.0; $R_{cv}^2 = .6497$; RMSECV = 3.06; $R_C^2 = .7675$; RMSEC = 2.41 | Chen et al., 2021 |
| ZS | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .917$ | Mutlu et al., 2011 |
| ZS | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 120 (80/40) | Collected from different regions | SD | MLR | SEP = 3.74 | Başlar & Ertugay, 2011 |
| ZS | Reflectance | Si-Ware NIR Systems | 1298–2606 | Flour | 196 (148/48) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .8185$; RMSEP = 1.80; $R_{cv}^2 = .8084$; RMSECV = 2.30; $R_C^2 = .8789$; RMSEC = 2.12 | Chen et al., 2021 |

(Continues)

TABLE 4 (Continued)

| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Chemometrics | | | References |
|-------------|------------------|-----------------------------------|----------------------------------|----------------|--------------------------------------|--|---------------|-------|---|---------------------|
| | | | | | | | Preprocess | Model | Performance | |
| ZS | Reflectance | Viavi MicroNIR 1700 | 908–1676 | Flour | 196 (148/48) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .7217$; RMSEP = 2.16; $R_{CV}^2 = .7155$; RMSECV = 2.73; $R_C^2 = .816$; RMSEC = 2.61 | Chen et al., 2021 |
| ZS | Reflectance | Spectral Engines NIR 2.0-W | 1550–1950 | Flour | 196 (148/48) | Various varieties collected in 2017 and 2019 | FD + SNV + OR | PLSR | $R_p^2 = .6309$; RMSEP = 2.41; $R_{CV}^2 = .6497$; RMSECV = 3.06; $R_C^2 = .7675$; RMSEC = 3.00 | Chen et al., 2021 |
| HD | Reflectance | NIR MPA™ Spectrometer | 800.7–1335.1 | Kernel | 128 (108/20) | Single variety | FD + LS | PLSR | $R_p^2 = .912$; RMSEP = 0.762; RPD = 3.29 | Mishra et al., 2018 |
| HD | Reflectance | Gu Jia Shen NIR camera | 900–1700 | Kernel | 1540 (880/660) | 22 varieties | SGA + FD | SVMR | $R_p^2 = .8492$; RMSEP = 0.1454; $R_{CV}^2 = .8505$; RMSECV = 0.1448 | Zhang et al., 2017 |
| HD | Reflectance | Gu Jia Shen NIR camera | Four intervals selected by i-PLS | Kernel | 1540 (880/660) | 22 varieties | SGA + FD | SVMR | $R_p^2 = .9174$; RMSEP = 0.1100; $R_{CV}^2 = .9180$; RMSECV = 0.1090 | Zhang et al., 2017 |
| HD | Reflectance | Gu Jia Shen NIR camera | 4 intervals selected by ACO | Kernel | 1540 (880/660) | 22 varieties | SGA + FD | SVMR | $R_p^2 = .9544$; RMSEP = 0.0590; $R_{CV}^2 = .9810$; RMSECV = 0.382 | Zhang et al., 2017 |
| HD | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | | 862 (755/107) | Seven different regions from 1998 to 2005 | None | RFR | $R_p^2 = .826$; SEP = 1/74.; RPD = 2.70 | Williams, 2020 |

(Continues)

TABLE 4 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|-----------------------------------|--------------------|----------------|--------------------------------------|---|------------|-------|---|---------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| HD | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | SVMR | $R_p^2 = .810$; SEP = 2.09; RPD = 2.25 | Williams, 2020 |
| HD | Reflectance | FOSS NIR System 6500 Spectrometer | 720–1700 | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | SGA + FD | PLSR | $R_p^2 = .813$; SEP = 2.07; RPD = 2.30 | Williams, 2020 |
| HD | Reflectance | Bruker MPA FT-NIR spectrometer | Full range | Kernel | 100 (80/20) | Clean samples and insect-infested samples | NL | PLSR | $R_{CV}^2 = .8078$ | Pandey et al., 2018 |
| TW | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | RFR | $R_p^2 = .756$; SEP = 0.78; RPD = 2.7 | Williams, 2020 |
| TW | Reflectance | FOSS NIR System 6500 Spectrometer | 600–2200 | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | SGA + FD | PLSR | $R_p^2 = .660$; SEP = 0.96; RPD = 1.6 | Williams, 2020 |

(Continues)

TABLE 4 (Continued)

| | | | | | | | Chemometrics | | | |
|-------------|------------------|-----------------------------------|-----------------------------|----------------|--------------------------------------|---|--------------|-------|--|---------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Preprocess | Model | Performance | References |
| TW | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | SVMR | $R_p^2 = .550$; SEP = 0.96; RPD = 1.6 | Williams, 2020 |
| KM | Reflectance | CDI 256L-1.7T1 spectrometer | 906–1683 | kernel | 300 (200/100) | Five varieties | SMC | PLSR | $R_p^2 = .86$; RMSEP = 2.8–4.0 | Armstrong, 2014 |
| TKW | Reflectance | NIR MPA™ Spectrometer | 800.7–1335.1; 2171.6–2361.3 | Kernel | 128 (108/20) | Single variety | SD | PLSR | $R_p^2 = .907$; RMSEP = 0.576; RPD = 3.03 | Mishra et al., 2018 |
| TKW | Reflectance | Bruker MPA FT-NIR spectrometer | Full range | Kernel | 100 (80/20) | Clean samples and insect-infested samples | NL | PLSR | $R_{CV}^2 = .903$ | Pandey et al., 2018 |

et al., 2018; Shi et al., 2019; Williams, 2020; Mishra et al., 2018; Hell et al., 2016; Ye et al., 2018). These calibration models have achieved stable and satisfactory determination, and the R_p^2 was generally above .9 except for the determination of wheat bran (Shi & Yu, 2017; Mishra et al., 2018; Armstrong, 2014; Shi et al., 2019; Chen et al., 2017; Williams, 2020; Hell et al., 2016). Overall, the performance of the PLSR model was mainly influenced by sample size and sample quality (e.g., the variance and uniformity) (Chen et al., 2017). Various wavelength selection methods have been employed and optimized the performance of the PLSR model. Representative selection methods, such as regression coefficient analysis (RCA), successive projections algorithm (SPA), improved simulated annealing (ISA), and MRBM, had powerful effects on PLSR models, making their performance comparable to those of nonlinear regression methods (Harrington, 2018; Shi & Yu, 2017). Two practical studies that promoted NIRS technology are worth mentioning. The first study designed an automatic NIR spectral collection instrument of single wheat kernels and a built-in grain quality evaluation system based on a PLSR model for use in breeding programs (Armstrong, 2014). This instrument has the potential to be combined with advanced PLSR models or other regression models to create an automatic and reliable instrument for breeding in the grain industry. The second study was conducted recently by Chen et al. (2021), in which three handheld NIRS devices were used to predict protein content, and two of their PLSR models gained satisfactory results ($R_p^2 > .94$) (Chen et al., 2021). This study illustrated the feasibility of using handheld NIRS devices to determine protein content and other parameters.

Besides wavelength selection methods, nonlinear regression methods were also introduced into model development to further improve model performance. As shown in Table 4, the performance of nonlinear regression methods exceeded the average performance of linear regression methods (Williams, 2020; Chen et al., 2017). In the early 2010s, nonlinear regression methods had shown its superiority and achieved better performance than linear regression methods with the same sample size (below 100) (Mutlu et al., 2011; Mao et al., 2014). Sample size in subsequent studies was increased as much as 10-fold, resulting in models' increasing robustness (Li et al., 2017; Cui & Fearn, 2018; Bian et al., 2018; Williams, 2020). In order to reach the largest range of protein content and therefore strengthen the generalizability of regression models, the source of the samples also should be diversified (Li et al., 2017; Cui & Fearn, 2018; Bian et al., 2018; Williams, 2020; Zhang et al., 2019). For example, according to studies by Bian et al. (2018) and Zhang et al. (2019), there was a total of 882 samples from five varieties collected in 1998–2005. Both studies made modifications

on the original neural network with random weights (NNRW) and convolutional neural network (CNN), which also contributed to the performance improvement (Bian et al., 2018; Zhang et al., 2019). These modifications were more likely to be seen in regression methods based on neural networks such as improved RBFNN and improved DBN (Li et al., 2017; Mao et al., 2014). In addition, support vector machine regression (SVMR) and random forest regression (RFR) were also reliable methods to gain satisfactory results according to the studies of Williams (2020) and Chen et al. (2017). Recently, combining NIR spectra with spatial images, obtained using an emerging imaging technique named hyperspectral imaging (HSI), became popular. The method was able to provide the spectral information in each pixel and therefore can predict the distribution of protein content in the sample instead of the average values of the spectral acquisition areas (Caporaso et al., 2018a, 2018b; Manley, 2014).

Moisture

Moisture analysis has also attracted considerable research interest (Table 4). PLSR as the mainstream application had been adopted to develop models for determination of moisture content in kernels, flour, and bran (Armstrong, 2014; Hell et al., 2016; Mishra et al., 2018; Peiris et al., 2017; Shi & Yu, 2017). PLSR model development is also relatively mature, which can be seen in the studies of flour and kernels conducted by Dong and Sun (2013), Armstrong (2014), Shi and Yu (2017), Peiris et al. (2017), and Mishra et al. (2018). These studies all gained satisfactory performance. Although small sample sizes were used in studies of Dong and Sun (2013) and Shi and Yu (2017), satisfactory models were obtained by using wavelength selection methods Pearson product-moment correlation coefficient (PCC), i-PLS, and RCA. In contrast, studies of Peiris et al. (2017) and Armstrong (2014) with large sample sizes (396 samples and 300 samples) gained better results for flour ($R_p^2 = .98$) and kernels ($R_p^2 = .978$), respectively. Excellent performance was also achieved by using the models based on PLSR and the canonical correlation analysis (CCA) (Gatius et al., 2017). PLSR performance in handheld NIRS devices still needs to be improved in terms of the spectral range and modeling methods (Chen et al., 2021). As nonlinear regression methods, artificial neural network (ANN) and improved DBN were used by Mutlu et al. (2011) and Li et al. (2017), respectively, to develop models for wheat flour, and both models achieved better accuracy and stability than PLSR.

Starch

Although starch accounts for the highest content in wheat, studies of starch are significantly less than those of protein and moisture. Owens et al. (2009) conducted a

comprehensive study using undried and dried (24 h at 100°C) wheat kernels and undried milled wheat from a wide range of varieties grown in various environments and years, but the performance was not ideal. The only calibration model with acceptable performance was the one built for wheat bran starch content determination with R_p^2 of .88 (Hell et al., 2016). Recently, there were also attempts using new chemometrics methods for starch content determination in other cereals (e.g., corn), and the progress in model performance was remarkable (Jiang & Lu, 2018). The correlation coefficient method plus PLSR was used to select optimal wavelengths, and the calibration model was based on RBNFF, which presented the smallest RMSEP (0.0497) and the highest R_p^2 (.9968) (Jiang & Lu, 2018).

Amylose

As for amylose content, the performance of PLSR model was inferior to the performance in wheat kernels, flour, and wheat bran (Owens et al., 2009; Hell et al., 2016). A novel research approach was to determine the mixture level of waxy wheat (amylose content <0.2%) and conventional wheat (typical amylose content around 20%). The first quantification of this mixture was conducted by Delwiche and Graybosch (2014). Nine pairs of waxy and conventional wheat varieties were used to form 29 different proportion (conventional/waxy) percentages (0%–100%), and the optimal PLSR model produced an R_p^2 in excess of .98, regardless of sample formats or spectral pretreatments (Delwiche & Graybosch, 2014). Delwiche and Graybosch furthered their study in 2016. The flour samples were from two separate seasons in consecutive years and were provided by breeders to conduct the same experiment (Delwiche & Graybosch, 2016). Three modeling methods—PLSR, one- and two-term linear regression, and SVMR—were used to model the spectral data and reference values; however, there was no distinct difference in model performance. What is worth mentioning is that the 2290 nm wavelength with the second derivative treatment showed great potential in amylose content determination, even linear regression also achieved SEP below 10% (Delwiche & Graybosch, 2016).

