



Original Research

Prediction of Cooking Time for Soaked and Unsoaked Dry Beans (*Phaseolus vulgaris* L.) Using Hyperspectral Imaging Technology

Fernando A. Mendoza,* Jason A. Wiesinger, Renfu Lu, Susan Nchimbi-Msolla, Phillip N. Miklas, James D. Kelly, and Karen A. Cichy*

Core Ideas

- Hyperspectral imaging predicted water uptake and cooking time of dry bean from a single scan.
- Hyperspectral imaging has potential for estimating cooking time of soaked and unsoaked dry bean.
- Models succeeded for a broad set of cultivars with different cooking properties grown in diverse environments.
- Periodical updates with new accessions are critical to produce robust sensing calibration models.

The cooking time of dry bean varies widely by genotype and is also influenced by the growing environment, storage conditions, and cooking method. Thus high-throughput phenotyping methods to assess cooking time would be useful to breeders interested in developing cultivars with a desired cooking time. The objective of this study was to evaluate the performance of hyperspectral imaging technology for predicting dry bean cooking time. Fourteen dry bean genotypes with a wide range of cooking times were grown in five environments over 2 yr. Hyperspectral images were taken from whole dry seeds, and partial least squares regression models based on the extracted hyperspectral image features were developed to predict water uptake and cooking time of soaked and unsoaked beans. Relatively good predictions of water uptake were obtained, as measured by the correlation coefficient for prediction ($R_{\text{pred}} = 0.789$) and standard error of prediction (SEP = 4.4%). Good predictions of cooking time for soaked beans (ranging between 19.9–95.5 min) were achieved giving $R_{\text{pred}} = 0.886$ and SEP = 7.9 min. The prediction models for the cooking time of unsoaked beans (ranging between 80–147 min) were less robust and accurate ($R_{\text{pred}} = 0.708$, SEP = 10.6 min). This study demonstrated that hyperspectral imaging technology has potential for providing a nondestructive, simple, fast, and economical means for estimating the water uptake and cooking time of dry bean. Moreover, a totally independent set of 110 similar dry bean samples confirmed the suitability of the technique for predicting cooking time of soaked beans after updating the partial least squares model with 20 of the new samples, giving $R_{\text{pred}} = 0.872$ and SEP = 3.7 min. However, due to the genotypic and phenotypic variability of water absorption and cooking time in dry bean, periodical updates of these prediction models with more samples and new bean accessions, as well as testing other multivariate prediction methods, are needed for further improving model robustness and generalization.

F.A. Mendoza, J.A. Wiesinger, and J.D. Kelly, Dep. of Plant, Soil and Microbial Sciences, Michigan State Univ., East Lansing, MI 48824; R. Lu and K.A. Cichy, USDA-ARS, Sugarbeet and Bean Research Unit, Michigan State Univ., East Lansing, MI 48824; S. Nchimbi-Msolla, Dep. of Crop Science and Production, Sokoine Univ. of Agriculture, Chuo Kikuu, Morogoro, MI 4882, Tanzania; P.N. Miklas, USDA-ARS, Grain Legume Genetics and Physiology Research Unit, Washington State Univ., Prosser, WA 99350.

Copyright © American Society of Agronomy and Crop Science Society of America. 5585 Guilford Rd., Madison, WI 53711 USA. This is an open access article distributed under the terms of the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)
Plant Phenome J. 1:180001 (2018)
doi:10.2135/tppj2018.01.0001

Received 10 Jan. 2018.
Accepted 20 July 2018.
Supplemental material online.

*Corresponding authors (fmendoza1830@gmail.com; Karen.Cichy@ars.usda.gov).

Dry beans (*Phaseolus vulgaris* L.) are a nutrient-dense food and a dietary staple in many regions of Africa and Latin America. One factor that limits greater utilization of beans is their long cooking times (Mishili et al., 2011). The cooking time of dry beans affects market price, processing costs, shelf life, and consumption patterns (Jones, 1999; Zamindar et al., 2013). Cooking time is influenced by multiple factors including genotypic differences in seed chemistry, growing environment, post-harvest handling, seed storage, and cooking conditions (i.e., soaking before cooking) (Schoeninger et al., 2014; Coelho et al., 2007). Because there exists a wide range of cooking times for dry beans that are sold in the market or from a supermarket shelf, it is difficult for the processor to provide exact cooking time recommendations and cooking instructions to the consumer. Genetic improvement can provide an effective tool for producing dry beans with more consistent or narrow cooking time ranges (Cichy et al., 2015a). One of the major limitations to incorporating cooking time in a bean breeding program is the difficulty of phenotyping this trait due to low throughput and high labor cost. A nondestructive high-throughput phenotyping method would increase the

Abbreviations: 1Der, first derivative; 2Der, second derivative; CWT, continuous wavelength transform; PLS, partial least squares; R_{cal} , average correlation coefficient of calibration; R_{pred} , average correlation coefficient of prediction; SEC, average standard error of calibration; SEP, average standard error of prediction; SFS, sequential forward selection; Vis/NIRS, visible and near-infrared spectroscopic.

likelihood of breeders incorporating cooking time in their breeding programs.

Currently, the cooking time of dry beans is experimentally measured using an automated Mattson pin drop cooker (Wang and Daun, 2005). With this method, cooking time is defined as the time it takes for 80% of the seeds to be completely pierced by a 2-mm stainless steel pin during heating in boiling water. While the Mattson cooker provides an excellent reference testing method for evaluating the cooking time of beans with high test reliability, this procedure is slow, labor intense, and expensive for practical applications in breeding programs and by industry. There is an urgent need for a fast, reliable, and nondestructive method for screening cooking time properties on intact dry bean seeds.

Automated nondestructive optical techniques have shown great potential to predict various quality traits in beans. Cichy et al. (2015b) assessed cooking time for a panel of 206 *P. vulgaris* accessions to identify genomic regions influencing this trait based on a genome-wide association analysis. A 5.5-fold variation for cooking time was measured among the panel of >200 *P. vulgaris* accessions. They also evaluated the suitability of visible and near-infrared (Vis/NIR) spectroscopic data gathered from intact dry seeds to predict cooking time. The Vis/NIR spectroscopic scanning of whole raw seed explained 68% of the phenotypic variation for cooking time. With further improvement, it may be possible to use Vis/NIR spectroscopy to nondestructively predict cooking time as part of a breeding program.

