Accepted Manuscript

Application of Near-infrared Spectroscopy to Predict the Cooking Times of Aged Common Beans (*Phaseolus vulgaris* L.)

Elizabeth N. Wafula, Irene N. Wainaina, Carolien Buvé, Nghia D.T. Nguyen, Peter K. Kinyanjui, Wouter Saeys, Daniel N. Sila and Marc E. Hendrickx

PII: S0260-8774(20)30154-0

DOI: [10.1016/j.jfoodeng.2020.110056](https://doi.org/10.1016/j.jfoodeng.2020.110056)

Reference: JFOE 110056

To appear in: *Journal of Food Engineering*

Please cite this article as: Elizabeth N. Wafula, Irene N. Wainaina, Carolien Buvé, Nghia D.T. Nguyen, Peter K. Kinyanjui, Wouter Saeys, Daniel N. Sila and Marc E. Hendrickx (2020). Application of near-infrared spectroscopy to predict the cooking times of aged common beans (*Phaseolus vulgaris* L.). *Journal of Food Engineering*, 284(110056). <https://doi.org/10.1016/j.jfoodeng.2020.110056>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Application of Near-infrared Spectroscopy to Predict the Cooking Times of Aged Common Beans (*Phaseolus vulgaris* L.)

Elizabeth Nakhungu Wafula^{a,b*}, Irene Njoki Wainaina^{a,b}, Carolien Buvé^a, Nghia-Do-Trong Nguyen^{c,d}, Peter Kahenya Kinyanjui^b, Wouter Saeys^e, Daniel Ndaka Sila^b and Marc Hendrickx^{a**}

Current Affiliations

^aKU Leuven, Department of Microbial and Molecular Systems (M²S), Laboratory of Food Technology, Kasteelpark Arenberg 22, Box 2457, 3001 Leuven, Belgium

^bJomo Kenyatta University of Agriculture and Technology, College of Agriculture and Natural Resources, School of Food and Nutritional Sciences, Department of Food Science and Technology, P.O. Box 62,000-00200, Nairobi, Kenya

^cTon Duc Thang University, Division of Computational Mathematics and Engineering, Institute for Computational Science, P.O. Box 19 Đường Nguyễn Hữu Thọ, Tân Hưng, Quận 7, Hồ Chí Minh 758307, Vietnam

^dTon Duc Thang University, Faculty of Applied Sciences, P.O. Box 19 Đường Nguyễn Hữu Thọ, Tân Hưng, Quận 7, Hồ Chí Minh 758307, Vietnam

^eKU Leuven, Department of Biosystems (BIOSYST), MeBioS, Kasteelpark Arenberg 30, Box 2456, 3001 Leuven, Belgium

*Author to whom correspondence should be addressed during the submission process

Email: elizabethnakhungu.wafula@kuleuven.be

**Author to whom correspondence should be addressed post-publication

Email: marceg.hendrickx@kuleuven.be

Abstract

The cooking time of beans is an important quality indicator which can change considerably during ageing. Therefore, this study investigated the potential of near-infrared spectroscopy to rapidly predict cooking times of aged common beans. Four bean varieties were aged under different storage conditions, resulting in a range of samples for each of these varieties. The cooking kinetics of the aged beans were determined by finger pressing and modelled using logistic regression to obtain the times it took to cook 95% of the beans. The cooking times obtained were predicted from the NIR spectra of milled raw bean samples. This was done using partial least squares regression, after carrying out wavelength selection. Model performance was improved up to an average prediction error of 8 minutes by de-hulling the beans and reducing the number of varieties included. In conclusion, NIR spectroscopy has high potential to predict the cooking times of aged beans.

Key words

Common beans; Ageing; Logistic regression; Cooking time; Near infrared spectroscopy; Partial least squares regression

List of Abbreviations

HTC – Hard-To-Cook

NIR – near-infrared

PLS-R – partial least squares regression

iPLS – interval partial least squares

R^2_C – coefficient of determination of calibration

- 46 R^2_{CV} – coefficient of determination of cross validation
- 47 R^2_P – coefficient of determination of prediction
- 48 RMSEC – root mean square of errors of calibration
- 49 RMSECV – root mean square of errors of cross validation
- 50 RMSEP – root mean square of errors of prediction
- 51 LV –latent variables
- 52 MC – mean centering
- 53 SNV – standard normal variate
- 54 OSC – orthogonal signal correction
- 55 DH – de-hulled beans
- 56 WSC – with seed coat
- 57 CW – Canadian wonder
- 58 RH – red haricot
- 59 RC – rose coco

60 **1. Introduction**

61 Legumes are widely considered important foods as part of diverse, nutritious, healthy and
 62 environmentally sustainable diets (Willett et al., 2019; Castro-Guerrero et al., 2016). This is true
 63 for the majority of lower-income populations in developing regions that highly depend on plant
 64 foods due to the high cost of foods from animal sources (Chávez-Mendoza and Sánchez, 2017;

Castro-Guerrero et al., 2016; Câmara et al., 2013) as well as for high-income populations (Willett et al., 2019). Among the legumes, common beans (*Phaseolus vulgaris* L.) are the most economically significant species and are widely accepted (Broughton et al., 2003). For all these reasons, common beans have great potential to contribute to food and nutrition security now and in the future.