Lipid

NIRS had been widely used in determination of lipid content for oilseed crops such as peanuts, soybean, sesame, and rapeseeds, but such studies in wheat were very limited (Li et al., 2020). The regression method most commonly adopted in lipid determination of oilseeds was PLSR, and the prediction results were excellent with some R^2 -values above .95 (Li et al., 2020). For wheat kernels and wheat flour, the content of free lipids and polar lipids was modeled by Dowell's group (2006). For flour, the highest R_p^2 was .74, while the results for wheat kernels were inferior and

not adequate for rough screening, which required an R_p^2 value above .7 (Dowell et al., 2006). Hell et al. (2016) used PLSR-based models with simple pretreatments to study wheat bran, and the performance in NIR spectra and mid-infrared (MIR) spectra significantly improved. In particular, the performance of lipid determination in the model for MIR spectra was improved (Hell et al., 2016). Similar to Hell et al (2016), Sujka et al. (2017) applied MIRS to build a model for crude fat and fatty acids including palmitic acid, stearic acid, arachidic acid, oleic acid, linoleic acid, and linoleic acid in flour, and only the model for crude fat had performance comparable to that reported by Hell et al. (2016). It is worth noting that the determination of arachidic acids had the highest accuracy, although its content was the lowest among these fatty acids in wheat flour (Sujka et al., 2017). Recently, researchers also experimented with wavelength selection methods and nonlinear methods for the assessment of free fatty acid value during storage of wheat flour. Variable combination population analysis (VCPA) was first introduced as a wavelength selection method for spectral processing of wheat flour, and the model based on extreme learning machine (ELM) only needed five variables selected by VCPA to achieve an R_p^2 above .95, which exhibited great potential in practical application for shelf-life evaluation (Jiang et al., 2020).

Fiber

Similar to starch, fiber content is relatively high in wheat, but its studies are relatively few. Models for fiber content in different commercial flour had been studied since 1990s (Kays et al., 1996; Horvath et al., 1984). In 2002, Kays and Barton (2002) adopted modified PLSR models to predict the contents of insoluble dietary fiber and soluble fiber, and the R_p^2 for insoluble fiber content prediction reached .98. Due to errors and deviations of the reference method, model performance for the soluble fiber content prediction was inferior (Kays & Barton, 2002). Subsequently, researchers paid more attention to using PLSR plus different wavelength selection methods including the variable importance for projection (VIP), competitive adaptive reweighted sampling (CARS), genetic algorithms (GA), and random frog (RF) (Ferreira et al., 2015; Liu et al., 2015). Similar to Kays and Barton's results, the prediction of soluble dietary fiber content in wheat was not ideal, while the result for insoluble dietary fiber was considerably better (Hell et al., 2016). In addition, the performance of NIRS was better than that of MIRS (Hell et al., 2016). Recently, HSI was also used to determine fiber content from wheat (Badaró et al., 2019). The R_p^2 and RMSEP were .98 and 0.52, respectively, with the adoption of eight selected wavelengths and PLSR (Badaró et al., 2019). Furthermore, with PLSR and spatial images from HSI, the

added fiber in semolina flour was visualized and formed a distribution map, and a reasonably accurate prediction map can be obtained in the range of 3%–75% (Badaró et al., 2019).

Polyphenols

Polyphenols are considered healthy components in whole grain products and are therefore receiving increasing attention. Overall, calibration model development for polyphenols is still at the initial stage, similar to the situation in fiber content determination, with the focus on the linear model PLSR. In addition, data accumulation for polyphenol compositions was not enough for CNN and other neural networks, which also limited the use of advanced nonlinear chemometrics (Cui & Fearn, 2018; Williams, 2020; Li et al., 2017). The first study of total polyphenol content determination in grains using NIRS was conducted in 2001 (Sato et al., 2001). In the study of Tian et al. (2021), the effects of different spectral ranges (full spectra or specific ranges from 1333 to 2222 nm) and different pretreatments on PLSR model performance were compared. The PLSR model pretreated by multiplicative scatter correction (MSC) plus first derivative (FD) had the best performance with an R_p^2 of .88, and this was also, to this day, the best performance of NIRS in polyphenol content prediction (Tian et al., 2021). Improved accuracy for polyphenol content prediction was achieved in studies of oats and green tea, with additional spectral pretreatment and selection methods. With the adoption of two-dimensional correlation spectroscopy (2DCOS) and Si-PLS for wavelength selection, performance was significantly improved (Chen et al., 2008; Zeng et al., 2019). This illustrates the importance of introducing new chemometrics methods for low concentration components because their signals are weak and more likely to be obscured by those of higher concentration (Li et al., 2020). NIRS applications in other bioactive compounds, including β -glucans, arabinoxylans, bound phenols, free phenols, and anthocyanins, are still rare (Albanell et al., 2021).

Ash

Ash content is an important flour milling quality indicator. The application of NIRS for ash determination has improved remarkably in recent years, and NIRS has been also a standard method for ash determination since 2000 (Cereals & Grains Association, 2020). Although ash does not have NIR energy absorption, most minerals are contained in the organic materials, which makes it possible to determine the ash content (Garnsworthy et al., 2000). Dowell et al. (2006) used four NIR spectrometers to acquire spectra and calibrated the model for wheat grain ash content and flour ash content, respectively, based on PLSR plus mean centering and FD. The R_p^2 of the two calibra-

tion models was below .45 and .32, respectively, which was insufficient for quantitative determination (Dowell et al., 2006). Another study conducted by Viljoen et al. (2005) had better performance by using a larger and processed dataset with 20 cultivars from 3 years as well as outliers removal (Viljoen et al., 2005). The comparison between the two studies showed that larger and processed datasets have a positive effect on model development. Subsequently, i-PLS was introduced into ash content determination (Dong & Sun, 2013). Although Dong and Sun's (2013) sample size was half of that in Viljoen et al. (2005), the performance of the PLSR model was the best and was the same as that acquired in Hell et al. (2016). The ash content in wheat bran was first studied in 2016, and the study compared the performance of MIRS and NIRS (Hell et al., 2016). It showed that the performance of MIRS was inferior to that of NIRS (Hell et al., 2016). Recently, a PLSR model for flour ash content determination was developed using MIRS, and the R_p^2 of the model reached .94, indicating its feasibility in determining flour ash content (Sujka et al., 2017).

Compared to linear regression methods, nonlinear regression methods achieved better performance in ash content determination. Back propagation neural network (BPNN), DBN, improved DBN, and CNN were evaluated by Li et al. (2017) and Cui and Fearn (2018). The model with the best performance in the study of Li et al. (2017) was developed by using wavelet transformation as a pretreatment method and then an improved DBN method instead of the original DBN method. Compared with these models, the performance achieved in the study of Chen et al. (2021) with handheld NIR devices was less satisfactory, but further exploration of nonlinear regression methods in handheld NIR devices will be worthwhile.

3.2.2 | Determination of physical parameters

Hardness

The hardness of wheat kernel samples is usually measured by the Single Kernel Characterization System (SKCS). NIRS can be used to measure hardness because the scatter characteristics of a sample shown in spectra are influenced by the particle size distribution of the powdered sample, which is in turn connected with wheat kernel texture or hardness (Norris et al., 1989). NIRS is recognized as a standard method for wheat kernel hardness determination (Cereals & Grains Association, 1999c). The high correlation between hardness and damaged starch in flour verified this conclusion (Finney et al., 1988). The first study to measure wheat kernel hardness using NIRS was conducted by Manley et al. in 2002, and the correlation of coefficient between the predicted results using PLSR and the reference values (particle size index [PSI] test) was

only .42 (Manley et al., 2002). Similar performance, regardless of wheat varieties, sample formats, and NIR instruments, was seen in the study of Dowell et al. (2006). Subsequent studies adopted pretreatment methods for PLSR model development and larger datasets, and model performance was improved (Owens et al., 2009; Mishra et al., 2018). A comprehensive study conducted by Williams (2020) compared nine prediction models built by participants with professional modeling backgrounds based on samples from 8 years. The nonlinear regression method, RFR, gave the best result ($R_p^2 = .826$) without any pretreatment, outlier removal, or wavelength selection. The calibration model was built on seven seasons' data and validated with one additional season. It exhibited great robustness, although the R_p^2 value was below that in the study of Mishra et al. (2018).

Newer chemometrics methods were used in the spectral dataset collected by HSI. Both Mahesh et al. (2014) and Erkinbaev et al. (2019) selected wheat varieties from Canada and used the same hardness test methods. The nonlinear regression method, ANN, was superior to PLSR in the study of Erkinbaev et al., and PLSR showed better performance than PCR in the study of Mahesh et al. (Mahesh et al., 2014; Erkinbaev et al., 2019). Ant colony optimization (ACO) algorithm combining SVMR gained the best performance in hardness prediction (Zhang et al., 2017). Overall, the accuracy of determination was mostly improved by the introduction of chemometrics rather than HSI, and HSI spectra relied more on advanced chemometrics methods than conventional NIR spectrometers to gain similar performance (Zhang et al., 2017; Mahesh et al., 2014).

Test weight and 1000 kernel weight

Test weight and 1000 kernel weight are both indicators of wheat quality. In the comprehensive study of Dowell et al. (2006), spectra from whole wheat kernels and flour of two wheat cultivars were used for PLSR model development to predict test weight (Dowell et al., 2006). The study showed that performance for the whole kernel was superior to that for flour, and the NIR spectrometer with a wider wavelength range was beneficial (Dowell et al., 2006). In the most recent study conducted by Williams (2020), the best R_p^2 was .756, which was achieved by a model based on RFR, and the experimental samples were collected from 7 years. In addition, the nonlinear regression model showed superiority in hardness prediction compared to the PLSR model ($R_p^2 = .414$) (Williams, 2020).

Armstrong (2014) designed an automatic NIR instrument for single-seed measurement of wheat kernels, and the developed PLSR model achieved a high R_p^2 (.86). The study concluded that there was no correlation between protein content and kernel mass (Armstrong, 2014). The single-seed measurement instrument can evaluate the sin-

gle seed at a rate of 4 seeds/s at least and achieved sufficient detection for breeding purposes (Armstrong, 2014). As for the 1000 kernel weight, Mishra et al. (2018) attempted to build a model to determine the relationship between NIR spectra and 1000 kernel weight measured by an automatic seed-counter machine (Mishra et al., 2018). From 2000 to 2018, R_p^2 improved from .66 to .907, which showed the potential to integrate 1000 kernel weight measurement into the NIRS assessment system (Garnsworthy et al., 2000; Mishra et al., 2018).

Germination/falling number

The falling number (the Hagberg number) represents the α -amylase activity in kernels, which increases sharply after germination and influences dough and end product quality (Caporaso et al., 2018b; Barbedo et al., 2018). Using NIRS to determine the falling number started in 2001, and it was subsequently studied in 2006. The performances of the models were all unacceptable regardless of the sample format and NIR equipment used (Hruskova et al., 2001; Dowell et al., 2006). The next effort was made in 2018. Delwiche et al. (2018) collected soft wheat samples in two formats: kernels and ground meals. Based on the PLSR models, the kernel format had better performance ($R_C^2 = .708$), which was still poor in accuracy. When the two formats were combined into a larger dataset, the model performance became worse (Delwiche et al., 2018). The only acceptable result was achieved by Sujka et al. (2017) using MIRS, where the R_p^2 was .95 as predicted by the PLSR model based on specific spectral ranges. In addition, the dataset in this study consisted of four different grains—wheat, rye, spelt, and triticale (Sujka et al., 2017). The performance of HSI-based models in falling number was not satisfactory either. With the adoption of various pretreatment methods, the R_p^2 s were below .5 in the study of Caporaso et al. (2017).

Gluten content and gluten index

Fewer studies have been performed on gluten content than protein content, and these studies can be divided into wet gluten determination and dry gluten determination. According to the studies of Başlar and Ertugay (2011) and Sinelli et al. (2011), models for dry gluten achieved better prediction performance (Başlar & Ertugay, 2011; Sinelli et al., 2011). The study of Gatius et al. (2017) was the only one using a new linear regression method (CCA) to predict dry gluten, and the model had better performance than PLSR.

The improvement brought by Si-PLS and SVMR on wet gluten content prediction models was shown in the study of Chen et al. (2017). The R_p^2 of the model based on SVMR plus Si-PLS was .85, while the single PLSR model R_p^2 was .675 (Chen et al., 2017). In a most recent study, Chen et al. (2021) achieved similar performance using the

PLSR model for wet gluten prediction without the help of wavelength selection methods, but the sample size was double that of the previous study (Chen et al., 2021). The spectra in the study were collected from handheld NIRS devices, which would aid practical application (Chen et al., 2021).

Gluten index is an indicator of gluten quality. In the study of Dowell et al. (2006) of gluten index, the values of R_p^2 were below .5 for both wheat kernels and wheat flour. Five years later, a double size dataset was used in the study of Sinelli et al. (2011), and the new model had better performance ($r_p = .82$) with extended MSC and FD (Sinelli et al., 2011).