Recently, Mendoza et al. (2018a) showed that the texture and hardness of canned black beans can be predicted from intact dry seeds by using Vis/NIR spectroscopy ($R_{\text{pred}} = 0.886$ and SEP = 40 kg kg⁻¹) and hyperspectral imaging ($R_{\text{pred}} = 0.844$ and SEP = 46 kg kg⁻¹) techniques. Since the end-use texture quality of dry beans is directly related to their hydration and cooking properties, these sensing technologies could help predict the cooking time properties of different bean seed genotypes, thus rapidly increasing the efficiency of screening. The objective of this study was therefore to evaluate the performance of hyperspectral imaging technology for predicting water uptake and cooking times for different dry bean market types grown across multiple environments, either soaked or unsoaked. Image processing programs were implemented to extract image features, which were then used for the development of partial least squares regression models, which were tested and compared for different spectral preprocessing methods.

Materials and Methods

Experimental Design, Field Design, and Storage Conditions

Fourteen *P. vulgaris* genotypes with wide ranges in cooking times, representing five dry bean market classes (yellow, cranberry, red mottled, light red kidney, and brown) of agricultural importance to eastern Africa, southern Africa, the Caribbean, and North America were evaluated across multiple environments. This group of genotypes represents a panel previously studied for cooking time in freshly harvested and stored dry beans (Wiesinger et al., 2016).

This panel was chosen to represent fast, moderate, and slow cooking genotypes discovered from screening >200 accession entries belonging to the Andean Diversity Panel (Cichy et al., 2015a), a genetic resource for pre-breeding and germplasm enhancement of dry beans of Andean origin (<http://arsfifbean.uprm.edu/bean/>). In 2014, the 14 dry bean genotypes were grown in three contrasting field sites located in Tanzania: Arusha (Selian Agriculture Institute), Mbeya (Uyole Research Station), and the heat/drought stress site in Morogoro (Sokoine University of Agriculture).

The same 14 genotypes were also grown in Othello, WA (Washington State University Research Station) in 2014 under normal field conditions, as well as under terminal drought stress conditions, in side-by-side trials. In 2015, 125 genotypes from the Andean Diversity Panel and breeding lines were grown at the Montcalm Research Farm located near Entrican, MI, under normal field conditions. This set of genotypes represented nine different dry bean market classes that included yellow, cranberry, red mottled, dark/light red kidney, brown, pink mottled, purple mottled, and small/large red beans. All genotypes were planted in a randomized complete block design with two field replicates in 2014 and 2015. Individual plots consisted of two rows 4.75 m long with 0.5-m spacing between rows. No fertilizer was applied to the plots, and recommended practices were followed for weed and insect control. Seeds were harvested at maturity by pulling the entire field plot and threshing by hand (Tanzania), or by threshing with a small-plot combine (Wintersteiger). Immediately on arrival at Michigan State University, 200 bean seeds from each field replicate were hand selected after elimination of discolored, wrinkled, or damaged seeds. Seeds were placed in dark storage under refrigerated conditions (4°C, 65% relative humidity) at standard atmospheric pressure for 3 wk prior to imaging analysis. Samples from 2014 were initially used for model building and prediction of water uptake and cooking time of soaked and unsoaked dry seeds. Bean samples from 2015 were used as a totally independent set for validating the performance of the cooking time prediction model built in 2014 for soaked beans.

Hydration Properties and Cooking Time Determination

Before image scanning, seeds were stored in a controlled-atmosphere cabinet (Storage Control Systems) at room temperature with a saturated sodium nitrite salt solution (63% relative humidity) until they reached the equilibrium moisture content of 10 to 12% (Moisture Check Plus, Deere and Company). To measure hydration properties, preweighed, moisture-equilibrated bean seeds were soaked in distilled water (1:8 raw seed weight/water weight) for 12 h at room temperature. The soaking water was then drained and superfluous water removed before reweighing. The water uptake percentage for each field replicate was calculated according to

$$\text{Water}_{\text{uptake}} = \left(\frac{\text{Seed weight after soak} - \text{Seed weight before soak}}{\text{Seed weight before soak}} \right) 100$$

Prior to cooking, the moisture-equilibrated bean seeds were either left unsoaked or soaked in distilled water (1:8 w/w) for 12 h at room temperature. Cooking time was determined using a pin drop Mattson cooking device (Customized Machining and Hydraulics Co.). The cooker utilizes a rack of 25 stainless steel piercing rods (70 g, 2-mm diameter) placed in contact with the middle surface of an unsoaked or presoaked bean, which is fitted into a 4-L stainless steel beaker containing 1.8 L of boiling distilled water heated over a Waring SB30 portable burner. Cooking time was measured as the number of minutes required for 80% of the 25 piercing rods to pass completely through each seed under a low-steady boil at 100°C (Wang and Daun, 2005). Figure 1 presents the experimental procedure for evaluating water uptake and cooking time from dry bean seeds using the hyperspectral imaging technique.

Hyperspectral Imaging

Reflectance images of dry bean samples were acquired for the wavelength range of 400 to 1000 nm. The imaging system consisted of a high-performance 12-bit digital CCD camera (SensiCam QE, The Cooke Corp.) with a 1376 by 1040 pixel array, an imaging spectrograph (ImSpector V10E OEM, Spectral Imaging Ltd.) with a 30- μ m entrance slit, a 23-mm fixed focal lens (Xenoplan, Schneider Optics), and two linear light sources with the 150-W tungsten halogen lamps for reflectance. Between 30 and 40 bean seeds (depending on the seed size) were placed on a petri dish, and the acquisition of hyperspectral images was performed using 2×2 binning operations and an exposure time of 80 ms. The acquired hyperspectral images had the spatial resolution of 1 mm/pixel along the scan line and a spectral resolution of 1.24 nm (Fig. 2A).