The cooking process (cooking time) is one of the major factors affecting the consumption convenience and general acceptability of common beans (Katungi et al., 2011; Bernal-Lugo et al., 1997). Conventional cooking is usually a two-step procedure of texture degradation which starts with soaking followed by thermal processing, mostly boiling at household or industrial level. During cooking, the rate limiting step of bean softening has been reported to be the solubilization of pectin present in the middle lamella and the cell walls of the cotyledon (Chigwedere et al., 2018). The cooking process of beans has been previously measured subjectively and/or objectively as the percentage of beans that have attained a stipulated endpoint of texture degradation or chemical change (e.g. starch gelatinization), then called cooked. The time it takes to reach the desired endpoint is an important quality indicator for breeders, seed distributors, food industries and consumers (Wood, 2017).

The cooking time of beans depends on variety (Bernal-Lugo et al., 1997; Kinyanjui et al., 2015) and even within the same variety they can differ depending on postharvest storage conditions (Kinyanjui et al., 2017). Cooking times can range from 30-120 minutes when beans are freshly harvested (Garcia et al., 2012) up to 8 hours due to inappropriate storage (Garcia et al., 1998). This has been attributed to the development of the Hard-To-Cook (HTC) textural defect. This defect causes the bean cotyledon to harden and although the bean properly hydrates, bean softening is hindered, resulting in a prolonged cooking time (Shiga and Lajolo, 2006; Shiga et al., 2004).

The mechanism of HTC development in beans has yet to be fully elucidated, nonetheless, several mechanisms have been suggested. The mechanisms depend on enzyme-catalyzed reactions and transport of small molecules largely controlled by the plasticizing effect of water. Therefore, temperature and relative humidity (typically above 25 °C and 65%) during storage (Stanley, 1992) and probably soaking are key factors of HTC development (Chigwedere et al., 2019a). The current overall hypothesis suggests that during adverse storage, membrane disruptions and (lipoxygenase catalyzed) lipid oxidation occur (Hussain et al., 1989). The pectin-cation-phytate theory (Kilmer et al., 1994; Hentges et al., 1991; Jones and Boulter, 1983), involves divalent cations released when phytase hydrolyzes phytates. The cations diffuse from protein bodies into the cell walls where they cross-link with pectin demethoxylated by pectin methylesterase. The lignification (Garcia et al., 1998; Hohlberg and Stanley, 1987) and multiple reactions theories (Aguilera and Rivera, 1992) involve the transport of phenolics to the cell walls resulting in lignification and/or phenolic-pectin cross-links. Besides the abovementioned theories, it has also been suggested that protein denaturation during storage and cooking hinders starch denaturation and thus bean softening (Garcia-Vela and Stanley, 1989).

To date, the approaches used to determine the cooking times of beans and eventually the HTC defect at the user level include finger pressing and texture analysis among others after cooking (Wood, 2017). Both these methods are time-consuming. Additionally, the finger pressing method requires trained personnel to obtain reliable results as it is subjective (Kinyanjui et al., 2015). This calls for further exploration of alternative techniques to address these challenges.

Near infra-red (NIR) spectroscopy is a high throughput analytical tool (Manley, 2014; Osborne, 2000b) that is widely applied in the quality check of raw materials and end-products in food research and industry (Nicolai et al., 2007; Williams, 2004; Osborne, 2000a;). The NIR region of

the electromagnetic spectrum ranges from 780 nm to 2800 nm. Samples are scanned to obtain spectra containing absorption information of functional groups (O-H, C-H, C=O, N-H) and scattering information induced by microstructural heterogeneities. Subsequently, chemometrics is used to correlate the spectra to sample quality attributes obtained using conventional methods. This results in calibration models used to predict quality attributes of new samples (Saeys et al., 2019; Pasquini, 2018; Williams, 2004). NIR spectroscopy has been successfully used to predict the chemical composition, physical characteristics and sensory properties of common beans (Wang, et al., 2014; Plans et al., 2014, 2013, 2012; Pojić et al., 2010).