Zeleny sedimentation value

Zeleny sedimentation value refers to the sedimentation volume of flour suspended in lactic acid solution, which is associated with the quality and quantity of gluten in flour (Cauvain & Young, 2009). Remarkable advances have been made in Zeleny sedimentation value prediction, with correlation coefficients increasing from .11–.50 in 2001 to .924 in 2011 (Hruskova et al., 2001; Hruskova & Famera, 2003; Dowell et al., 2006; Jirsa et al., 2008; Başlar & Ertugay, 2011). The initial studies were based on PLSR for model development, and performance improvement to a large extent depended on increasing the quality and size of datasets. In 2011, the introduction of ANNs broke the bottleneck and achieved the most satisfactory performance without using any pretreatment method (Mutlu et al., 2011). Chen et al. (2021) compared the performances of three handheld NIRS spectrometers on Zeleny sedimentation value. With a larger dataset, pretreatment, and outlier removal, the performance of the best model in the study was similar to that in the study of Mutlu et al. (2011).

Damaged starch

Native starch exists as discrete granules in kernels or flour. The granules may be damaged during milling, thereby forming damaged starch that will increase water absorption and thus have an effect on dough rheology and end use product quality. The first determination of starch damage content was conducted in 1981, and the performance was acceptable ($R_p^2 = .90$; $SEP = 4.2$); the calibration model was subsequently optimized (Dowell et al., 2006; Pojić & Mastilović, 2013). The application of discrete wavelengths and full wavelength range for model development were respectively studied by Finney et al. in 1988 and Morgan and Williams in 1995, and both approaches improved performance (Finney et al., 1988; Morgan & Williams, 1995). A breakthrough in spectral data acquisition mode and sample type was made by Miralbés in 2004. The spectra were collected using the NIR transmittance mode, and the sam-

ple was commercial wheat flour (a mix of different cultivars). A modified PLSR model with standard normal variate (SNV), detrending, and the FD achieved record high performance ($R_p^2 = .94$; $SEP = 1.63$) (Miralbés, 2004).

3.3 | Quantitative determination of rheological parameters

Rheological parameters are concerned with the properties of dough made from wheat flour. Rheological parameters can be studied using many methods, and NIRS application mainly focuses on four aspects: farinograph characteristics, mixograph characteristics, extensograph characteristics, and alveograph characteristics (Table 5).

Farinograph characteristics refer to dough's resistance against the mixing action of two paddles (blades). Relevant measurements include farinograph water absorption (FWA), farinograph dough development time (FDDT), farinograph dough stability (FDS), and farinograph mixing tolerance index (FMTI) (Cauvain & Young, 2009). Mixograph characteristics are also an indicator of dough resistance but against more intensive mixing action of pins. Relevant measurements include mixograph water absorption (MWA), mixograph development time (MDT), and mixograph tolerance (MT) (Cauvain & Young, 2009). Extensograph characteristics refer to the stretch behaviors of dough, and relevant measurements include resistance to extension (R) and extensibility (E) (Cauvain & Young, 2009). The alveograph measures the gas retention capacity when air is injected into dough to form bubbles. Relevant measurements include dough to deformation/tenacity (P), deformation energy (W), and configuration ratio (P/L) (Cauvain & Young, 2009).

NIRS was first applied to determine rheological parameters, including water absorption and mixing time, in the study of Rubenthaler and Pomeranz (1987). Subsequent studies extended the diversity of models and NIR spectrometers and introduced more and newer chemometrics methods (Hruskova et al., 2001; Miralbés, 2003, 2004; Dowell et al., 2006; Jirsa et al., 2008; Mutlu et al., 2011; Sinelli et al., 2011; Arazuri et al., 2012; Gatus et al., 2017; Mutlu et al., 2011; Williams, 2020). Overall, the performance of NIRS models to assess rheological parameters was not ideal. A comprehensive summary of rheological parameter studies from the past 10 years is shown in Table 5. The challenge in rheological parameter prediction is how to achieve generalizability of the prediction model, which is influenced by sample formats and composition proportions, the quality of the dataset for calibration model development, and system errors in reference values and chemometrics methods.

TABLE 5 An overview of NIR application in quantitative determination of wheat rheological parameters

| Chemometrics | | | | | | | | | | |
|--------------|------------------|--------------------------------------|--------------------|----------------|--------------------------------------|---|----------------|-------|---|----------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | preprocess | Model | Performance | References |
| FWA | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | RFR | $R_p^2 = .914$; SEP = 0.99; RPD = 4.83 | Williams, 2020 |
| FWA | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | SVMR | $R_p^2 = .910$; SEP = 1.41; RPD = 3.39 | Williams, 2020 |
| FWA | Reflectance | FOSS NIR System 6500 Spectrometer | 720–1700 | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | SGA + FD + MSC | PLSR | $R_p^2 = .914$; SEP = 1.4; RPD = 3.41 | Williams, 2020 |
| DDT | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | RFR | $R_p^2 = .581$; SEP = 0.94; RPD = 1.85 | Williams, 2020 |
| DDT | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | SVMR | $R_p^2 = .52$; SEP = 1.02; RPD = 1.7 | Williams, 2020 |
| DDT | Reflectance | FOSS NIR System 6500 Spectrometer | 720–1700 | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | SGA + MSC | PLSR | $R_p^2 = .498$; SEP = 1.22; RPD = 1.42 | Williams, 2020 |

(Continues)

TABLE 5 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|-----------------------------------|--------------------|----------------|--------------------------------------|---|------------|-------|---|----------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | preprocess | Model | Performance | References |
| FMTI | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | RFR | $R_p^2 = .972$; SEP = 9.6; RPD = 3.72 | Williams, 2020 |
| FMTI | Reflectance | FOSS NIR System 6500 Spectrometer | Full range | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | SVMR | $R_p^2 = .890$; SEP = 18.1; RPD = 1.96 | Williams, 2020 |
| FMTI | Reflectance | FOSS NIR System 6500 Spectrometer | 1108–2492 | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | None | BPANN | $R_p^2 = .963$; SEP = 15.3; RPD = 2.31 | Williams, 2020 |
| FMTI | Reflectance | FOSS NIR System 6500 Spectrometer | 720–1700 | Kernel | 862 (755/107) | Five varieties collected in seven different regions from 1998 to 2005 | SGA | PLSR | $R_p^2 = .890$; SEP = 16.3; RPD = 2.18 | Williams, 2020 |
| L | Reflectance | Foss InfraX-act™ Lab/Pro | 570–1850 | Kernel | 411 (300/111) | Collected from 50 localities from 2007 to 2009 | FD + OR | PLSR | $R_p^2 = .66$; SEP = 8.05; $R_{CV}^2 = .81$; SECV = 5.89 | Arazuri et al., 2012 |
| P | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 722 (644/100) | Collected from a commercial flour company | None | PLSR | RMSEP = 4.63; RMSECV = 5.26 | Gatius et al., 2017 |
| P | Reflectance | FOSS NIR System 6500 Spectrometer | 1100–2500 | Flour | 722 (644/100) | Collected from a commercial flour company | None | CCA | RMSEP = 3.91; RMSECV = 4.94 | Gatius et al., 2017 |

(Continues)

TABLE 5 (Continued)

| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | Chemometrics | | | References |
|-------------|------------------|-----------------------------------|--------------------|----------------|--------------------------------------|--|---------------|-------|---|----------------------|
| | | | | | | | preprocess | Model | Performance | |
| P | Reflectance | Foss InfraX-act™ Lab/Pro | 570–1850 | Kernel | 411 (300/111) | Collected from 50 localities from 2007 to 2009 | SD + OR | PLSR | $R_p^2 = .27$; SEP = 4.87; $R_{CV}^2 = .77$; SECV = 4.80 | Arazuri et al., 2012 |
| P | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79; (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .948$ | Mutlu et al., 2011 |
| W | Reflectance | Bruker FT-NIR spectrometer | 830–2700 | Kernel | 547 (289/157) | 30 varieties collected from eight different agro-climate areas | MSC + SD + OR | PLSR | $r_p = .92$; RMSEP = 17.60; $r_{CV} = .91$; RMSECV = 18.9; RPD = 2.3 | Sinelli et al., 2011 |
| W | Reflectance | Foss InfraX-act™ Lab/Pro | 570–1850 | Kernel | 411 (300/111) | Collected from 50 localities from 2007 to 2009 | OR | PLSR | $R_p^2 = .60$; SEP = 13.07; $R_{CV}^2 = .86$; SECV = 11.53 | Arazuri et al., 2012 |
| P/L | Reflectance | Foss InfraX-act™ Lab/Pro | 570–1850 | Kernel | 411 (300/111) | Collected from 50 localities from 2007 to 2009 | OR | PLSR | $R_p^2 = .61$; SEP = 0.049; $R_{CV}^2 = .86$; SECV = 0.027 | Arazuri et al., 2012 |
| P/L | Reflectance | Bruker FT-NIR spectrometer | 830–2700 | Kernel | 547 (289/157) | 30 varieties collected from eight different agro-climate areas | MSC + OR | PLSR | $r_p = .88$; RMSEP = 0.19; RPD = 2.0; $r_{CV} = .90$; RMSECV = 0.23; RPD = 2.2 | Sinelli et al., 2011 |
| P/G | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .933$ | Mutlu et al., 2011 |

(Continues)

TABLE 5 (Continued)

| | | Chemometrics | | | | | | | | |
|-------------|------------------|-----------------------------------|--------------------|----------------|--------------------------------------|--------------------------------------|------------|-------|----------------|--------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Additional sample information | preprocess | Model | Performance | References |
| FWA | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .832$ | Mutlu et al., 2011 |
| DDT | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .174$ | Mutlu et al., 2011 |
| DS | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .034$ | Mutlu et al., 2011 |
| STA | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .009$ | Mutlu et al., 2011 |
| W | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .049$ | Mutlu et al., 2011 |
| L | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .465$ | Mutlu et al., 2011 |
| Ie | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .000$ | Mutlu et al., 2011 |

3.3.1 | Linear regression methods

Early applications of NIRS to determine rheological parameters used linear regression methods. Hruskova et al. (2001) conducted a comprehensive study of farinograph and extensograph characteristics based on 39 wheat flour samples from 1998 and 75 samples from 1999. Only FWA values were acceptable in two models for samples from 1998 and 1999 and their joint dataset, while FMTI, FDS, and E performance were unsatisfactory with correlation coefficients below .5 (Hruskova et al., 2001). Farinograph characteristics' prediction was improved somewhat in 2006 but still not enough for rough screening (Dowell et al., 2006). Mixograph characteristics were first modeled in 2006, and the best performance was shown in the model for MWA ($R_p^2 = .79$), while the prediction for the other parameters was not satisfactory (Dowell et al., 2006).

As for alveograph characteristics, performance was generally poor. The NIR transmittance mode was employed in Miralbés's two studies (Miralbés, 2003, 2004), which contained 236 and 358 samples from different varieties and regions, respectively. In 2003, only W gained satisfactory results ($R_p^2 = .84$). When the dataset was divided into two using W values, and respective models were built for the two subsets, the R_p^2 s for P, P/L, and W all greatly improved. In Miralbés's study conducted in 2004, R_p^2 s for P and W were above .9, while the prediction in P/L was still low with $R_p^2 = .79$ (Miralbés, 2004). A breakthrough for P/L prediction ($r_p = .88$) was shown in the study of Sinelli et al. (2011) where the spectra were from 547 durum wheat samples and the NIR reflectance mode was used. W prediction also reached a high level ($r_p = .92$) in the PLSR model (Sinelli et al., 2011). Another linear regression method, canonical correlation analysis (CCA), was used in P prediction, and the result in the CCA-based model outperformed that in the PLSR model (Gatius et al., 2017).

Low generalizability was seen in prediction of alveograph characteristics. Generalizability of the calibration model for W and P in the prediction dataset (spectra from commercial flour) was unsatisfactory, although it was acceptable in the validation dataset (Jirsa et al., 2008). The study of Arazuri et al. (2012) yielded similar results. The performance of the developed model deteriorated when an independent dataset was used to test the model for P, P/L, W, and extensibility (L) (Arazuri et al., 2012).