Image preprocessing was first performed on the original hyperspectral images for flat-field correction by subtraction of the dark current image and then division of the difference between the

reference image and dark current image. After flat-field correction, an image mask was created to remove the background based on five selected spectral images (at wavelengths of 823.9, 836.9, 848.6, 860.3, and 872.0 nm). For this, the intensity images were all individually preprocessed and converted to binary images using a suitable global threshold value for each case and then combined to build the mask. The mask was applied to each wavelength image, resulting in a stack of masked images with pixels covering only the bean seeds (true spectral intensity) on a black background. This procedure allows accurate discrimination of individual beans from the background. The segmentation procedure and its resultant spectrum for a bean seed are depicted in Fig. 2. For analysis, mean spectra were calculated by averaging the intensity pixels corresponding to all bean seeds at each wavelength image, yielding a spectrum of 520 data points for each scanned sample. The automatic segmentation and mean spectra computation algorithms were developed in MATLAB 7.5.0 (The MathWorks).

Model Development for Prediction

Prediction models were developed using the 2014 data set in three steps: spectral data preprocessing, selection of the best 100 variables (or wavelengths), and modeling.

Spectral Data Preprocessing

To ensure the most appropriate analysis strategy, four different spectral preprocessing techniques were tested, which include smoothing, first derivative (1Der), second derivative (2Der), and the multi-resolution wavelet transform method (CWT) based on continuous one-dimensional decomposition using the symlet wavelet at 64-scales. The resultant vector after applying continuous wavelet transform represents the root mean square of the transformed data at each single wavelet. The details of calculating these spectral features can be found in Mendoza et al. (2012). Additionally all combinations of two-band ratios for each preprocessed data set (i.e., smooth, 1Der, 2Der, and CWT) were also obtained and used for the optimal wavebands selection.

Selection of the Best Hundred Variables

For removing redundant spectral information and reducing computational time, the sequential forward selection (SFS) method was used for selecting the best 100 variables from the smooth, 1Der, 2Der and CWT data sets, as well as from the resultant vector of features after calculating the two-band ratios for each preprocessed data set.

Model Building

Quantitative calibration models were developed using partial least squares regression. In this step, the optimum numbers of latent variables (from the 100 selected in the previous step) were selected on the basis of minimizing the standard error of cross-validation, which was calculated

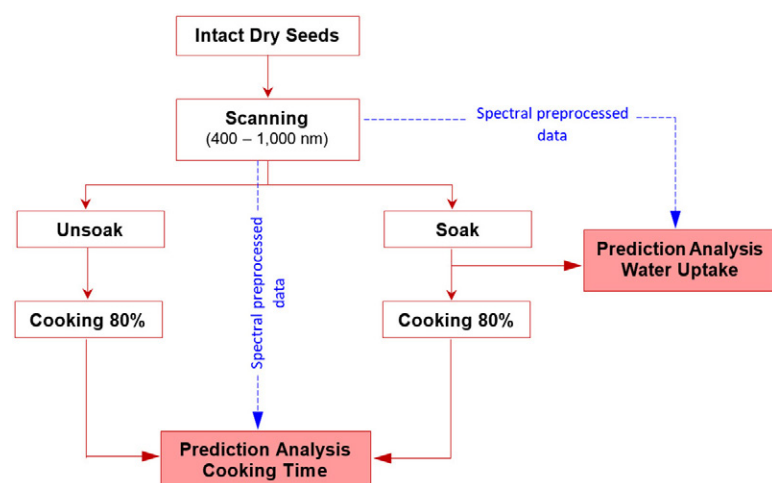


Fig. 1. Experimental procedure for evaluating water uptake after soaking and the determination of cooking time from dry bean seeds using hyperspectral imaging technology. Cooking time is defined as the number of minutes for 80% of the seeds to be completely pierced by a 70-g stainless steel rod during heating in boiling water.

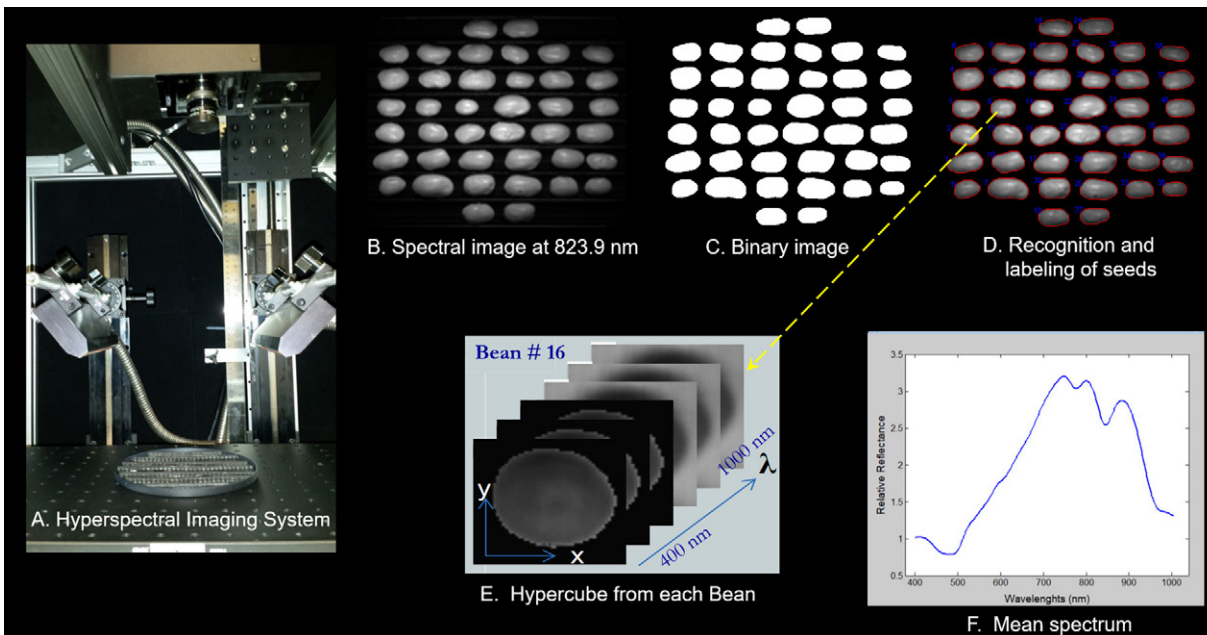


Fig. 2. (A) Hyperspectral imaging system developed by the USDA–ARS postharvest engineering laboratory at Michigan State University, East Lansing, to acquire reflectance images in the wavelength range of 400 to 1000 nm. Image processing steps for: (B) flat-field correction, (C) binarization (mask), (D) recognition of seed, (E) segmentation of individual seeds along the 520 image wavelengths, and (F) typical mean spectrum for an individual seed (Sample 16).