NIR spectroscopy and hyperspectral imaging have also been successfully used to predict cooking times of Andean Diversity Panel accession covering a wide range of fresh common bean seed types (Mendoza et al., 2018b; Cichy et al., 2015). However, these studies excluded the development of storage-induced HTC. Also, the spectra were acquired for whole beans, which to the best of our knowledge likely put limitations on the vis/NIR approach when using the reflectance mode to evaluate cooking times associated with an interior structure, the cotyledon. Therefore, the aim of this study was to evaluate to what extent NIR spectroscopy can be used to determine the cooking time for aged beans. Hence, the experimental design involved using a limited number of varieties stored under different conditions to generate aged samples with a broad range of cooking times within each variety. As it has been shown that HTC development is cotyledon related, the aged samples were separated into two groups, the first with seed coats and the second as cotyledons only. The samples were milled before the acquisition of spectra. By doing this, we hypothesized that the molecules described in one or more of the mechanisms leading to HTC development could be identified during model development.

2. Materials and methods

The flow of research activities undertaken during the study is schematically presented in **Figure 1**. Kinyanjui et al., (2017) conducted the storage and cooking experiments summarized in **2.1-2.2**.

2.1 Bean samples and storage treatment

Fresh beans of four varieties including Canadian wonder, pinto, red haricot and rose coco were purchased from KARLO, Thika station in Kenya. All the varieties had been planted in the same geographical area during the same season and thus belonged to the same harvest. Sieving of the beans was done to remove unwanted foreign objects. Hand sorting was also performed to remove split, rotten (moldy) and discolored seeds and polishing was done to clean off residual dust particles from the beans. Afterwards, the beans were conditioned at 25 °C and 75% relative humidity for two weeks to equilibrate the moisture content. A reference sample ('0 month storage') of each bean variety was set aside. The remaining beans from each variety were divided into four equal batches and stored under different conditions of temperature and relative humidity in thermostatic chambers, as depicted in **Table 1**. To maintain relative humidity of 75% and 83%, saturated salt solutions of sodium chloride and potassium chloride, respectively, were placed in the storage chambers (Kinyanjui et al., 2017). For each variety 11-12 samples stored under different conditions were included in this study. Herein, a sample referred to beans from one variety, stored at a particular temperature and relative humidity and sampled at a particular storage time. For example, red haricot stored at 25 °C and 75% relative humidity for 4 months is a sample denoted as RH 4/25/75.

2.2 Reference cooking times

There was impetus to employ finger pressing as a technique to determine the cooking kinetics of beans in spite of its somewhat subjective nature. The method's ability to produce reliable results had already been validated using an objective method in an earlier study. The findings indicated that there were no significant differences between the cooking times of beans obtained using finger pressing and cutting tests (Kinyanjui et al., 2015). The cooking kinetics of beans included in this study were determined using the finger pressing method in a previous study (Kinyanjui et al., 2017). The method was as follows.

One hundred seeds from each sample were separately soaked in demineralised water in a ratio of 1:5 (w/v) at ambient conditions for 16 hours. The soaking water was drained from the beans before cooking (Memmert WBU-45, Germany). Ten seeds were taken out after intervals of 30 minutes without interrupting boiling and cooled in a water bath for one minute. Cooked beans were identified by exerting moderate pressure on the beans using the forefinger and thumb. The beans were classified as cooked when the cotyledons disintegrated. The beans were cooked from 0 to 300 minutes. Cooking of beans was done in triplicate by two different persons. The experimental results obtained from that study were expressed as the number of cooked beans versus the number of non-cooked beans as a function of time.

The data analysis approach employed in the aforementioned work was, however, different from that employed in this study. Logistic regression was used to describe the kinetic data by modelling the relationship between time of cooking (explanatory variable) and the percentage of cooked beans (outcome). Logistic regression utilizes a link function to equate the probability that an outcome will occur to a linear equation containing parameters (α and β) that describe the influence of explanatory variable(s). The link function in logistic regression is the logit transform, and it is

the natural logarithm of the odds that an outcome will occur. In this study, the odds that a bean will be cooked is the ratio of the probability that the bean will be cooked to the probability that the bean will be non-cooked after a given time of cooking. The parameters α and β are then estimated (estimates of α and β , respectively) from the data based on the maximum likelihood criterion. Parameter estimation was carried out for each sample separately. The logistic regression model is described by **Equations 1 and 2**. **Equation 1** represents a simple logit transform, with one explanatory variable (T –time of cooking).

$$\text{Logit } (Y) = \text{natural log(odds)} = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta T \quad \text{Equation 1}$$

The antilog of **Equation 1** gives the probability that a bean is cooked after a given time of cooking (**Equation 2**). In this study, sufficient cooking (further indicated as ‘cooking time’) was defined as the time corresponding to a probability that 95% of the beans would be cooked.

$$\pi = \text{Probability } (Y = \text{outcome} \mid T = t, \text{ a specific value of time to cook}) = \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}} \quad \text{Equation 2}$$

In **Equations 1 and 2**, π is the probability that a bean will be cooked, α is the Y intercept, β is the regression coefficient which represents the direction and scale to which a unit change in the time of cooking affects the log odds of the probability that the bean will be cooked. Then, 1 is the conditional operator, $1 \mid T$ is the range of times of cooking (0-300 minutes). While t is one specific time of cooking, which is an independent continuous variable, Y is the percentage of cooked beans, which is a dependent variable (Kutner et al., 2005; Peng et al., 2002). The R Studio (version 3.5.2, 2018) software was used to carry out the statistical analysis.