3.3.2 | Nonlinear regression methods

Regression methods were first used in 2011 by Mutlu et al. who adopted ANNs to predict farinograph and alveograph characteristics (Mutlu et al., 2011). Based on reflectance spectra from 79 samples, P, FWA, and the

ratio of P/G (dough swelling) had satisfactory accuracy, while the performance in FDDT, FDS, alveograph elasticity index (I_e), W, and L was unsatisfactory ($R_p^2 < .4$) (Mutlu et al., 2011). Most recently, Williams (2020) conducted a comprehensive study using three nonlinear regression methods and six PLSR methods to compare their performances in terms of three farinograph characteristics FWA, FDDT, and FMTI (Williams, 2020). Each parameter was individually modeled by nine different methods. The model based on RFR had the best accuracy in three parameters. Although R_p^2 in FDDT was still unsatisfactory, the improvement was remarkable compared to previous results (Williams, 2020). As for FWA and FMTI, R_p^2 reached .914 and .972, respectively, both of which show their strong robustness and generalizability (Williams, 2020).

3.4 | Quantitative determination of end product quality

3.4.1 | Bread

Wheat-based products include bread, cookies, pasta, noodle, cake, and others. As the major end product, bread has been mostly studied (Sinelli et al., 2011; Abu-Ghoush, Herald, Dowell, Xie, Aramouni, & Madl, 2008; Abu-Ghoush, Herald, Dowell, Xie, Aramouni, & Walker, 2008). Quality assessment of bread includes loaf tests (e.g., loaf volume, weight, specific volume, and height), texture analyses (e.g., crumb pore size, crumb grain score, crumb hardness, and crust brittleness), and other analyses (e.g., moisture content) (Ahmad et al., 2016). Because of its non-destructive feature, NIRS has gained increasing interest for bread quality assessment (Abu-Ghoush, Herald, Dowell, Xie, Aramouni, & Madl, 2008; Abu-Ghoush, Herald, Dowell, Xie, Aramouni, & Walker, 2008; Amigo et al., 2019; Nallan Chakravartula et al., 2019). A comprehensive summary of bread quality assessment studies in the past 10 years is shown in Table 6.

Baking test parameters

Among different quality prediction parameters, baking test parameters, in particular loaf volume, had gained the most attention. The first study of loaf volume prediction and crumb grain score prediction was conducted by Dowell et al. in 2006 and achieved acceptable results on loaf volume, while the results on crumb grain score were not satisfactory (Dowell et al., 2006). Spectra from hard red spring/winter wheat kernel and flour were used to build calibration models for loaf volume prediction. The results showed that the flour from hard red winter wheat had the best performance (2006). In the study of Jirsa et al. (2008), flour samples were from different manufacturers

TABLE 6 An overview of NIR application in quantitative determination of end products

| Chemometrics | | | | | | | | | | |
|------------------|------------------|-----------------------------------|--------------------|---------------------|--------------------------------------|--|------------|-------|--|----------------------|
| Application mode | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Other sample information | Preprocess | Model | Performance | References |
| Loaf volume | Reflectance | Polytec PSS 1720 spectrometers | 850–1650 | Kernel, no cleaning | 591 (443/148) | Collected from different field trials in 2 years | None | CPLSR | $R_p^2 = .65$; RMSEP = 53.9; RPD = 1.69; RMSECV = 48.8 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Polytec PSS 2120 spectrometers | 1100–2100 | Kernel, no cleaning | 591 (443/148) | Collected from different field trials in 2 years | None | PLSR | $R_p^2 = .66$; RMSEP = 52.6; RPD = 1.71; RMSECV = 54.3 | Gabriel et al., 2017 |
| Loaf volume | Transmittance | FOSS Infracore 1241 spectrometers | 850–1050 | Kernel, no cleaning | 591 (443/148) | Collected from different field trials in 2 years | None | CPLSR | $R_p^2 = .69$; RMSEP = 50.8; RPD = 1.8; RMSECV = 51.4 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Polytec PSS 1720 spectrometers | 850–1650 | Kernel | 591 (443/148) | Collected from different field trials in 2 years | None | PLSR | $R_p^2 = .70$; RMSEP = 49.7; RPD = 1.83; RMSECV = 49.2 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Polytec PSS 2120 spectrometers | 1100–2100 | Kernel | 591 (443/148) | Collected from different field trials in 2 years | None | CPLSR | $R_p^2 = .69$; RMSEP = 51.8; RPD = 1.80; RMSECV = 49.3 | Gabriel et al., 2017 |
| Loaf volume | Transmittance | FOSS Infracore 1241 spectrometers | 850–1050 | Kernel | 591 (443/148) | Collected from different field trials in 2 years | None | PLSR | $R_p^2 = .66$; RMSEP = 52.5; RPD = 1.71; RMSECV = 50.3 | Gabriel et al., 2017 |

(Continues)

TABLE 6 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|----------------------------------|--------------------|--------------------|--------------------------------------|--|------------|-------|--|----------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Other sample information | Preprocess | Model | Performance | References |
| Loaf volume | Transmittance | FOSS Infratec 1241 spectrometers | 850–1050 | Kernel | 591 (443/148) | Collected from different field trials in 2 years | None | CPLSR | $R_p^2 = .67$ RMSEP = 51.6; RPD = 1.74; RMSECV = 50.8 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Polytec PSS 1720 spectrometers | 850–1650 | Coarse meal | 591 (443/148) | Collected from different field trials in 2 years | None | PLSR | $R_p^2 = .68$; RMSEP = 50.3; RPD = 1.68; RMSECV = 49.6 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Polytec PSS 2120 spectrometers | 1100–2100 | Coarse meal | 591 (443/148) | Collected from different field trials in 2 years | None | PLSR | $R_p^2 = .70$; RMSEP = 49.3; RPD = 1.73; RMSECV = 50.2 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | FOSS XDS spectrometers | 400–2500 | Coarse meal | 591 (443/148) | Collected from different field trials in 2 years | None | CPLSR | $R_p^2 = .69$; RMSEP = 50.5; RPD = 1.83; RMSECV = 45.3 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Polytec PSS 1720 spectrometers | 850–1650 | Standardized flour | 591 (443/148) | Collected from different field trials in 2 years | None | CPLSR | $R_p^2 = .72$; RMSEP = 48.2; RPD = 1.89; RMSECV = 48.2 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Polytec PSS 2120 spectrometers | 1100–2100 | Standardized flour | 591 (443/148) | Collected from different field trials in 2 years | None | PLSR | $R_p^2 = .75$; RMSEP = 45.6; RPD = 2; RMSECV = 47.5 | Gabriel et al., 2017 |

(Continues)

TABLE 6 (Continued)

| | | Chemometrics | | | | | | | | |
|-------------|------------------|-----------------------------------|--------------------|--------------------|--------------------------------------|--|------------|-------|--|-------------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Other sample information | Preprocess | Model | Performance | References |
| Loaf volume | Reflectance | FOSS XDS spectrometers | 400–2500 | Standardized flour | 591 (443/148) | Collected from different field trials in 2 years | None | CPLSR | $R_p^2 = .74$; RMSEP = 46.2; RPD = 1.96; RMSECV = 42.8 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Polytec PSS I720 spectrometers | 850–1650 | Dough | 591 (443/148) | Collected from different field trials in 2 years | None | PLSR | $R_p^2 = .81$; RMSEP = 39.4; RPD = 2.29; RMSECV = 46.0 | Gabriel et al., 2017 |
| Loaf volume | Reflectance | Hamamatsu Photonics NIR II CCD | 1380–2250 | Flour | 10 (10/0) | Four varieties harvested in 2009 | SGS + SNV | nPLSR | $R_{CV}^2 = .29$; RMSECV = 17; RMSEC = 8 | Li Vigni & Cocchi, 2013 |
| Loaf volume | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .687$ | Mutlu et al., 2011 |
| Loaf weight | Reflectance | Hamamatsu Photonics NIR II CCD | 1380–2250 | Flour | 10 (10/0) | Four varieties harvested in 2009 | SGS + SNV | nPLSR | $R_{CV}^2 = .02$; RMSECV = 0.5; RMSEC = 0.3 | Li Vigni & Cocchi, 2013 |
| Loaf weight | Reflectance | FOSS NIR System 6500 Spectrometer | 400–2500 | Flour | 79 (58/21) | Collected in different areas in 2006 | None | BPNN | $R_p^2 = .714$ | Mutlu et al., 2011 |
| Loaf height | Reflectance | Hamamatsu Photonics NIR II CCD | 1380–2250 | Flour | 10 (10/0) | Four varieties harvested in 2009 | SGS + SNV | nPLSR | $R_{CV}^2 = .31$; RMSECV = 0.2; RMSEC = 0.1 | Li Vigni & Cocchi, 2013 |

(Continues)

TABLE 6 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|---|--------------------|----------------|--------------------------------------|----------------------------------|----------------|-------|---|-------------------------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Other sample information | Preprocess | Model | Performance | References |
| Loaf density | Reflectance | Hamamatsu Photonics NIR II CCD | 1380–2250 | Flour | 10 (10/0) | Four varieties harvested in 2009 | SGS + SNV | nPLSR | $R_{CV}^2 = .33$; RMSECV = 0.01; RMSEC = 0.05 | Li Vigni & Cocchi, 2013 |
| Chewiness | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Bread | 90 (63/27) | A single layer of coating | SNV + MSC + | PLSR | $R_p^2 = .879$; RMSEP = 3.4; $R_{CV}^2 = .901$; RMSECV = 2.9 | Nallan Chakravar- tula et al., 2019 |
| Chewiness | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Bread | 90 (63/27) | Two layers of coating | SNV + MSC + FD | PLSR | $R_p^2 = .892$; RMSEP = 3.7; $R_{CV}^2 = .915$; RMSECV = 3.7 | Nallan Chakravar- tula et al., 2019 |
| HD | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Bread | 90 (63/27) | A single layer of coating | SNV + MSC + | PLSR | $R_p^2 = .878$; RMSEP = 8.8; $R_{CV}^2 = .886$; RMSECV = 8.1 | Nallan Chakravar- tula et al., 2019 |
| HD | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Bread | 90 (63/27) | Two layers of coating | SNV + MSC + FD | PLSR | $R_p^2 = .915$; RMSEP = 10.7; $R_{CV}^2 = .924$; RMSECV = 10.6 | Nallan Chakravar- tula et al., 2019 |
| HD | Reflectance | Headwall photonics model 1002A-00371 Spectrometer | 938–1630 | Bread | 108 (99/9) | Six different storage times | MSC + SNV + | PLSR | $R_{CV}^2 = .86$; RMSECV = 85.2 | Amigo et al., 2019 |

(Continues)

TABLE 6 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|--|--------------------|-----------------|--------------------------------------|---------------------------|----------------|-------|--|--|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Other sample information | Preprocess | Model | Performance | References |
| Moisture | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Top of bread | 90 (63/27) | A single layer of coating | SNV + MSC + FD | PLSR | $R_p^2 = .938$; RMSEP = 0.90; $R_{CV}^2 = .948$; RMSECV = 0.90 | Nallan Chakravar- tula et al., 2019 |
| Moisture | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Bottom of bread | 90 (63/27) | A single layer of coating | SNV + MSC + | PLSR | $R_p^2 = .921$; RMSEP = 1.88; $R_{CV}^2 = .955$; RMSECV = 0.77 | Nallan Chakravar- tula et al., 2019 |
| Moisture | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Centre of bread | 90 (63/27) | A single layer of coating | SNV + MSC + FD | PLSR | $R_p^2 = .946$; RMSEP = 1.52; $R_{CV}^2 = .959$; RMSECV = 1.40 | Nallan Chakravar- tula et al., 2019 |
| Moisture | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Top of bread | 90 (63/27) | Two layers of coating | SNV + MSC + | PLSR | $R_p^2 = .8451$; RMSEP = 1.36; $R_{CV}^2 = .852$; RMSECV = 1.31 | Nallan Chakravar- tula et al., 2019 |
| Moisture | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Bottom of bread | 90 (63/27) | Two layers of coating | SNV + MSC + FD | PLSR | $R_p^2 = .862$; RMSEP = 1.45; $R_{CV}^2 = .899$; RMSECV = 1.35 | Nallan Chakravar- tula et al., 2019 |
| Moisture | Reflectance | Bruker FT-NIR spec- tropho- tometer (MATRIX™ -F) | 800–2500 | Centre of bread | 90 (63/27) | Two layers of coating | SNV + MSC + | PLSR | $R_p^2 = .921$; RMSEP = 1.88; $R_{CV}^2 = .945$; RMSECV = 1.72 | Nallan Chakravar- tula et al., 2019 |

(Continues)