with the leave-one-out method. Partial least squares (PLS) regression calibration models were developed in MATLAB 7.5.0 (The MathWorks) with the PLS Toolbox (Eigenvector Research). Once the calibration model was established, it was then used to predict a set of samples from the same 2014 lot. For this, samples were divided into the calibration (75%) and independent validation or prediction (25%) sets. Since the calibration and prediction performance may vary depending on how the calibration and prediction samples are actually selected, calibrations and predictions were run four separate times. After calibration and prediction were conducted for the first time, a second 25% set of samples was removed from the original calibration set of samples and was used for prediction in the second run. The procedure was repeated until all samples had been removed for prediction once. Average values for the number of latent variables or factors, correlation coefficient, and standard error for the calibration and prediction data sets (i.e., R_{cal} and R_{pred} , SEC and SEP, respectively) were calculated to test the performance of the prediction models.

Finally, a model validation test for predicting the cooking time of soaked seeds was performed using a totally independent set of samples harvested in 2015. In this case, the hyperspectral images and their computed spectral features were preprocessed and analyzed based on the two-band ratios procedure for CWT data, as was done for the calibration model in 2014. Then, after extracting the same best 100 two-band ratios defined for the PLS calibration model built in 2014, the new data were tested on this original model, as well as on updated PLS models after adding 5, 10, 15, 20, 25, and 30 samples selected at random along the cooking time distribution of the new data. The reported prediction results for the independent set of samples were obtained from the average of 10 runs.

Results and Discussion

Characterization of Different Bean Types for Water Uptake and Cooking Time

Significant differences (p value < 0.05) were observed among the bean genotypes for water uptake and cooking time, as shown in Fig. 3 and 4. Water uptake after 12 h of soaking ranged from 86.8 to 118.9%, with the average water absorption of 103.4% and standard deviation of 6.4%. Water uptake was normally distributed among the 135 samples measured in this study for 2014 (Fig. 3). Genetic diversity is a prerequisite for the conventional genetic improvement of a crop. The rational use and analysis of germplasm collections requires an accurate knowledge of their phenotypic characteristics and variability. Moreover, in evaluating biological materials, the number of samples (size and/or variability) and their distribution directly affect the robustness and accuracy of prediction models (Mendoza et al., 2014). Therefore, we characterized the variability and shape distribution of water uptake and cooking time to develop suitable statistical models.

The average cooking times were 55.1 ± 28.7 min (range 19.9–160.1 min) for soaked and 114.2 ± 38.4 min (range 80.1–396.3 min) for unsoaked bean seeds for 2014. Skews toward longer times for both distributions resulted in large standard deviations, with a large tail for the soaked beans (Fig. 4). The histogram distributions for both cooking conditions (soaked and unsoaked) are well described by a lognormal fit (p value > 0.05), although the two histogram distributions were significantly different (p value < 0.05) as determined by a non-parametric Kolmogorov–Smirnov test. Unsoaked cooking times were correlated with soaked cooking times ($R = 0.67$, p value < 0.05). While the correlation is relatively

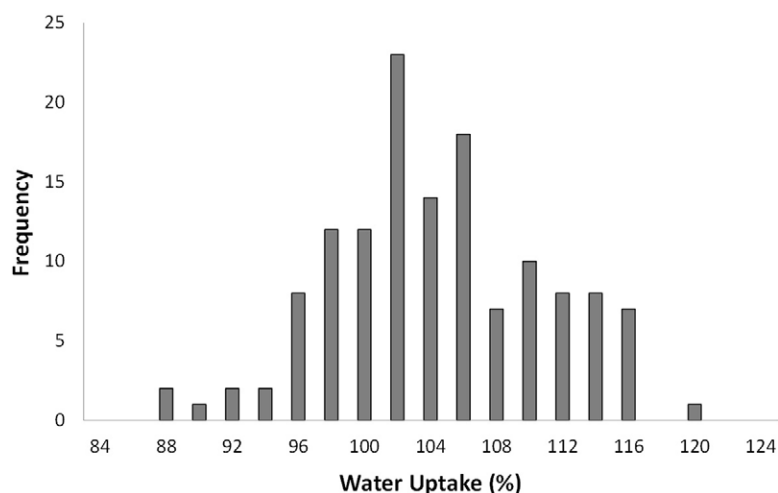


Fig. 3. Histogram distribution for the water uptake of dry bean seeds after soaking for 12 h. Samples represent 14 dry bean accession entries from five market classes that were grown in five contrasting environments during the course of 2 yr ($n = 135$).

high, there is some indication from these data that it is worthwhile to evaluate unsoaked cooking times during cultivar development, especially if the beans are intended for consumption by consumers who do not soak beans prior to cooking. It should also be noted that the histograms for soaked and unsoaked cooking times overlapped between 83 and 163 min. There are some bean types that can be cooked without soaking in a reasonable period of time (<90 min). While water uptake has been used as a predictor of cooking time in beans, our results indicate that water uptake after soaking for 12 h is a weak indicator of cooking time ($R = 0.67$).

The moderate correlation between soaked and unsoaked cooking times is indicative of multiple genetic mechanisms controlling these traits. Hydration or imbibition (water diffusivity) of seeds is a complex function of the microstructure, chemical composition, moisture content, and temperature of the beans. Microstructural features, such as seed coat thickness, seed coat porosity, micropyle diameter, and hilum physiology tend to influence sorption rates

in dry beans. Bean seed coat chemical composition and structural architecture may also affect the imbibition rate (Vertucci, 1989; Souza and Marcos-Filho, 2001). During the soaking process, the concentration of water at the outer coat increases exponentially with time during wetting and remains constant after saturation. First, the seed coat provides a physical barrier to the cotyledons, especially if impermeable, such as in non-hydrating seeds (also known as “hard seed defect”), which is inferred by both the structure and composition of the seed coat. Water penetration through the seed coat can occur via the micropyle or hilum (especially when seed coats have a waxy surface) or through the seed coat itself. Because of the high fiber content, the seed coat has a high water-holding capacity. Once water reaches the cotyledons, the bean seed starts to imbibe and swell until a plateau (maximum hydration) is reached (Wood, 2017). As the cooking process continues, the pectic substances decompose and the connection between the cotyledon cells weakens and the shearing strength decreases. Therefore, the seed coat and microstructure of seeds may be responsible for facilitating a rapid softening of seeds during soaking (Taiwo et al., 1998).