2.3 Acquisition of spectral data

2.3.1 Sample preparation of raw beans

All samples obtained from the storage experiment were further divided into two separate batches (**Figure 1**). The first batch referred to as “with seed coat” (WSC) beans were milled with their seed coats intact using a micro mill (IKA A10 Basic, IKA(r)-Werke GmbH & Co. KG, Staufen, Germany). The seed coats of the second batch were removed manually before milling and referred to as “de-hulled” (DH) beans. After milling, the samples were passed through a 1 mm mesh size sieve to remove large particles. The bean flours were placed in petri-dishes. They were then transferred to airtight cabinets containing desiccant to condition at room temperature for 10 days. This was done to reduce variation in spectral absorbance brought about by differing moisture contents (Williams, 2004). In addition, these conditions prohibit any further evolution of enzymatic reactions that are responsible for the hard to cook development. The resulting moisture content of the samples was 6% on average (data not shown).

2.3.2 Spectral data acquisition

From each sample, sub-samples were transferred to a transparent glass vial of diameter 2.2 cm and put up to a depth of 1 cm. The sub-samples were scanned over the 3,595 to 12,490 cm^{-1} range using the Multi-Purpose Analyzer Fourier Transformer NIR spectrophotometer (MPA-FT Bruker Optics, Ettlingen, Germany). Scanning was done in reflectance mode using the integrating sphere and each spectrum obtained was an average of 32 scans. The wavelength resolution employed was 16 cm^{-1} . Seven replicates (independent spectra) per sample were obtained by recording one spectrum for each of seven sub-samples (prerequisite sample depth and replicates were determined

during preliminary research). The wavelength range of spectra was converted to nm and reduced to 1000-2600 nm to remove noisy regions based on visual inspection of the spectra (Conzen, 2014).

2.4 Development of NIRS models for predicting cooking times

2.4.1 Selection of samples for calibration and external validation sets

The DH bean samples spectra were separated from those for the WSC bean samples. The spectra of each sample type were further grouped into four-, two- and single-variety data sets (the basis upon which this was done is explained in **section 3.3**). This resulted in seven data sets both for DH and WSC aged bean samples. Thus, 14 PLS-R models were developed in total. The four- and two-variety data sets were separated into the calibration set and an external validation set, herein referred to as the “test set”. Due to the limited number of samples, two and four samples were manually selected based on variety, storage treatment and cooking times for the two- and four-variety data sets, respectively, to form the test set. The samples were pinto 0/25/75 (cooking time: 79 minutes), RC 2/45/75 (cooking time: 150 minutes) and RH 12/25/75 (cooking time: 206 minutes) and CW 6/35/83 (cooking time: 247 minutes). Manual selection was done to represent variability in the calibration set in terms of the aforementioned criteria. To better understand the calibration and test sets, we refer to Appendix B Table B.1. for information on samples included in this study and their cooking times. All samples of the one-variety models were used in calibration development.

2.4.2 PLS-R model development

Partial least squares regression (PLS-R) was used to predict the cooking times (Y) from the acquired NIR spectra (X) (Naes et al., 2002; Saeys et al., 2019). To test the reliability of the models, leave-one-sample out cross validation was used. This cross validation approach was required to

keep the spectral replicates belonging to the same sample, for example RH 0/25/75, together in either the calibration or validation set (Naes et al., 2002). The model complexity in terms of the number of latent variables (LVs) in the PLS-R model was selected based on the RMSECV. When the plot of the RMSECV as a function of the number of LVs showed a clear minimum (which was the case for all PLS-R models developed in this study), this number of LVs was selected. Outliers were detected based on Hotelling T^2 and Q residual values, after performing PLS-R on mean centered spectra (Andersen and Bro, 2010). This was followed by verification of reference values allocated to the outlier and inspection of its spectra to find anomalies. An outlier (a replicate) was removed if the correct reference value had been allocated yet its spectrum was visually different from the other replicates. In this study not more than 3% of the data was removed as outliers. To optimize the model performance and aid interpretation, wavelength or variable selection was then carried out on mean centered spectral data using interval partial least squares (iPLS). This method iteratively adds (forward) or removes (backward) intervals as long as the prediction performance in cross-validation (RMSECV) keeps improving. In this way, the spectral regions which contain the most relevant information were identified (Andersen and Bro, 2010; Nørgaard et al., 2000). Forward and reverse iPLS with an interval width of either 50 or 100 variables were explored. This was followed by further pre-processing of the spectral data. The pre-processing methods were limited to those that were suited for use after wavelength selection (Andersen and Bro, 2010) which were standard normal variate and orthogonal signal correction always followed by mean centering. Afterwards, PLS-R was performed on the selected wavelength regions before final model selection. Important statistics put into consideration to validate model performance were the coefficient of determination (R^2) and the root mean square of the errors obtained in cross-validation (RMSECV) and prediction (RMSEP). The coefficient of determination indicates the percentage of