TABLE 6 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|---|---------------------------------|-------------------------|--------------------------------------|---|------------|-------|---|----------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Other sample information | Preprocess | Model | Performance | References |
| Storage time | Reflectance | Bruker Spectrometer Vector 22/N | 1200–2500 | Bread | 84 (72/12) | Seven different storage times | MSC + FD | PLSR | $R_p^2 = .969$; RMSEP = 1.87; $R_{CV}^2 = .972$; RMSECV = 1.66 | Cevoli et al., 2015 |
| Protein | Reflectance | Bruker MPA™. Multipurpose FT-NIR analyzer | 1100–1245; 1590–2600 | Various bakery products | 102 (102/0) | Ground, homogenous, and sieved powder samples | FD | PLSR | $R_{CV}^2 = .9893$; RMSECV = 0.16 | Szigedi et al., 2011 |
| Lipid | Reflectance | Bruker MPA™. Multipurpose FT-NIR analyzer | 1330–1840; 2170–2350 | Various bakery products | 102 (102/0) | Ground, homogenous, and sieved powder samples | FD + NL | PLSR | $R_{CV}^2 = .9907$; RMSECV = 0.79 | Szigedi et al., 2011 |
| Sugar | Reflectance | Bruker MPA™. Multipurpose FT-NIR analyzer | 1400–1630; 2000–2170; 2230–2570 | Various bakery products | 102 (102/0) | Ground, homogenous, and sieved powder samples | MSC | PLSR | $R_{CV}^2 = .9981$; RMSECV = 0.28 | Szigedi et al., 2011 |
| Energy | Reflectance | Viavi MicroNIR 1700 handheld spectrometer | 908–1676 | Pasta/sauce blend | 125 (100/25) | Five different proportion mixtures (0%–100%) | SGA + EMSC | PLSR | $R_p^2 = .86$; RPD = 2.02 | Neves et al., 2019 |
| Carbohydr | Reflectance | Viavi MicroNIR 1700 handheld spectrometer | 908–1676 | Pasta/sauc blend | 125 (100/25) | Five different proportion mixtures (0%–100%) | SGA + EMSC | PLSR | $R_p^2 = .85$; RPD = 2.54 | Neves et al., 2019 |

(Continues)

TABLE 6 (Continued)

| Chemometrics | | | | | | | | | | |
|--------------|------------------|---|--------------------|-------------------|--------------------------------------|--------------------------------------|------------|-------|---------------------------------|-----------------------|
| Application | Acquisition mode | NIR instrument | Spectral range, nm | Type of sample | Sample size (calibration/prediction) | Other sample information | Preprocess | Model | Performance | References |
| Fat | Reflectance | Viavi MicroNIR 1700 handheld spectrometer | 908–1676 | Pasta/sauce blend | 125 (100/25) | Five different proportions (0%–100%) | SGA + EMSC | PLSR | $R_p^2 = .89$; RPD = 2.77 | Neves et al., 2019 |
| Fiber | Reflectance | Viavi MicroNIR 1700 handheld spectrometer | 908–1676 | Pasta/sauce blend | 125 (100/25) | Five different proportions (0%–100%) | SGA + EMSC | PLSR | $R_p^2 = .90$; RPD = 2.45 | Neves et al., 2019 |
| Protein | Reflectance | Viavi MicroNIR 1700 handheld spectrometer | 908–1676 | Pasta/sauce blend | 125 (100/25) | Five different proportions (0%–100%) | SGA + EMSC | PLSR | $R_p^2 = .86$; RPD = 2.26 | Neves et al., 2019 |
| Sugar | Reflectance | Viavi MicroNIR 1700 handheld spectrometer | 908–1676 | Pasta/sauce blend | 125 (100/25) | Five different proportions (0%–100%) | SGA + EMSC | PLSR | $R_p^2 = .88$; RPD = 2.19 | Neves et al., 2019 |
| Storage time | Reflectance | Buchi NIRLAB N-200 FT-NIR | 1000–2500 | Pasta | 120 (58/62) | Storage at 0°C | Unknown | PLSR | $r_{cv} = .984$; RMSECV = 2.12 | Zardetto et al., 2021 |
| Storage time | Reflectance | Buchi NIRLAB N-200 FT-NIR | 1000–2500 | Pasta | 102 (54/48) | Storage at 5°C | Unknown | PLSR | $r_{cv} = .987$; RMSECV = 2.02 | Zardetto et al., 2021 |
| Storage time | Reflectance | Buchi NIRLAB N-200 FT-NIR | 1000–2500 | Pasta | 108 (55/53) | Storage at 10°C | Unknown | PLSR | $r_{cv} = .984$; RMSECV = 2.09 | Zardetto et al., 2021 |

and dataset division was inhomogeneous, R_p^2 of loaf volume decreased sharply to .096, and generalizability was poor (Jirsa et al., 2008). It was difficult to build a robust model based on a small calibration dataset, which was shown in loaf volume, weight, height, and density determination (Li Vigni & Cocchi, 2013).

A comprehensive study conducted by Gabriel et al. (2017) on prediction of baking quality had the most practical value. The study tested three NIR spectrometers plus PLSR and canonical PLSR (CPLSR) in five different levels of processed wheat ranging from no cleaning wheat kernels to dough (Gabriel et al., 2017). The study showed that NIR spectrometers affected model performance due to their collection spectral range, while CPLSR, a new method combining CCA and PLSR, was not significantly better than PLSR. The extent of processing played the most important role in model performance, which was consistent with the findings in Dowell's study (2006), and the best R_p^2 was .81, which was achieved by the PLSR model in the 148 independent spectra of dough (Gabriel et al., 2017). In addition, protein content can only explain 59% variability of loaf volume, and that number decreased to 15% when protein content was above 12%, which was inconsistent with previous reports (Jirsa et al., 2008; Gabriel et al., 2017).

The only attempt using nonlinear regression models for loaf volume and loaf weight prediction was based on 58 spectral samples from wheat flour (Mutlu et al., 2011). The performances in both parameters were inferior to that in the study of Gabriel et al. (Mutlu et al., 2011; Gabriel et al., 2017). Considering that ANN's superiority was based on bulk data, ANN still has potential for bread quality prediction and needs further exploration using a larger dataset (Mutlu et al., 2011).

Bread texture analysis and other parameters

Bread storage is of great importance. Bread staling is associated with a complicated interaction between starch, water, and protein in bread, and it is because of this that NIRS may be used to measure staling-related parameters (Amigo et al., 2019; Nallan Chakravartula et al., 2019). The first study to explore the possibility of using NIRS for measuring staling was conducted by Wilson et al., 1991, and the results show that NIRS was able to follow bread staling in a specific time scale. A further study compared NIRS with the texture analyzer for measuring bread change during storage (Xie et al., 2003). The results showed that NIRS was comparable to the textual analyzer, even better in terms of measuring texture hardness (for bread slices and bread loaves), which was possibly due to NIR spectra simultaneously containing chemical and physical information (Xie et al., 2003). In addition, the results indicated that the prediction for loaves was better than that for the slices both in

hardness and storage time (Xie et al., 2003). In the study of Ringsted et al. (2017), changes in bread crumb hardness during aging were correlated with spectra obtained from NIRS and MIRS, respectively. The Pearson's correlation coefficient was .98, demonstrating its potential in bread staling detection (Ringsted et al., 2017). Another study on the application of NIRS in storage time prediction was conducted in 2015. A total of 72 samples distributed in seven different storage times were used to build PLSR models, and the optimal PLSR model performance achieved an R_p^2 of .969 (Cevoli et al., 2015).

Application as an assistant-characterization tool

Besides the direct application of NIRS in bread quality assessment, in some studies of food additives, NIRS was used as an assistant tool for quality assessment. In the studies of Abu-Ghoush et al. on antimicrobial agents and preservation addition for shelf-life extension, NIRS was successful to model loaf freshness based on the spectra collected from the bottom and top of the loaves (Abu-Ghoush, Herald, Dowell, Xie, Aramouni, & Madl, 2008; Abu-Ghoush, Herald, Dowell, Xie, Aramouni, & Walker, 2008). Recently, NIRS was used to detect crumb texture (hardness) and assess the staling of white bread and the effects of antistaling enzymes (Amigo et al., 2019). For the untreated loaf samples, the PLSR model obtained satisfactory results ($R_p^2 = .86$), but the results were inferior for the samples treated with antistaling enzymes, possibly because enzyme treatment led to the increase of nonlinearity in the staling process (Amigo et al., 2019).

In a study of the edible coating of mini-buns, PLSR obtained satisfactory results predicting both crumb chewiness and hardness in one-layer coating and two-layer coating buns, with the latter results being better (Nallan Chakravartula et al., 2019). In addition, moisture contents of the top, bottom, and center of the one- or two-layer coating buns were modeled using PLSR, and the results for one-layer coating buns were better than those for two-layer coating buns (Nallan Chakravartula et al., 2019).

3.4.2 | Other wheat-based products

In the past 10 years, NIRS was also used to study the baking parameters in cake, pastas, and cookies (Table 6). Nutritional parameters that consider the daily needs of consumers gained the most attention. In 2011, the protein, lipid, and sugar contents in different types of bread, cake, doughnuts, and croissants were modeled by PLSR based on spectra collected from bakery product powders (Szigedi et al., 2011). Unfortunately, the study had no prediction dataset. Subsequently, a study conducted in 2019 by Neves et al. added a prediction dataset for model evaluation and

also extended nutritional parameters. NIRS was applied in pasta blended with tomato sauce to measure nutritional parameters including energy, carbohydrate, fat, fiber, protein, and sugar. The pasta and sauce were produced by different manufacturers. A handheld NIR spectrometer was used to collect five blend proportions (0%, 25%, 50%, 75%, and 100%) (Neves et al., 2019). The R_p^2 s were all above .85 and even reached .90 for fiber content, which proved that not only could these parameters be determined in wheat or flour, but also in final processed products (Szigedi et al., 2011; Neves et al., 2019). Recently, shelf life of pasta stored at different temperatures was modeled based on NIRS information, and the results showed that absorption wavelengths related to gluten, starch, and water played an important role in the great model performance ($r_{CV} > .98$) (Zardetto et al., 2021).

The spatial imaging ability of HSI made it possible to detect the entire samples instead of local areas. Previous work in NIRS determined wavelengths that are highly correlated to water, and Andresen et al. (2013) used this conclusion to select wavelengths to measure water content in butter cake. The water content in each visualized pixel of the cake was predicted. The average water content of the whole cake was then calculated with the RPD = 0.22%. Another strategy was to first calculate the average spectra region of interest (ROI) and then use it for model development for global parameters. Multiple combinations of wavelengths of the average spectra of ROI were respectively modeled by PLSR methods for hardness and moisture determination (Polak et al., 2019). Moisture determination gained good results with even single wavelengths achieving an $R_p^2 = .9449$, while the R_p^2 for hardness was .7956 (Polak et al., 2019). Using the full wavelength range of the spectra gained the best performance both in moisture and hardness determination, which was consistent with previous predictions based on spectra collected from flour and kernels (Polak et al., 2019). Recently, water activity and storage time of cakes were modeled, and the PLSR models resulted in $R_p^2 = .767$ and RMSEP = 0.013, and $R_p^2 = .835$ and RMSEP = 1.242, respectively, both of which still need further improvement (Sricharoonratana & Teerachaichayut, 2020; Sricharoonratana et al., 2021).

4 | CONCLUSION AND PERSPECTIVES

This review provides an overview of NIRS methodology in biological material analysis and its application in the quantitative determination of wheat and wheat product quality. The pretreatment methods, spectral wavelength selection methods, outlier disposal, dataset division, regression methods, model evaluation, and industry applications are introduced. There exist stable and highly accurate mod-

els for the determination of most of the wheat composition and associated parameters. Performance for some rheological properties and end product quality is satisfactory, though further improvement is expected in such aspects. Most studies still depend on common pretreatment methods (e.g., SNV, MSC, and FD), manual spectral wavelength selection or full ranges, and linear regression methods (e.g., PLSR and multiple linear regression [MLR]). Although some of these studies achieved good accuracies, the models that adopt cutting-edge pretreatment and wavelength selection methods as well as regression methods are more likely to achieve better results. On the other hand, the main limitations of NIRS are its high dependence on dataset properties and model development through chemometrics methods (Gatius et al., 2017; Manley, 2014).