Prediction Analysis

Partial least squares regression models were built by selecting the best 100 features according to the forward selection method from preprocessed spectral data using smoothing, 1Der, 2Der, and CWT, as well as from the two-band ratios from each preprocessed data set. Due to the long-tail distributions for cooking time in both soaked and unsoaked 2014 data sets (Fig. 4) with very few samples toward longer times, the prediction analysis was performed only in the range of cooking times of 19.9 to 95.5 min (120 samples) for soaked beans and 80.1 to 147.0 min (127 samples) for unsoaked beans.

Table 1 presents the calibration and prediction results for water uptake using the best 100 single wavelengths from each pre-processing method and those from their corresponding best 100

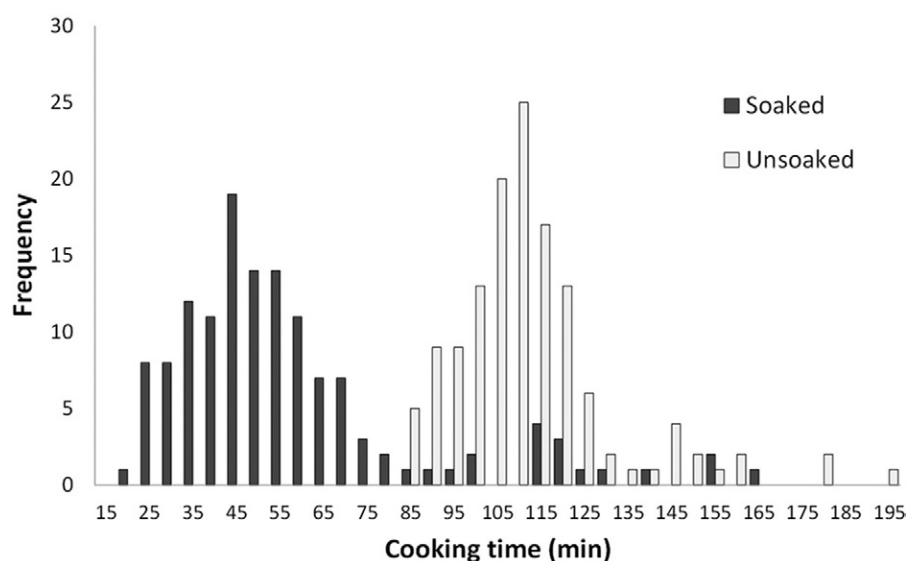


Fig. 4. Histogram distributions for cooking time on a panel of soaked and not soaked dry beans. Samples represent 14 dry bean accession entries from five market classes that were grown in five contrasting environments during the course of 2 yr ($n = 135$).

Table 1. Calibration and prediction results of water uptake (%) after 12-h soaking using hyperspectral imaging on a panel of intact dry beans, including the average number of features required for the partial least square model after optimization (Avg. feature), the average correlation coefficients of calibration and prediction over four calculations (R_{cal} and R_{pred} , respectively), and the average standard error for calibration and prediction over four calculations (SEC and SEP, respectively).

Parameter†	Best single wavelengths					Best ratios				
	Avg. features	R_{cal}	SEC	R_{pred}	SEP	Avg. features	R_{cal}	SEC	R_{pred}	SEP
Smooth	10	0.650	4.6	0.370	6.0 A‡	8	0.719	4.3	0.577	5.1 D
1Der	9	0.707	4.3	0.437	5.9 B	11	0.767	3.9	0.495	5.8 B
2Der	8	0.716	4.2	0.466	5.8 B	10	0.832	3.0	0.584	5.2 D
CWT	6	0.601	4.8	0.309	6.4 C	16	0.965	1.6	0.786	4.4 E

† Smooth, 1Der, 2Der, and CWT: Smoothing, first derivative, second derivative, and continuous wavelet transform decomposition, respectively, combined with the two-band ratios preprocessing method.

‡ SEP values followed by the same letter in a rows or column are not significantly different ($p \geq 0.05$).

ratios. Overall the two-band ratios approach gave significantly better predictions than the best single wavelengths approach (p value < 0.05). The best single wavelengths showed low correlation coefficients of prediction (R_{pred}) ranging from 0.309 to 0.466 and large standard errors of prediction (SEP) ranging from 5.8 to 6.4%, with the best results for the 1Der and 2Der preprocessing methods. To the contrary, the best ratios approach showed the lowest performance for 1Der, followed by the smooth and 2Der preprocessing methods, using 11, 8, and 10 latent variables for model building, respectively. The CWT preprocessing method yielded the highest performance with a R_{pred} value of 0.786 and SEP of 4.4%, although in this case, 16 latent variables were needed for model building. In terms of SEP, CWT improved the prediction accuracy of the model by at least 13.7% compared with the smoothing preprocessing method, which had the second best results. A previous study of predicting water uptake using PLS regression models for the Vis/NIR data of intact dry beans produced in three consecutive harvest seasons (2010–2012) showed R_{pred} values of 0.733 and 0.758 for 1Der and CWT spectral preprocessing methods, respectively. The SEP value was 11% in both cases (Mendoza et al., 2014). Figure 5 is a graphic display of the performance of the two-band ratios with CWT preprocessing for water uptake. The calibration and prediction plots for water uptake exhibit good linear relationships between the actual and predicted values, and the two regression lines overlap very well within the scatterplot.

Table 2 summarizes the calibration and prediction results for the cooking time of soaked and unsoaked dry beans using the best 100 ratios with each preprocessing spectral method. Results for the best single wavelengths were not included because, in previous tests, their prediction models showed significantly lower performances than those for the best ratios (p value < 0.05). Improvements in prediction for the computation of ratios were expected because the two-band ratios could remove the effect of non-uniform illumination and also minimize the effect of seed curvature and seed size in the spectral scanning. Predicted cooking times for unsoaked beans were less accurate and had larger scatter ($R_{pred} = 0.513$ – 0.708 and $SEP = 10.6$ – 11.9 min) than the predictions of cooking time for soaked beans ($R_{pred} = 0.668$ – 0.886 and $SEP = 7.9$ – 11.0 min).