variance present in the reference values which are reproduced in the cross validation and prediction. The RMSE gives an average uncertainty that is expected for the predicted values. Reliable PLS-R models show high R^2 values and low RMSE values (Saeys et al., 2019; Conzen, 2014; Naes et al., 2002).

Solo version 8.7, the stand alone version of the PLS toolbox for Matlab (Eigenvector Research Inc., Manson, USA) was used for model development.

3. Results and Discussion

3.1 Logistic regression modelling of cooking kinetic data

Logistic regression modelling resulted in cooking times as summarized in **Table 2**. The corresponding cooking curves (experimental data points and predicted lines) are illustrated in **Appendix A Figure A.1**. The curves had an initial lag phase followed by an exponential phase and a final stationary phase. The duration of each phase was largely dependent on the variety, storage conditions and storage time. For all varieties the logistic regression predictor curves of samples stored for 0 months and at 25 °C and 75% relative humidity (reference sample) had the steepest exponential phases. This was an indication that fresh beans cook in a more uniform manner than aged beans. Whereas the exponential phases of the aged beans rose more gradually. This suggests that even beans within the same variety having undergone the same storage conditions will eventually cook differently. This might be because the beans may have different sensitivity to develop the HTC defect and/or ageing conditions were not completely uniform throughout the samples during storage. This observation necessitated the milling of the samples to ensure NIR measurements were carried out for homogenous samples as only one cooking time was determined per sample (i.e. RH 0/25/75). Additionally, it has been reported that spectra obtained from milled beans resulted in better predictions of physical and chemical characteristics

than spectra obtained from whole beans (Wang, et al., 2014; Plans et al., 2013). It could be possible that carrying out NIR spectroscopy measurements on whole beans using NIR in transmittance or transreflectance mode may provide sufficient information on bean quality for the prediction of cooking times of aged beans.

Among the reference samples, red haricot had the shortest cooking time while Canadian wonder showed the longest cooking time (**Table 2**). In all varieties except rose coco, the cooking curves of beans stored for 6 months, at 45 °C and 75% relative humidity had the longest cooking times. For this storage treatment, red haricot had the shortest cooking time (233 minutes) whereas pinto showed the longest cooking time (306 minutes). Additional information on cooking times for other samples included in this study can be found in **Appendix B Table B.1**. It should be noted that the differences in cooking times of the samples were attributed to the variety and storage conditions only. The influence of the individual explanatory variables; temperature, relative humidity and period of storage as well as variety, for this data was already extensively discussed in the work of Kinyanjui et al., (2017).

3.2 NIR spectra of aged beans

The raw spectral data for the DH and WSC bean samples, separately, in the wavelength region between 1000-2600 nm are illustrated in **Figure 2**. The general shape of the spectral data of the common bean samples obtained in this work was similar to those obtained in other NIR spectroscopy studies on milled raw Chinese faba beans and common beans based on visual comparison (Plans et al., 2014; Wang et al., 2014). The observed prominent peaks were inspected to associate them to functional groups and compounds present in beans. The wavelengths 1195 nm and 1762 nm are associated with aliphatic hydrocarbons, 1465 nm and 1564 nm are associated with secondary amides/proteins (R-NH-R), 1938 nm (O-H) are associated with water, and 2105

nm and, 2322 nm and 2507 nm (C-H stretching) are associated with polysaccharides (Workman and Weyer, 2012). These peaks of the raw spectra could be attributed to the fact that beans contain relatively high quantities of proteins, carbohydrates and an appreciable amount of aliphatic hydrocarbons including alcohols and aldehydes (Rui et al., 2011; Carbonaro et al., 2008). The shapes of the spectra for both DH and WSC beans were similar having peaks at the same wavelengths. However, there was a larger variation in the absorbance values of WSC beans compared to DH beans. The differences may be attributed not only to expected differences in composition (Stordrange, 2003; Plans et al., 2014), but also to increased light scattering (Rinnan et al., 2009) in WSC beans possibly due to the presence of distinctive and visible patterns exhibited by the surface of seed coats of common beans (Berrios et al., 1998).