Possible directions for future studies on the application of NIRS in wheat are proposed as follows:

1. Integration of different kinds of spectroscopic techniques and establishing different types of spectral databases can be the basis for more robust calibration models. For example, FS has shown feasibility for predicting analytical, rheological, and baking parameters (Ahmad et al., 2016). The HSI technique can provide extra information for determining the spatial distribution of quality parameters, which is in great demand for wheat and wheat products assessments (Caporaso et al., 2018a).
2. Introduction of new pretreatment methods and nonlinear regression methods for model development is essential. Spectral data have high dimensions, and how to pretreat and extract representative wavelengths and perform regression analysis is the focus of data science. Experiments should be conducted to test the performance of new methods and modify them for wheat quality assessments, for example, the applications of restricted Boltzmann machine (RBM) and DBN, which had made breakthroughs in model performance (Li et al., 2017; Harrington, 2018).
3. It is necessary to extend the range of reference values beyond wet chemistry experiments by using manual interventions in samples. Larger dataset, which refers to more samples and wider range of reference values, has shown its superiority in reliable and robust model building when combining with advanced chemometrics methods (Bian et al., 2018; Cui & Fearn, 2018; Delwiche et al., 2018; Dong & Sun, 2013; Gatius et al., 2017; Owens et al., 2009; Li et al., 2017; Mutlu et al., 2011; Salimi Khorshidi et al., 2018; Williams, 2020; Zhang et al., 2019). By manual intervention in cultivation or postharvest processing, it is possible to gain a larger dataset. For example, extraneous wheat proteins could

be added into wheat flour to increase the protein content and collect its associated NIR spectra. The compositions of natural wheat, its flour, and commercial products are usually at a relatively narrow range, so manual intervention is in demand for creating different gradients of a specific composition for acquisition of spectral data associated with a larger range of reference values.

4. A secondary calibration model may be taken into consideration to calibrate prediction values due to the interactive effects of different compositions. External factors (e.g., growing season) and prediction values (e.g., protein content and baking parameters) should be combined to correct the predicted values. For example, it has been shown that when protein content reached a certain level, baking quality no longer increases with the increase of protein content (Gabriel et al., 2017). Therefore, after prediction of the first calibration, the weight of protein content in baking quality should be revised in the second calibration model to gain more accurate assessment.
5. Applications in micronutrient analysis, such as polyphenols, vitamins, and minerals, need to be extended. There were few studies on this topic, and the methods for calibration model development also need to be advanced for these low concentration components (Tian et al., 2021).
6. Efforts should be made for practical and commercial application of NIRS in shelf life prediction of wheat products, since the performance of application of NIRS in analyzing compositions as well as the end products has been improved gradually over the years (Gao et al., 2017; Pandey et al., 2018; Jiang et al., 2020; Sricharoonratana & Teerachaichayut, 2020; Sricharoonratana et al., 2021; Zardetto et al., 2021).
7. Reference values from the sensory evaluation of end products can supplement current reference values from wet chemistry experiments. Given the consistency in industrial manufacturing parameters, it is feasible to employ a professional sensory evaluation group to assess end products based on standard manufacturing processes and then combine these assessments with spectral data to develop calibration models.

ACKNOWLEDGMENTS

This is contribution no. 22-121-J from the Kansas Agricultural Experimental Station. This work is in part supported by the USDA National Institute of Food and Agriculture Hatch project KS17HA1008 and USDA Agricultural Research Service Cooperative Agreement 58-3020-9-017. 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.

AUTHOR CONTRIBUTIONS

Zhenjiao Du: Conceptualization; formal analysis; investigation; methodology; validation; visualization; writing – original draft; writing – review & editing. **Wenfei Tian:** Writing – review & editing. **Michael Tilley:** Resources; writing – review & editing. **Donghai Wang:** Resources; writing – review & editing. **Guorong Zhang:** Resources; writing – review & editing. **Yonghui Li:** Conceptualization; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; writing – original draft; writing – review & editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Yonghui Li  <https://orcid.org/0000-0003-4320-0806>

REFERENCES

- Abu-Ghoush, M., Herald, T. J., Dowell, F., Xie, F., Aramouni, F. M., & Madl, R. (2008). Effect of preservatives addition on the shelf-life extensions and quality of flat bread as determined by near-infrared spectroscopy and texture analysis. *International Journal of Food Science and Technology*, 43(2), 357–364. <https://doi.org/10.1111/j.1365-2621.2007.01594.x>
- Abu-Ghoush, M., Herald, T. J., Dowell, F., Xie, F., Aramouni, F. M., & Walker, C. (2008). Effect of antimicrobial agents and dough conditioners on the shelf-life extension and quality of flat bread, as determined by near-infrared spectroscopy. *International Journal of Food Science and Technology*, 43(2), 365–372. <https://doi.org/10.1111/j.1365-2621.2007.01625.x>
- Agelet, L. E., & Hurburgh, C. R. (2010). A tutorial on near infrared spectroscopy and its calibration. *Critical Reviews in Analytical Chemistry*, 40(4), 246–260. <https://doi.org/10.1080/10408347.2010.515468>
- Ahmad, M. H., Nache, M., Waffenschmidt, S., & Hitzmann, B. (2016). A fluorescence spectroscopic approach to predict analytical, rheological and baking parameters of wheat flours using chemometrics. *Journal of Food Engineering*, 182, 65–71. <https://doi.org/10.1016/j.jfoodeng.2016.03.006>
- Albanell, E., Martínez, M., De Marchi, M., & Manuelian, C. L. (2021). Prediction of bioactive compounds in barley by near-infrared reflectance spectroscopy (NIRS). *Journal of Food Composition and Analysis*, 97, 103763. <https://doi.org/10.1016/j.jfca.2020.103763>
- Amigo, J. M., Del Olmo, A., Engelsen, M. M., Lundkvist, H., & Engelsen, S. B. (2019). Staling of white wheat bread crumb and effect of maltogenic α -amylases. Part 2: Monitoring the staling process by using near infrared spectroscopy and chemometrics. *Food Chemistry*, 297, 124946. <https://doi.org/10.1016/j.foodchem.2019.06.013>
- Andresen, M. S., Dissing, B. S., & Løje, H. (2013). Quality assessment of butter cookies applying multispectral imaging. *Food Science & Nutrition*, 1(4), 315–323. <https://doi.org/10.1002/fsn3.46>

- Arazuri, S., Ignacio Arana, J., Arias, N., Arregui, L. M., Gonzalez-Torralba, J., & Jaren, C. (2012). Rheological parameters determination using near infrared technology in whole wheat grain. *Journal of Food Engineering*, *111*(1), 115–121. <https://doi.org/10.1016/j.foodeng.2012.01.017>
- Armstrong, P. R. (2014). Development and evaluation of a near-infrared instrument for single-seed compositional measurement of wheat kernels. *Cereal Chemistry*, *91*(1), 23–28. <https://doi.org/10.1094/CCHEM-07-13-0132-R>
- Badaró, A. T., Morimitsu, F. L., Ferreira, A. R., Clerici, M. T. P. S., & Fernandes Barbin, D. (2019). Identification of fiber added to semolina by near infrared (NIR) spectral techniques. *Food Chemistry*, *289*, 195–203. <https://doi.org/10.1016/j.foodchem.2019.03.057>
- 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>
- Başlar, M., & Ertugay, M. F. (2011). Determination of protein and gluten quality-related parameters of wheat flour using near-infrared reflectance spectroscopy (NIRS). *Turkish Journal of Agriculture and Forestry*, *35*(2), 139–144. <https://doi.org/10.3906/tar-0912-507>
- Besançon, L., Rondet, E., Grabulos, J., Lullien-Pellerin, V., Lhomond, L., & Cuq, B. (2020). Study of the microstructure of durum wheat endosperm using X-ray micro-computed tomography. *Journal of Cereal Science*, *96*, 103115. <https://doi.org/10.1016/j.jcs.2020.103115>
- Bian, X., Diwu, P., Zhang, C., Lin, L., Chen, G., Tan, X., Guo, Y., & Cheng, B. (2018). Robust boosting neural networks with random weights for multivariate calibration of complex samples. *Analytica Chimica Acta*, *1009*, 20–26. <https://doi.org/10.1016/j.aca.2018.01.013>
- Botosoa, E. P., Chéné, C., & Karoui, R. (2013). Use of front face fluorescence for monitoring lipid oxidation during ageing of cakes. *Food Chemistry*, *141*(2), 1130–1139. <https://doi.org/10.1016/j.foodchem.2013.03.059>
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>
- 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*, 1–15. <https://doi.org/10.1255/jsi.2017.a4>
- Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2018a). 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>
- Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2018b). Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chemistry*, *240*, 32–42. <https://doi.org/10.1016/j.foodchem.2017.07.048>
- Cauvain, S. P., & Young, L. S. (2009). *Bakery food manufacture and quality: Water control and effects*. John Wiley & Sons. <https://doi.org/10.1002/9781444301083.ch1>
- Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, *18*(2), 72–83. <https://doi.org/10.1016/j.tifs.2006.09.003>
- Centner, V., De Noord, O. E., & Massart, D. L. (1998). Detection of nonlinearity in multivariate calibration. *Analytica Chimica Acta*, *376*(2), 153–168. [https://doi.org/10.1016/S0003-2670\(98\)00543-1](https://doi.org/10.1016/S0003-2670(98)00543-1)
- Cereals & Grains Association. (1999a). *AACC Approved Methods of Analysis, 11th Ed. Method 39-11.01: Near-infrared reflectance method for protein determination in wheat flour*. <http://doi.org/10.1094/AACCIntMethod-39-11.01>
- Cereals & Grains Association. (1999b). *AACC Approved Methods of Analysis, 11th Ed. Method 39-25.01: Near-infrared reflectance method for protein content in whole-grain wheat*. <http://doi.org/10.1094/AACCIntMethod-39-25.01>
- Cereals & Grains Association. (1999c). *AACC Approved Methods of Analysis, 11th Ed. Method 39-70.02: Near-infrared reflectance method for hardness determination in wheat*. <http://doi.org/10.1094/AACCIntMethod-39-70.02>
- Cereals & Grains Association. (2020). *AACC Approved Methods of Analysis, 11th Ed. Method 08-21.01: Prediction of ash content in wheat flour—Near-infrared method*. <http://doi.org/10.1094/AACCIntMethod-08-21.01>
- Cevoli, C., Gianotti, A., Troncoso, R., & Fabbri, A. (2015). Quality evaluation by physical tests of a traditional Italian flat bread Piadina during storage and shelf-life improvement with sourdough and enzymes. *European Food Research and Technology*, *240*(6), 1081–1089. <https://doi.org/10.1007/s00217-015-2429-7>
- Che, W., Sun, L., Zhang, Q., Zhang, D., Ye, D., Tan, W., & Dai, C. (2017). Application of visible/near-infrared spectroscopy in the prediction of azodicarbonamide in wheat flour. *Journal of Food Science*, *82*(10), 2516–2525. <https://doi.org/10.1111/1750-3841.13859>
- Chen, J., Zhu, S., & Zhao, G. (2017). Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR. *Food Chemistry*, *221*, 1939–1946. <https://doi.org/10.1016/j.foodchem.2016.11.155>
- Chen, Q., Ding, J., Cai, J., & Zhao, J. (2012). Rapid measurement of total acid content (TAC) in vinegar using near infrared spectroscopy based on efficient variables selection algorithm and nonlinear regression tools. *Food Chemistry*, *135*(2), 590–595. <https://doi.org/10.1016/j.foodchem.2012.05.011>
- Chen, Q., Guo, Z., Zhao, J., & Ouyang, Q. (2012). Comparisons of different regressions tools in measurement of antioxidant activity in green tea using near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, *60*, 92–97. <https://doi.org/10.1016/j.jpba.2011.10.020>
- Chen, Q., Zhao, J., Liu, M., Cai, J., & Liu, J. (2008). Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. *Journal of Pharmaceutical and Biomedical Analysis*, *46*(3), 568–573. <https://doi.org/10.1016/j.jpba.2007.10.031>
- Chen, X., Siesler, H. W., & Yan, H. (2021). Rapid analysis of wheat flour by different handheld near-infrared spectrometers: A discussion of calibration model maintenance and performance comparison. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *252*, 119504. <https://doi.org/10.1016/j.saa.2021.119504>
- Cui, C., & Fearn, T. (2018). Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. *Chemometrics and Intelligent Laboratory Systems*, *182*, 9–20. <https://doi.org/10.1016/j.chemolab.2018.07.008>
- Delwiche, S. R., & Graybosch, R. A. (2014). Measurement of blend concentrations of conventional and waxy hard wheats using NIR Spectroscopy. *Cereal Chemistry*, *91*(4), 358–365. <https://doi.org/10.1094/CCHEM-09-13-0195-R>
- Delwiche, S. R., & Graybosch, R. A. (2016). Binary mixtures of waxy wheat and conventional wheat as measured by NIR reflectance.