Similar to the water uptake analysis, applying the two-band ratios approach to the CWT data was advantageous for predicting the cooking time for both soaked and unsoaked beans, as indicated by the higher R_{cal} and R_{pred} , and lower SEC and SEP, compared with the smooth, 1Der, and 2Der methods, although larger numbers of variables were needed for building their PLS regression models. Comparing the SEP results for CWT and 2Der (which was the second best prediction for both unsoaked and soaked beans), CWT improved the cooking time prediction accuracy by 2.8% for unsoaked bean seeds and 13.7% for soaked bean seeds. Furthermore, the prediction accuracies of CWT models for soaked beans were at least 25.5% better than those for unsoaked seeds.

Wood (2017) stated that when samples are cooked without presoaking, hydration occurs simultaneously with cooking (in one hydrothermal step). In contrast, presoaking seeds prior to cooking separates hydration (Step 1) and cooking (Step 2), thus shortening the cooking time and reducing the variability among samples. The cooking time results for unsoaked and soaked beans presented in Table 2 confirms this statement. For better visualization of PLS

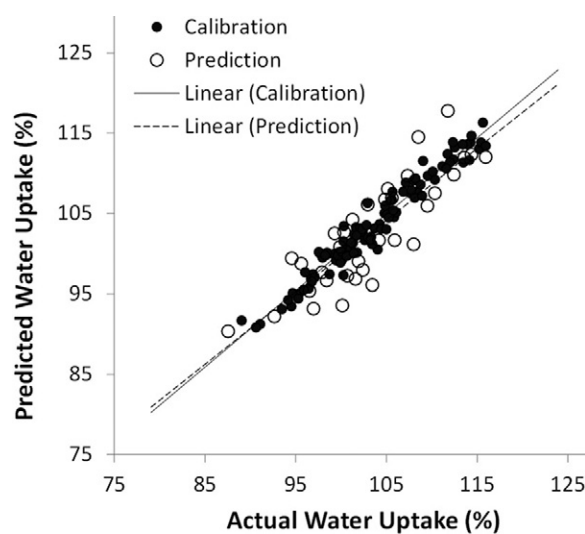


Fig. 5. Actual and predicted water uptake for a panel of intact dry seeds based on hyperspectral imaging data using the combination of continuous wavelet transform decomposition and the two-band ratios preprocessing methods ($R_{pred} = 0.786$ and $SEP = 4.4\%$).

Table 2. Calibration and prediction results of cooking time (min) using hyperspectral imaging on a panel of intact dry beans for soaked and unsoaked seeds, including the average number of features required for the partial least square model after optimization (Avg. feature), the average correlation coefficients of calibration and prediction over four calculations (R_{cal} and R_{pred} , respectively), and the average standard error for calibration and prediction over four calculations (SEC and SEP, respectively).

Parameter†	Unsoaked (80.1–147.0 min, 127 samples)					Soaked (19.9–95.5 min, 120 samples)				
	Avg. features	R_{cal}	SEC	R_{pred}	SEP	Avg. features	R_{cal}	SEC	R_{pred}	SEP
Smooth	7	0.711	9.6	0.573	11.4 A‡	7	0.783	9.7	0.694	11.0 E
1Der	5	0.652	10.4	0.513	11.9 B	7	0.769	10.0	0.668	10.8 C
2Der	8	0.765	8.8	0.637	10.9 C	14	0.946	4.2	0.739	9.8 F
CWT	10	0.945	4.5	0.708	10.6 D	14	0.983	2.9	0.886	7.9 G

† Smooth, 1Der, 2Der, and CWT: Smoothing, first derivative, second derivative, and continuous wavelet transform decomposition, respectively, combined with the two-band ratios preprocessing method.

‡ SEP values followed by the same letter in a rows or column are not significantly different ($p \geq 0.05$).

regression model performance for unsoaked and soaked dry seeds, Fig. 6A and 6B illustrate the plots of cooking time predictions for unsoaked and soaked beans, respectively, using the best 100 ratios obtained from the CWT preprocessing method. The two fitted regression lines for the calibration and prediction data sets overlap well across the entire ranges. These plots again confirm that better predictions of cooking time ($R_{pred} = 0.886$ and $SEP = 7.9$ min) were obtained for soaked beans than for unsoaked beans ($R_{pred} = 0.708$ and $SEP = 10.6$ min).

Prediction of Independent Samples

The PLS regression model from 2014 for soaked beans that was built with the best 100 ratios from the CWT preprocessing method was tested again using a new data set from 2015 ranging from 20.5 to 98.1 min (110 samples). Considering the inherent variability of samples between two harvest seasons, a model updating procedure with 5, 10, 15, 20, 25, and 30 new samples taken at random was also implemented and tested. Table 3 shows the range and average SEP after running the model updating procedure 10 times, and Fig. 7 plots the cooking time predictions for soaked beans from 2015 after updating the 2014 calibration model with 20 new samples from 2015. As could be expected, the original 2014 calibration model (i.e., without updating) produced large standard errors of prediction ($SEP = 18.4$ min) and variations (from 14.3 to 19.9 min in 10 runs) predicting the 2015 data set. However, adding to the model five new samples and removing the same number consecutively, the updated calibration model significantly improved the prediction performance until 20 new samples were added (Table 3), reaching cooking time predictions of $R_{pred} = 0.872$ and $SEP = 3.7$ min. Similarly, as was observed for the 2014 lot, the two fitted regression lines for the calibration and prediction data sets overlap well across the entire ranges (Fig. 7).