3.3 Prediction of cooking times of beans by the PLS-R models

The most desirable outcome would have been to successfully develop a PLS-R model that could predict cooking times in aged beans over a range of varieties for DH or WSC bean samples. However, as depicted in **Figure 3** for both DH and WSC the four-variety models demonstrated relatively poor predictive abilities and large RMSE values. It was observed, however, that the PLS-R model of the DH bean samples in general performed slightly better than of the WSC bean samples in terms of having lower RMSECV and RMSEP values. Additionally, as illustrated in **Figures 3A, 3B, 3D and 3E** the replicate spectra for the DH beans samples resulted in more similar predictions than those of the WSC beans samples. Hence de-hulling beans during sample preparation resulted in a model with better predictive performance for the four-variety models. The better performance of the DH beans could probably be explained by the fact that cooking times in aged beans are majorly attributed to changes in cotyledon composition/structure, with the seed coat having minimal contribution (Chigwedere et al., 2018). Thus, the inclusion of seed coat

material may have created additional sources of noise/interferences rather than useful information about cooking time changes as a consequence of bean ageing. From **Figures 3D and 3E**, it can be noted that the four-variety models particularly struggled to predict short cooking times. These models could thus be adopted for the rough screening of aged beans with expected longer cooking times. They would, however, find little or no application in cases where highly accurate cooking times of fresh beans, which tend to exhibit short cooking times, are required. It was also observed that the spectra of both DH and WSC bean samples clustered into two different groups; Canadian wonder and pinto on the one hand, and red haricot and rose coco on the other (**Figures 3C and 3F**). Therefore, based on this observation, two-variety PLS-R models were developed with the intention of improving model performance. Later on one-variety models were also developed.

In **Table 3**, an overview of the different DH and WSC bean samples using one- and two-variety PLS-R models is given. From this table, it is clear that there was a reduction in the RMSECV and RMSEP values of the two-variety models compared to the four-variety models. The greatest reductions in prediction errors of 71% and 75% were achieved by the one-variety RH DH and WSC bean samples, respectively, compared to the four-variety models. To illustrate the improvements attained by developing one- and two-variety models, RH and RC (two-variety models) and RH (one-variety models) are used as examples (**Figures 4 and 5**). Short cooking times were also better predicted by the two- and one-variety models compared to the four-variety models. Thus, the cooking times in aged beans were better predicted if PLS-R modelling was carried out for systems with reduced biological complexity. This finding is similar to another study where models developed based on the visible/NIR and hyperspectral NIR imaging to predict texture of canned beans were affected by genetic diversity of the beans (Mendoza et al., 2018a). In this study, they reported that a decrease in diversity resulted in a better model performance.

The difference in RMSECV and RMSEP values for the same model can be used as a measure for the stability of the prediction performance, with small differences indicating higher stability. Given the difference in the RMSECV and RMSEP values of all the two-variety models was approximately 0-2 minutes (**Table 3**), these models can be considered relatively stable in their prediction performance compared to the four-variety models (5-7 minutes). Furthermore, a comparison of RMSEC and RMSECV values for models for DH bean samples with models for the WSC bean samples showed small differences (1 to 5 minutes) for the one- and two-variety models compared to the four-variety models (8 to 20 minutes) (**Table 3**). This is probably because the models could better cope with the reduced influence of seed coat materials which translated in less biological diversity. This suggests that de-hulling may not be a necessary step in sample preparation during the development of the one- and two-variety models. This information seems relevant in the context of developing a fast method for predicting cooking times in beans.

3.4 Selected variables and pre-processing techniques during PLS-R model development

The results for the different variable selection and pre-processing strategies applied to the spectral data of the bean samples are summarized in **Table 4**. The variable selection methods that led to the best final models varied. This is because different variable selection methods apply better to certain sets of spectral data than others depending on the biological complexity of the samples (Andersen and Bro, 2010). In spite of this, in the PLS-R models of both DH and WSC bean samples the wavelength ranges within 1000-1083 nm (11 out of 14 models) and 1446-1860 nm (8 out of 14 models) were mostly selected. This indicated that these regions contain useful information for predicting cooking time in aged beans.