- Talanta, 146, 496–506. <https://doi.org/10.1016/j.talanta.2015.08.063>
- 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(3), 469–477. <https://doi.org/10.1002/cche.10049>
- Dixon, W. J. (1950). Analysis of extreme values. *The Annual Mathematics Statistics*, 21(4), 488–506. <https://doi.org/10.1214/aoms/1177729747>
- Dhanoa, M. S., Lister, S. J., Sanderson, R., & Barnes, R. J. (1994). The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. *Journal of Near Infrared Spectroscopy*, 2(1), 43–47. <https://doi.org/10.1255/jnirs.30>
- Dong, X., & Sun, X. (2013). A case study of characteristic bands selection in near-infrared spectroscopy: Nondestructive detection of ash and moisture in wheat flour. *Journal of Food Measurement and Characterization*, 7(3), 141–148. <https://doi.org/10.1007/s11694-013-9149-0>
- Dowell, F. E., Maghirang, E. B., Xie, F., Lookhart, G. L., Pierce, R. O., Seabourn, B. W., & Chung, O. K. (2006). Predicting wheat quality characteristics and functionality using near-infrared spectroscopy. *Cereal Chemistry*, 83(5), 529–536. <https://doi.org/10.1094/CC-83-0529>
- Du, Y. P., Liang, Y. Z., Jiang, J. H., Berry, R. J., & Ozaki, Y. (2004). Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares. *Analytica Chimica Acta*, 501(2), 183–191. <https://doi.org/10.1016/j.aca.2003.09.041>
- Du, Z., Hu, Y., Ali Buttar, N., & Mahmood, A. (2019). X-ray computed tomography for quality inspection of agricultural products: A review. *Food Science and Nutrition*, 7, 3146–3160. <https://doi.org/10.1002/fsn3.1179>
- Erkinbaev, C., Derksen, K., & Paliwal, J. (2019). Single kernel wheat hardness estimation using near infrared hyperspectral imaging. *Infrared Physics and Technology*, 98, 250–255. <https://doi.org/10.1016/j.infrared.2019.03.033>
- Ezeanaka, M. C., Nsor-Atindana, J., & Zhang, M. (2019). Online low-field nuclear magnetic resonance (LF-NMR) and magnetic resonance imaging (MRI) for food quality optimization in food processing. *Food and Bioprocess Technology*, 12(9), 1435–1451. <https://doi.org/10.1007/s11947-019-02296-w>
- Ferreira, D. S., Poppi, R. J., & Lima Pallone, J. A. (2015). Evaluation of dietary fiber of Brazilian soybean (Glycine max) using near-infrared spectroscopy and chemometrics. *Journal of Cereal Science*, 64, 43–47. <https://doi.org/10.1016/j.jcs.2015.04.004>
- Finney, P. L., Kinney, J. E., & Donelson, J. R. (1988). Prediction of damaged starch in straight-grade flour by near-infrared reflectance analysis of whole ground wheat. *Cereal Chemistry*, 65, 449–452.
- Fox, G., & Manley, M. (2014). Applications of single kernel conventional and hyperspectral imaging near infrared spectroscopy in cereals. *Journal of the Science of Food and Agriculture*, 94, 174–179. <https://doi.org/10.1002/jsfa.6367>
- Gabriel, D., Pfitzner, C., Haase, N. U., Hüsken, A., Prüfer, H., Greef, J. M., & Rühl, G. (2017). New strategies for a reliable assessment of baking quality of wheat – Rethinking the current indicator protein content. *Journal of Cereal Science*, 77, 126–134. <https://doi.org/10.1016/j.jcs.2017.08.002>
- Gao, D., Jian, G., Cheng, W., Kehong, L., Lingang, L., Jing, W., & Dazhou, Z. (2017). Differentiation of storage time of wheat seed based on near infrared hyperspectral imaging. *International Journal of Agricultural and Biological Engineering*, 10(2), 251–258. <https://doi.org/10.3965/j.ijabe.20171002.1619>
- Garnsworthy, P. C., Wisenman, J., & Fegeros, K. (2000). Prediction of chemical, nutritive and agronomic characteristics of wheat by near infrared spectroscopy. *The Journal of Agricultural Science*, 135(4), 409–417. <https://doi.org/10.1017/S0021859699008382>
- Gatus, F., Miralbés, C., David, C., & Puy, J. (2017). Comparison of CCA and PLS to explore and model NIR data. *Chemometrics and Intelligent Laboratory Systems*, 164, 76–82. <https://doi.org/10.1016/j.chemolab.2017.03.011>
- Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *The Annals of Mathematical Statistics*, 21(1), 27–58. <https://doi.org/10.1214/aoms/1177729885>
- Harrington, P. d. B. (2018). Feature expansion by a continuous restricted Boltzmann machine for near-infrared spectrometric calibration. *Analytica Chimica Acta*, 1010, 20–28. <https://doi.org/10.1016/j.aca.2018.01.026>
- Hell, J., Prückler, M., Danner, L., Henniges, U., Apprich, S., Rosenau, T., & Böhmendorfer, S. (2016). A comparison between near-infrared (NIR) and mid-infrared (ATR-FTIR) spectroscopy for the multivariate determination of compositional properties in wheat bran samples. *Food Control*, 60, 365–369. <https://doi.org/10.1016/j.foodcont.2015.08.003>
- Horvath, L., Norris, K. H., Horvath-Mosonyi, M., Rigo, J., & Hegedus-Volgyesi, E. (1984). Study into determining dietary fiber of wheat bran by NIR-technique. *Acta Alimentaria*, 13, 355–382.
- Hruskova, M., Bednarova, M., & Novotny, F. (2001). Wheat flour dough rheological characteristics predicted by NIRSystems 6500. *Czech Journal of Food Sciences*, 19, 213–218. <https://doi.org/10.17221/6610-cjfs>
- Hruskova, M., & Famera, O. (2003). Prediction of wheat and flour zeleny sedimentation value using NIR technique. *Czech Journal of Food Sciences*, 21, 91–96.
- Isaksson, T., & Kowalski, B. (1993). Piece-wise multiplicative scatter correction applied to near-infrared diffuse transmittance data from meat products. *Applied Spectroscopy*, 47(6), 702–709. <https://doi.org/10.1366/0003702934066839>
- Jiang, H., & Lu, J. (2018). Using an optimal CC-PLSR-RBFNN model and NIR spectroscopy for the starch content determination in corn. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 196, 131–140. <https://doi.org/10.1016/j.saa.2018.02.017>
- Jiang, H., Liu, T., & Chen, Q. (2020). Quantitative detection of fatty acid value during storage of wheat flour based on a portable near-infrared (NIR) spectroscopy system. *Infrared Physics and Technology*, 109, 103423. <https://doi.org/10.1016/j.infrared.2020.103423>
- Jiang, J., Zhao, M., Liu, Q., & Thermo Fisher, S. (2007). *Rapid analysis of multiple components in tobacco using the antaris II FT-NIR analyzer*. Thermo Fisher, Scientific. <https://assets.thermofisher.com/TFS-Assets/CAD/Application-Notes/D14198~.pdf>
- Jirsa, O., Hrušková, M., & Švec, I. (2008). Near-infrared prediction of milling and baking parameters of wheat varieties. *Journal of Food Engineering*, 87(1), 21–25. <https://doi.org/10.1016/j.jfoodeng.2007.09.008>

- Kays, S. E., & Barton, F. E. (2002). Near-infrared analysis of soluble and insoluble dietary fiber fractions of cereal food products. *Journal of Agricultural and Food Chemistry*, *50*(10), 3024–3029. <https://doi.org/10.1021/jf0116552>
- Kays, S. E., Windham, W. R., & Barton, F. E. (1996). Prediction of total dietary fiber in cereal products using near-infrared reflectance spectroscopy. *Journal of Agricultural and Food Chemistry*, *44*(8), 2266–2271. <https://doi.org/10.1021/jf960053t>
- Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, *120*, 126–135. <https://doi.org/10.1016/j.chemolab.2012.11.009>
- Lasztity, R., & Abonyi, T. (2009). Prediction of wheat quality past, present, future. A review. *Food Reviews International*, *25*(2), 126–141. <https://doi.org/10.1080/87559120802682706>
- Leardl, R., & Norgaard, L. (2004). Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *Journal of Chemometrics*, *18*(11), 486–497. <https://doi.org/10.1002/cem.893>
- Li, H. D., Xu, Q. S., & Liang, Y. Z. (2012). Random frog: An efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification. *Analytica Chimica Acta*, *740*, 20–26. <https://doi.org/10.1016/j.aca.2012.06.031>
- Li Vigni, M., & Cocchi, M. (2013). Near infrared spectroscopy and multivariate analysis to evaluate wheat flour doughs leavening and bread properties. *Analytica Chimica Acta*, *764*, 17–23. <https://doi.org/10.1016/j.aca.2012.12.018>
- Li, W., Lin, M., Huang, Y., Liu, H., & Zhou, X. (2017). Near infrared spectroscopy detection of the content of wheat based on improved deep belief network. *Journal of Physics: Conference Series*, *887*(1), 012046. <https://doi.org/10.1088/1742-6596/887/1/012046>
- Li, X., Zhang, L., Zhang, Y., Wang, D., Wang, X., Yu, L., & Li, P. (2020). Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. *Trends in Food Science & Technology*, *101*, 172–181. <https://doi.org/10.1016/j.tifs.2020.05.002>
- Liang, N., Sun, S., Zhang, C., He, Y., & Qiu, Z. (2022). Advances in infrared spectroscopy combined with artificial neural network for the authentication and traceability of food. *Critical Reviews in Food Science and Nutrition*, *62*(11), 2963–2984. <https://doi.org/10.1080/10408398.2020.1862045>
- Liu, X., Rong, Y. Z., Zhang, X., Mao, D. Z., Yang, Y. J., & Wang, Z. W. (2015). Rapid determination of total dietary fiber and minerals in coix seed by near-infrared spectroscopy technology based on variable selection methods. *Food Analytical Methods*, *8*(7), 1607–1617. <https://doi.org/10.1007/s12161-014-0037-y>
- Luo, Q., Yun, Y., Fan, W., Huang, J., Zhang, L., Deng, B., & Lu, H. (2015). Application of near infrared spectroscopy for the rapid determination of epimedin A, B, C and icariin in Epimedium. *RSC Advances*, *5*(7), 5046–5052. <https://doi.org/10.1039/c4ra11421c>
- Luypaert, J., Heuerding, S., Vander Heyden, Y., & Massart, D. L. (2004). The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams. *Journal of Pharmaceutical and Biomedical Analysis*, *36*(3), 495–503. <https://doi.org/10.1016/j.jpba.2004.06.023>
- Luypaert, J., Massart, D. L., & Vander Heyden, Y. (2007). Near-infrared spectroscopy applications in pharmaceutical analysis. *Talanta*, *72*(3), 865–883. <https://doi.org/10.1016/j.talanta.2006.12.023>
- Mahesh, S., Jayas, D. S., Paliwal, J., & White, N. D. G. (2014). comparison of partial least squares regression (PLSR) and principal components regression (PCR) methods for protein and hardness predictions using the near-infrared (NIR) hyperspectral images of bulk samples of canadian wheat. *Food and Bioprocess Technology*, *8*(1), 31–40. <https://doi.org/10.1007/s11947-014-1381-z>
- Manley, M., Van Zyl, L., & Osborne, B. G. (2002). Using fourier transform near infrared spectroscopy in determining kernel hardness, protein and moisture content of whole wheat flour. *Journal of Near Infrared Spectroscopy*, *10*(1), 71–76. <https://doi.org/10.1255/jnirs.323>
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews*, *43*(24), 8200–8214. <https://doi.org/10.1039/c4cs00062e>
- Mao, X., Sun, L., Hui, G., & Xu, L. (2014). Modeling research on wheat protein content measurement using near-infrared reflectance spectroscopy and optimized radial basis function neural network. *Journal of Food and Drug Analysis*, *22*(2), 230–235. <https://doi.org/10.1016/j.jfda.2014.01.023>
- Mark, H., & Workman, J. (2005). Chemometrics in spectroscopy linearity in calibration: The Durbin-Watson statistic. *Spectroscopy-trend*, *20*(3), 34–39.
- Miralbés, C. (2003). Prediction chemical composition and alveograph parameters on wheat by near-infrared transmittance spectroscopy. *Journal of Agricultural and Food Chemistry*, *51*(21), 6335–6339. <https://doi.org/10.1021/jf034235g>
- Miralbés, C. (2004). Quality control in the milling industry using near infrared transmittance spectroscopy. *Food Chemistry*, *88*(4), 621–628. <https://doi.org/10.1016/j.foodchem.2004.05.004>
- Mishra, G., Srivastava, S., Panda, B. K., & Mishra, H. N. (2018). Rapid assessment of quality change and insect infestation in stored wheat grain using FT-NIR spectroscopy and chemometrics. *Food Analytical Methods*, *11*(4), 1189–1198. <https://doi.org/10.1007/s12161-017-1094-9>
- Morgan, J. E., & Williams, P. C. (1995). Starch damage in wheat flours: A comparison of enzymatic, lodometric, and near-infrared reflectance techniques. *Cereal Chemistry*, *72*, 209–212.
- Mutlu, A. C., Boyaci, I. H., Genis, H. E., Ozturk, R., Basaran-Akgul, N., Sanal, T., & Evlice, A. K. (2011). Prediction of wheat quality parameters using near-infrared spectroscopy and artificial neural networks. *European Food Research and Technology*, *233*(2), 267–274. <https://doi.org/10.1007/s00217-011-1515-8>
- Nallan Chakravartula, S. S., Cevoli, C., Balestra, F., Fabbri, A., & Rosa, M. D. (2019). Evaluation of the effect of edible coating on mini-buns during storage by using NIR spectroscopy. *Journal of Food Engineering*, *263*, 46–52. <https://doi.org/10.1016/j.jfoodeng.2019.05.035>
- Neves, M. D. G., Poppi, R. J., & Siesler, H. W. (2019). Rapid determination of nutritional parameters of pasta/sauce blends by handheld near-infrared spectroscopy. *Molecules*, *24*(11), 2029. <https://doi.org/10.3390/molecules24112029>
- Norris, K. H., Hruschka, W. R., Bean, M. M., & Slaughter, D. C. (1989). A definition of wheat hardness using near-infrared spectroscopy. *Cereal Foods World*, *34*, 696–705.
- Norris, K. H., & Williams, P. C. (1984). Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. influence of particle size. *Cereal Chemistry*, *61*, 158–165.