Overall, considering the fact that intact dry seeds were used to predict the cooking time, these results reveal promising opportunities for predicting the cooking time of either soaked or unsoaked beans using hyperspectral imaging technology. Moreover, hyperspectral imaging technology showed consistent results that successfully predicted the water uptake and cooking

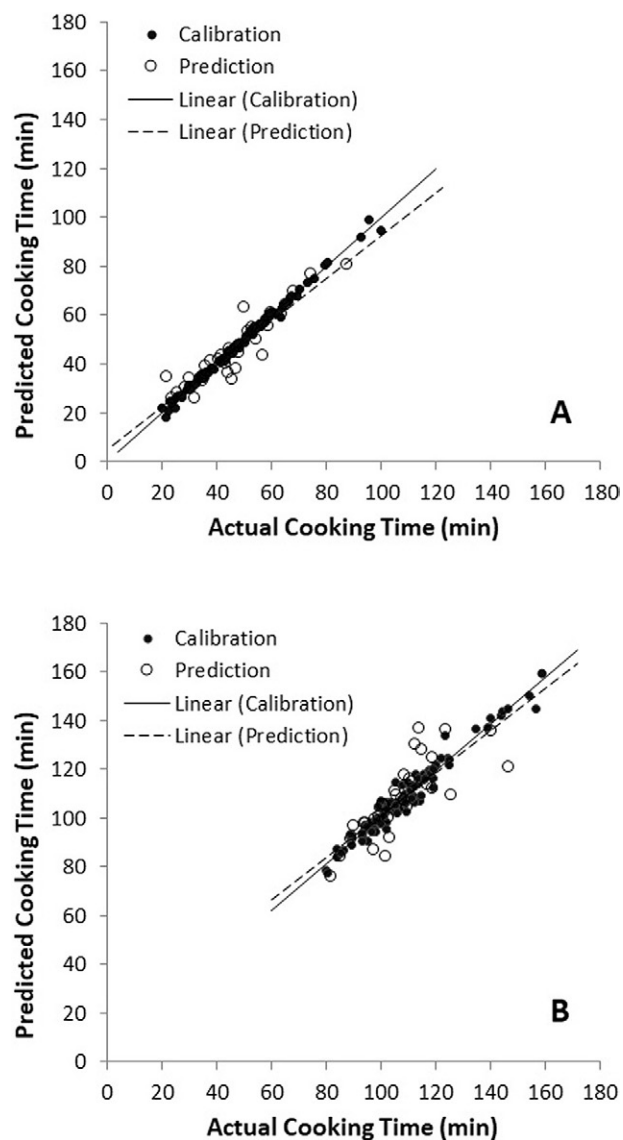


Fig. 6. Actual and predicted cooking time for a panel of intact dry seeds based on hyperspectral imaging data using the combination of continuous wavelet transform decomposition and the two-band ratios preprocessing methods: (A) soaked beans ($R_{pred} = 0.886$ and $SEP = 7.9$ min) and (B) unsoaked beans ($R_{pred} = 0.708$ and $SEP = 10.6$ min).

Table 3. Sample numbers from 2014 and 2015 for model updating and their resultant standard error of prediction (SEP) for cross-year evaluation.

Calibration (19.9–95.5 min)		Prediction (20.5–98.1 min)	SEP		
2014 samples†	2015 samples added‡	2015 samples 2015§	Min.	Max.	Avg.¶
120	0	80	14.3	19.9	16.4
115	5	80	5.2	9.5	7.2
110	10	80	4.6	5.8	5.2
105	15	80	3.6	5.7	4.6
100	20	80	3.2	4.3	3.7
95	25	80	2.9	4.4	3.7
90	30	80	2.9	4.3	3.6

† Samples from 2014 selected at random along the data distribution.

‡ Samples (5, 10, 15, 20, 25, and 30) from 2015 selected at random along the data distribution for model updating.

§ Samples selected at random from 2015 data.

¶ Average SEP from 10 runs.

time of soaked and unsoaked beans for a broad set of cultivars and landraces from five different market classes, with different cooking properties, and grown under four diverse environments including under well-watered and drought conditions in one of the environments. Because of the large variability in bean seed types and environmental conditions represented in this data set, in addition to the phenotypic complexity of microstructural and physicochemical changes during soaking and cooking among the different seed types, successful use of the method is assumed to be achieved by using a carefully selected, well characterized set of germplasm to produce robust calibration models (Cichy et al., 2015b; Wiesinger et al., 2016).

Finally, hyperspectral imaging technology is suitable for use in the breeding process as well as for facilitating processing decisions by the industry. This particular study with two independent panels of dry bean seeds—120 samples for calibration from 2014 and 110 samples for independent validation from 2015—showed that updating the original calibration model with 16.7% of new data can estimate with decent accuracy the cooking time of soaked dry bean seeds. The combination of CWT and the two-band ratios preprocessing method produced an important improvement in the prediction of cooking time compared with the traditional preprocessing methods for model building, and hence this procedure should be further tested in other studies with more data and for other bean traits. Phenotypic evaluations of multiple bean traits can be conducted from a single scan simultaneously and quickly, after appropriate chemometric and model updating procedures have been built and optimized.

Conclusions

Hyperspectral imaging technology has great potential for estimating cooking time properties of both soaked and unsoaked

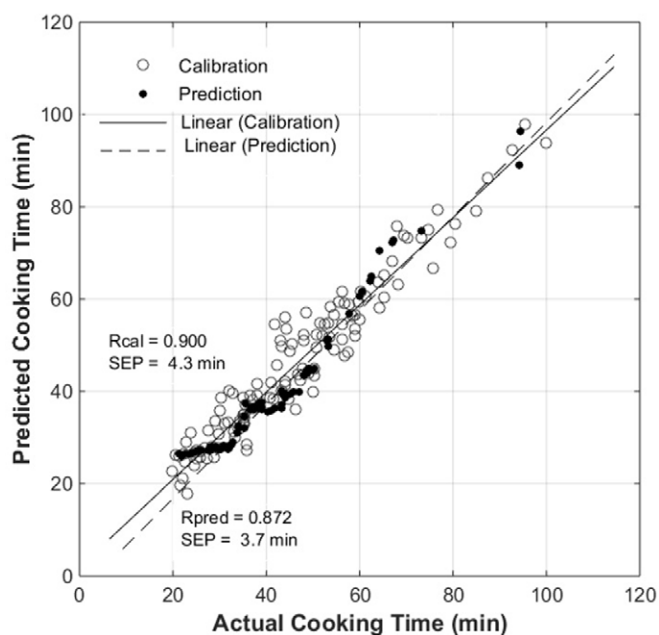


Fig. 7. Actual and predicted cooking time for soaked beans from 2015 after updating the 2014 calibration model with 20 new samples from 2015. The new hyperspectral imaging data were analyzed using the same procedure used in 2014.

dry bean seeds. Applying the two-band ratios method to the CWT data with a single feature selection method PLS regression model provided the best estimate of the cooking time of dry, unprocessed bean seeds. More accurate and consistent results were found for seeds cooked after soaking ($R_{\text{pred}} = 0.886$ and $\text{SEP} = 7.9$ min) than for unsoaked beans ($R_{\text{pred}} = 0.708$ and $\text{SEP} = 10.6$ min). Additionally, a totally independent panel of similar dry bean seeds confirmed the suitability of the technique for predicting the cooking time of soaked beans after updating the PLS calibration model from 2015 with 20 of these new samples. This cross-year validation showed predictions of $R_{\text{pred}} = 0.872$ and $\text{SEP} = 3.7$ min. Although this predictive model is based on a wide range of production environments and seed types with large variations in cooking properties, further validation across an even broader set of environments and genotypes is warranted, and slight changes or periodical model updates that improve robustness of the model are expected as these additional data are processed.