Variable selection was found to be useful for simplifying the interpretability of the obtained PLS-R models. The one-variety RH DH beans samples model had the lowest RMSECV value (**Table**

4) and its regression vector (**Figure 6**) was used to identify wavelengths that were highly correlated to the cooking times of aged beans. The wavelengths that positively correlated to cooking times of aged beans included 1142 nm (aromatic hydrocarbons (C-H)), 1440 nm (aliphatic hydrocarbons (C-H)), and 1821 nm (cellulose). This is probably because aromatic and aliphatic hydrocarbons are volatiles found in common beans (Oomah et al., 2007) and these compounds evolve and increase with ageing (Chigwedere et al., 2019b). Hence, a positive correlation of these volatiles to cooking times could have been expected. The wavelengths that were negatively correlated to cooking times were 1690 nm (protein β sheets structures) and 2010 (amides (N-H)). It has been suggested that the secondary structure of bean protein (β sheets) influence protein digestibility (Carbonaro et al., 2012). The negative correlation could be attributed to the fact that protein digestibility decreases as beans age (Parmar et al., 2017). Increased cooking times with ageing of beans is a classic indicator of HTC. It was therefore hypothesized in this study that the identification of wavelengths that correlate to increasing cooking times in aged common beans could help explain the mechanism of HTC development. However, even though there were correlations observed between wavelengths and the cooking times, it is difficult to link the corresponding compounds or the changes they undergo with ageing of beans to any of the theories that have been put forward to explain the development of the HTC (see introduction). The fact that the models rely on spectral changes other than the ones directly related to the mechanism of hard to cook development but occurring in parallel to these (and thus responsible for the correlation between the observed wavelengths and the time to cook) may partially explain the difficulty to obtain models that cover an extensive pallet of varieties. In this study, no correlation was observed between wavelengths representing water (-OH) and the cooking times of aged beans. This suggests that it may not be necessary to condition samples, to reduce their moisture variation, before

carrying out NIR measurements. This would allow to shorten the sample preparation time substantially by 10 days. For the two- and one-variety models sample preparation would therefore entail milling beans with their seed coats on.

With respect to pre-processing, in the majority of the cases SNV in combination with mean centering resulted in the best performing models. For some cases, OSC together with mean centering or mean centering only resulted in the best performing model (**Table 4**). These findings are different to those of Plans et al. (2014) and Mendoza et al. (2018a), who identified a single pre-processing method as best suited to the different beans spectral data sets. This different outcome could be attributed to the process of variable selection which was carried out in this study and not in the other studies mentioned. Therefore, in addition to the fact that pre-processing enhances differences between different samples (Gautam et al., 2015), it can be suggested that different pre-processing techniques were suitable for different combinations of wavelength regions.

4. Conclusions

This study investigated the potential of NIR spectroscopy to predict cooking times of four varieties (Canadian wonder, pinto, red haricot and rose coco) of aged common beans. Firstly, evaluating cooking of beans using finger pressing followed by analyzing the kinetic data with logistic regression seems to be a suitable methodology to obtain cooking times. Secondly, these cooking times of whole aged beans can be predicted from the NIR spectra of milled raw beans. NIR spectra from milled de-hulled beans gave slightly better prediction performance of four-variety models than those milled from whole beans at the expense of longer sample preparation times. Further improvements in prediction performance were achieved by developing PLS-R models for those varieties with similar NIR spectra. Two-variety models including Canadian wonder and pinto in one model, and red haricot and rose coco reduces prediction errors and sample preparation time

by eliminating the need to remove seed coats. Even better results were obtained for single-variety models (RMSECV=8 minutes, $R^2_C=0.98$, $R^2_{CV}=0.93$). In the future, the robustness of single-variety PLS-R models to predict cooking times in aged beans should be investigated for samples grown in different geographic locations and different seasons. Unfortunately, it was not possible to identify the molecules at the basis of the HTC defect from the NIR spectra or prediction models. To this end, alternative strategies should be explored.

Acknowledgments

“The authors acknowledge the financial support from VLIR-UOS through funding of the Jomo Kenyatta University of Agriculture and Technology Institutional University Cooperation (IUC) programme ‘Legume Centre of Excellence for Food and Nutrition Security (LCEFoNS)’, grant number KE2017IUC037A101. Opinions of the author(s) do not automatically reflect those of either the Belgian government or VLIR-UOS, and can neither bind the Belgian Government nor VLIR-UOS.”

The funding source played no role in the study design, collection, analysis and interpretation of data, the writing of this manuscript or in the decision to submit the manuscript for publication.

References

- Aguilera, J. M., & Rivera, J. M. (1992). Hard-to-cook defect in black beans : hardening rates, water imbibition and multiple mechanism hypothesis. *Food Research International*, 25, 101–108.
- Andersen, C. M., & Bro, R. (2010). Variable selection in regression — a tutorial. *Journal of Chemometrics*, 24, 728–737. <https://doi.org/10.1002/cem.1360>
- Bernal-Lugo, I., Parra, C., Portilla, M., Peña-Valdivia, C. B., & Moreno, E. (1997). Cotyledon

thermal behavior and pectic solubility as related to cooking quality in common beans. *Plant Foods for Human Nutrition*, 50(2), 141–150. <https://doi.org/10.1007/BF02436033>

Berrios, J.J., Swanson, B.G., & Cheong, W.A. (1998). Structural Characteristics of Stored Black Beans (*Phaseolus vulgaris* L.). *Scanning*, 20:410-417.