- Owens, B., McCann, M. E. E., McCracken, K. J., & Park, R. S. (2009). Prediction of wheat chemical and physical characteristics and nutritive value by near-infrared reflectance spectroscopy. *British Poultry Science*, 50(1), 103–122. <https://doi.org/10.1080/00071660802635347>
- Pandey, P., Srivastava, S., & Mishra, H. N. (2018). Comparison of FT-NIR and NIR for evaluation of physico-chemical properties of stored wheat grains. *Food Quality and Safety*, 2(3), 165–172. <https://doi.org/10.1093/fqsafe/fyy015>
- Peiris, K. H. S., Dong, Y., Davis, M. A., Bockus, W. W., & Dowell, F. E. (2017). Estimation of the deoxynivalenol and moisture contents of bulk wheat grain samples by FT-NIR spectroscopy. *Cereal Chemistry*, 94(4), 677–682. <https://doi.org/10.1094/CCHEM-11-16-0271-R>
- Pojić, M. M., & Mastilović, J. S. (2013). Near infrared spectroscopy-advanced analytical tool in wheat breeding, trade, and processing. *Food and Bioprocess Technology*, 6(2), 330–352. <https://doi.org/10.1007/s11947-012-0917-3>
- Polak, A., Coutts, F. K., Murray, P., & Marshall, S. (2019). Use of hyperspectral imaging for cake moisture and hardness prediction. *IET Image Processing*, 13(7), 1152–1160. <https://doi.org/10.1049/iet-ipr.2018.5106>
- Ringsted, T., Siesler, H. W., & Engelsen, S. B. (2017). Monitoring the staling of wheat bread using 2D MIR-NIR correlation spectroscopy. *Journal of Cereal Science*, 75, 92–99. <https://doi.org/10.1016/j.jcs.2017.03.006>
- Rinnan, Å., Berg, F. v. d., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, 28(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Rubenthaler, G. L., & Pomeranz, Y. (1987). Near-infrared reflectance spectra of hard red winter wheats varying widely in protein content and breadmaking potential. *Cereal Chemistry*, 64(6), 407–411.
- Sadat, A., Corradini, M. G., & Joye, I. J. (2019). Molecular spectroscopy to assess protein structures within cereal systems. *Current Opinion in Food Science*, 25, 42–51. <https://doi.org/10.1016/j.cofs.2019.02.001>
- Salgó, A., & Gergely, S. (2012). Analysis of wheat grain development using NIR spectroscopy. *Journal of Cereal Science*, 56(1), 31–38. <https://doi.org/10.1016/j.jcs.2012.04.011>
- Salimi Khorshidi, A., Storsley, J., Malunga, L. N., Thandapilly, S. J., & Ames, N. (2018). Advancing the science of wheat quality evaluation using nuclear magnetic resonance (NMR) and ultrasound-based techniques. *Cereal Chemistry*, 95(3), 347–364. <https://doi.org/10.1002/cche.10040>
- Sato, T., Morishita, T., Hara, T., Suda, I., & Tetsuka, T. (2001). Near-infrared reflectance spectroscopic analysis of moisture, fat, protein, and physiological activity in buckwheat flour for breeding selection. *Plant Production Science*, 4(4), 270–277. <https://doi.org/10.1626/pps.4.270>
- Sendin, K., Williams, P. J., & Manley, M. (2018). Near infrared hyperspectral imaging in quality and safety evaluation of cereals. *Critical Reviews in Food Science and Nutrition*, 58(4), 575–590. <https://doi.org/10.1080/10408398.2016.1205548>
- Shi, H., Lei, Y., Louzada Prates, L., & Yu, P. (2019). Evaluation of near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy techniques combined with chemometrics for the determination of crude protein and intestinal protein digestibility of wheat. *Food Chemistry*, 272, 507–513. <https://doi.org/10.1016/j.foodchem.2018.08.075>
- Shi, H., & Yu, P. (2017). Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. *Food Control*, 82, 57–65. <https://doi.org/10.1016/j.foodcont.2017.06.015>
- Sinelli, N., Pagani, M. A., Lucisano, M., D'Egidio, M. G., & Mariotti, M. (2011). Prediction of semolina technological quality by FT-NIR spectroscopy. *Journal of Cereal Science*, 54(2), 218–223. <https://doi.org/10.1016/j.jcs.2011.06.002>
- Sricharoonratana, M., & Teerachaichayut, S. (2020). Prediction of water activity in mamón (Filipino sponge) cakes by near infrared hyperspectral imaging. *Key Engineering Materials*, 862, 7–11. <https://doi.org/10.4028/www.scientific.net/KEM.862.7>
- Sricharoonratana, M., Thompson, A. K., & Teerachaichayut, S. (2021). Use of near infrared hyperspectral imaging as a nondestructive method of determining and classifying shelf life of cakes. *LWT - Food Science and Technology*, 136, 110369. <https://doi.org/10.1016/j.lwt.2020.110369>
- Sujka, K., Koczoń, P., Ceglińska, A., Reder, M., & Ciemnińska-Zytkeiwicz, H. (2017). The application of FT-IR spectroscopy for quality control of flours obtained from polish producers. *Journal of Analytical Methods in Chemistry*, 2017, 4315678. <https://doi.org/10.1155/2017/4315678>
- Szigedi, T., Dernovics, M., & Fodor, M. (2011). Determination of protein, lipid and sugar contents of bakery products by using fourier-transform near infrared spectroscopy. *Acta Alimentaria*, 40, 222–229. <https://doi.org/10.1556/AAlim.40.2011.Suppl.21>
- Teye, E., & Huang, X. (2015). Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection multivariate regression. *Food Analytical Methods*, 8(4), 945–953. <https://doi.org/10.1007/s12161-014-9933-4>
- Teye, E., Huang, X. Y., Lei, W., & Dai, H. (2014). Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans. *Food Research International*, 55, 288–293. <https://doi.org/10.1016/j.foodres.2013.11.021>
- Tian, W., Chen, G., Zhang, G., Wang, D., Tilley, M., & Li, Y. (2021). Rapid determination of total phenolic content of whole wheat flour using near-infrared spectroscopy and chemometrics. *Food Chemistry*, 344, 128633. <https://doi.org/10.1016/j.foodchem.2020.128633>
- USDA. (2020). World agricultural production. *Foreign Agricultural Service*, 7, 22–23. <https://doi.org/10.32317/2221-1055.201907059>
- Viljoen, M., Brand, T. S., Brandt, D. A., & Hoffman, L. C. (2005). Prediction of the chemical composition of winter grain and maize with near infrared reflectance spectroscopy. *South African Journal of Plant and Soil*, 22(2), 89–93. <https://doi.org/10.1080/02571862.2005.10634687>
- Westerhuis, J. A., De Jong, S., & Smilde, A. K. (2001). Direct orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems*, 56(1), 13–25. [https://doi.org/10.1016/S0169-7439\(01\)00102-2](https://doi.org/10.1016/S0169-7439(01)00102-2)
- Williams, P., Manley, M., & Antoniszyn, J. (2019). Near infrared technology: Getting the best out of light. *African Sun Media under the Sun Press*, <https://doi.org/10.1590/S0103-50532003000200006>

- Williams, P. C. (2020). Application of chemometrics to prediction of some wheat quality factors by near-infrared spectroscopy. *Cereal Chemistry*, 97(5), 958–966. <https://doi.org/10.1002/cche.10318>
- Wilson, R. H., Goodfellow, B. J., Belton, P. S., Osborne, B. G., Oliver, G., & Russell, P. L. (1991). Comparison of fourier transform mid infrared spectroscopy and near infrared reflectance spectroscopy with differential scanning calorimetry for the study of the staling of bread. *Journal of the Science of Food and Agriculture*, 54(3), 471–483. <https://doi.org/10.1002/jsfa.2740540318>
- Xie, F., Dowell, F. E., & Sun, X. S. (2003). Comparison of near-infrared reflectance spectroscopy and texture analyzer for measuring wheat bread changes in storage. *Cereal Chemistry*, 80(1), 25–29. <https://doi.org/10.1094/CCHEM.2003.80.1.25>
- Yang, M., Chen, Q., Kutsanedzie, F. Y. H., Yang, X., Guo, Z., & Ouyang, Q. (2017). Portable spectroscopy system determination of acid value in peanut oil based on variables selection algorithms. *Measurement: Journal of the International Measurement Confederation*, 103, 179–185. <https://doi.org/10.1016/j.measurement.2017.02.037>
- Ye, D., Sun, L., Zou, B., Zhang, Q., Tan, W., & Che, W. (2018). Non-destructive prediction of protein content in wheat using NIRS. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 189, 463–472. <https://doi.org/10.1016/j.saa.2017.08.055>
- Zardetto, S., Pasini, G., Romani, S., Rocculi, P., & D Rosa, M. (2021). Evaluation of physico-chemical changes and FT-NIR spectra in fresh egg pasta packed in modified atmosphere during storage at different temperatures. *Food Packaging and Shelf Life*, 28, 100648. <https://doi.org/10.1016/j.fpsl.2021.100648>
- Zareef, M., Chen, Q., Hassan, M. M., Arslan, M., Hashim, M. M., Ahmad, W., & Agyekum, A. A. (2020). An overview on the applications of typical non-linear algorithms coupled with NIR spectroscopy in food analysis. *Food Engineering Reviews*, 12(2), 173–190. <https://doi.org/10.1007/s12393-020-09210-7>
- Zeng, X. Y., Zhao, W. Q., Hu, X. Z., Li, X. P., Qiao, Y. Y., Ma, Z., & Zhang, Q. A. (2019). Determination of polyphenols in oats by near-infrared spectroscopy (NIRS) and two-dimensional correlation spectroscopy. *Analytical Letters*, 52(6), 962–971. <https://doi.org/10.1080/00032719.2018.1508295>
- Zhang, H., Gu, B., Mu, J., Ruan, P., & Li, D. (2017). Wheat hardness prediction research based on NIR hyperspectral analysis combined with ant colony optimization algorithm. *Procedia Engineering*, 174, 648–656. <https://doi.org/10.1016/j.proeng.2017.01.202>
- Zhang, X., Lin, T., Xu, J., Luo, X., & Ying, Y. (2019). DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. *Analytica Chimica Acta*, 1058, 48–57. <https://doi.org/10.1016/j.aca.2019.01.002>

How to cite this article: Du, Z., Tian, W., Tilley, M., Wang, D., Zhang, G., & Li, Y. (2022). Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review. *Comprehensive Reviews in Food Science and Food Safety*, 21, 2956–3009. <https://doi.org/10.1111/1541-4337.12958>