Data Availability

Due to the large size of the raw hyperspectral image data, only the MATLAB codes and relevant processed data files from 2014 and 2015 have been uploaded in the Dryad Digital Repository (Mendoza et al. 2018b). Original hyperspectral images are available upon request from the corresponding authors.

Acknowledgments

This research was carried out as a part of the USDA–ARS in-house project 3635-21430-009-00D, Improved Quality in Dry Bean Using Genetic and Molecular Approaches. Funding was also received through MSU Project GREEN and the USAID Legume Innovation Lab (Cooperative Agreement no. EDH-A-00-07-00005-00). We thank Mr. Scott Shaw and Evan Wright for their technical support in testing bean samples. The contents of this publication do not necessarily reflect the views or policies of

the USDA nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

- Cichy, K.A., T.G. Porch, J.S. Beaver, P. Cregan, D. Fourie, R.P. Glahn, et al. 2015a. A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Crop Sci.* 55:2149–2160. doi:10.2135/cropsci2014.09.0653
- Cichy, K.A., J.A. Wiesinger, and F.A. Mendoza. 2015b. Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 128:1555–1567. doi:10.1007/s00122-015-2531-z
- Coelho, C.M.M., C. de Mattos Bellato, J.C.P. Santos, E.M.M. Ortega, and S.M. Tsai. 2007. Effect of phytate and storage conditions on the development of the 'hard-to-cook' phenomenon in common beans. *J. Sci. Food Agric.* 87:1237–1243. doi:10.1002/jsfa.2822
- Jones, A.L. 1999. Phaseolus bean: Post-harvest operations. FAO, Rome.
- Mendoza, F.A., R. Lu, and H. Cen. 2012. Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content. *Postharvest Biol. Technol.* 73:89–98. doi:10.1016/j.postharvbio.2012.05.012
- Mendoza, F.A., K.A. Cichy, R. Lu, and J.D. Kelly. 2014. Evaluation of canning quality traits in black beans (*Phaseolus vulgaris* L.) by visible/near-infrared spectroscopy. *Food Bioproc. Technol.* 7:2666–2678. doi:10.1007/s11947-014-1285-y
- Mendoza, F.A., K.A. Cichy, C. Sprague, A. Goffnett, R. Lu, and J.D. Kelly. 2018a. Prediction of canned black bean texture (*Phaseolus vulgaris* L.) from intact dry seeds using visible/near infrared spectroscopy and hyperspectral imaging data. *J. Sci. Food Agric.* 98:283–290. doi:10.1002/jsfa.8469
- Mendoza, F.A., J.A. Wiesinger, R. Lu, S. Nchimbi-Msolla, P.N. Miklas, J.D. Kelly, and K.A. Cichy. 2018b. Data from: Prediction of cooking time for soaked and unsoaked dry beans (*Phaseolus vulgaris* L.) using hyperspectral imaging technology. Dryad Digital Repository. doi:10.5061/dryad.ch4ns27
- Mishili, F.J., A. Temu, J. Fulton, and J. Lowenberg-DeBoer. 2011. Consumer preferences as drivers of the common bean trade in Tanzania: A marketing perspective. *J. Int. Food Agribus. Mark.* 23:110–127. doi:10.1080/08974438.2011.558761
- Schoeninger, V., S.R. Machado Coelho, D. Christ, and S.C. Sampaio. 2014. Processing parameter optimization for obtaining dry beans with reduced cooking time. *LWT—Food Sci. Technol.* 56:49–57. doi:10.1016/j.lwt.2013.11.007
- Souza, F.H., and J. Marcos-Filho. 2001. The seed coat as a modulator of seed–environment relationships in Fabaceae. *Braz. J. Bot.* 24:365–375. doi:10.1590/S0100-84042001000400002
- Taiwo, K.A., C.T. Akanbi, and O.O. Ajibola. 1998. Regression relationships for the soaking and cooking properties of two cowpea varieties. *J. Food Eng.* 37:331–344. doi:10.1016/S0260-8774(98)00034-X
- Vertucci, C.W. 1989. The kinetics of seed imbibition: Controlling factors and relevance to seeding vigor. In: P.C. Stanwood and M.B. McDonald, editors, Seed moisture. *CSSA Spec. Publ.* 14. CSSA, Madison, WI. p. 93–115. doi:10.2135/cssaspecpub14.c6
- Wang, N., and J.K. Daun. 2005. Determination of cooking times of pulses using an automated Mattson cooker apparatus. *J. Sci. Food Agric.* 85:1631–1635. doi:10.1002/jsfa.2134
- Wiesinger, J.A., K.A. Cichy, R.P. Glahn, M.A. Grusak, M.A. Brick, H.J. Thompson, and E. Tako. 2016. Demonstrating a nutritional advantage to the fast-cooking dry bean (*Phaseolus vulgaris* L.). *J. Agric. Food Chem.* 64:8592–8603. doi:10.1021/acs.jafc.6b03100
- Wood, J.A. 2017. Evaluation of cooking time in pulses: A review. *Cereal Chem.* 94:32–48. doi:10.1094/CCHEM-05-16-0127-FI
- Zamindar, N., M.S. Baghekhanda, A. Nasirpour, and M. Sheikhzeinoddin. 2013. Effect of line, soaking and cooking time on water absorption, texture and splitting of red kidney beans. *J. Food Sci. Technol.* 50:108–114. doi:10.1007/s13197-011-0234-2