Broughton, W. J., Hernandez, G., Blair, M., Beebe, S., Gepts, P., & Vanderleyden, J. (2003). Spp.) – Model Food Legumes. *Plant and Soil*, 252(55), 55–128.

Câmara, C., Urrea, C., & Schlegel, V. (2013). Pinto Beans (*Phaseolus vulgaris* L.) as a Functional Food: Implications on Human Health. *Agriculture*, 3(1), 90–111. <https://doi.org/10.3390/agriculture3010090>

Carbonaro, M., Maselli, P., Dore, P., & Nucara, A. (2008). Application of Fourier transform infrared spectroscopy to legume seed flour analysis. *Food Chemistry*, 108(1), 361–368. <https://doi.org/10.1016/j.foodchem.2007.10.045>

Carbonaro, M., Maselli, P., & Nucara, A. (2012). Relationship between digestibility and secondary structure of raw and thermally treated legume proteins : a Fourier transform infrared (FT-IR) spectroscopic study. *Amino Acids*, 43, 911–921. <https://doi.org/10.1007/s00726-011-1151-4>

Castro-Guerrero, N. A., Isidra-Arellano, M. C., Mendoza-Cozatl, D. G., & Valdés-López, O. (2016). Common Bean: A Legume Model on the Rise for Unraveling Responses and Adaptations to Iron, Zinc, and Phosphate Deficiencies. *Frontiers in Plant Science*, 7, 1–7. <https://doi.org/10.3389/fpls.2016.00600>

- Chávez-Mendoza, C., & Sánchez, E. (2017). Bioactive compounds from mexican varieties of the common bean (*Phaseolus vulgaris*): Implications for health. *Molecules*, 22(8). <https://doi.org/10.3390/molecules22081360>
- Chigwedere, C. M., Njoroge, D. M., Loey, A. M. Van, & Hendrickx, M. E. (2019a). Understanding the Relations Among the Storage , Soaking , and Cooking Behavior of Pulses : A Scientific Basis for Innovations in Sustainable Foods for the Future. *Reviews in Food Science and Food Safety*, 18, 1135–1165. <https://doi.org/10.1111/1541-4337.12461>
- Chigwedere, C. M., Olaoye, T. F., Kyomugasho, C., Jamsazzadeh Kermani, Z., Pallares Pallares, A., Van Loey, A. M., & Hendrickx, M. E. (2018). Mechanistic insight into softening of Canadian wonder common beans (*Phaseolus vulgaris*) during cooking. *Food Research International*, 106(October 2017), 522–531. <https://doi.org/10.1016/j.foodres.2018.01.016>
- Chigwedere, C. M., Tadele, W. W., Yi, J., Wibowo, S., Kebede, B. T., Van Loey, A. M., & Hendrickx, M. E. (2019b). Insight into the evolution of flavor compounds during cooking of common beans utilizing a headspace untargeted fingerprinting approach. *Food Chemistry*, 275(June 2018), 224–238. <https://doi.org/10.1016/j.foodchem.2018.09.080>
- Cichy, K. A., Wiesinger, J. A., & Mendoza, F. A. (2015). Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 128(8), 1555–1567. <https://doi.org/10.1007/s00122-015-2531-z>
- Conzen, J. (2014). *Multivariate Calibration: A practical guide for developing methods in the quantitative analytical chemistry* (3rd ed. pp 23-49). Bruker Optik GmbH

- Garcia-Vela, L. A., & Stanley, D. W. (1989). Water-Holding Capacity in Hard-to-Cook Beans (*Phaseolus vulgaris*): Effect of pH and Ionic Strength. *Journal of Food Science*, 54(4), 1080–1081.
- Garcia, E., Filisetti, T. M. C. C., Udaeta, J. E. M., & Lajolo, F. M. (1998). Hard-To-Cook Beans (*Phaseolus vulgaris*): Involvement of Phenolic Compounds and Pectates. *Journal of Agricultural and Food Chemistry*, 46(6), 2110–2116. <https://doi.org/10.1021/jf970848f>
- Garcia, R. A. V., Rangel, P. N., Bassinello, P. Z., Brondani, C., Melo, L. C., Sibov, S. T., & Vianello-Brondani, R. P. (2012). QTL mapping for the cooking time of common beans. *Euphytica*, 186(3), 779–792. <https://doi.org/10.1007/s10681-011-0587-7>
- Gautam, R., Vanga, S., Ariese, F., & Umapathy, S. (2015). Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation*, 2(1). <https://doi.org/10.1140/epjti/s40485-015-0018-6>
- Guevara-González, R. G., Torres-Pacheco, I., Reynoso-Camacho, R., Ramos-Gomez, M., & Loarca-Pina, G. (2006). Bioactive components in common beans (*Phaseolus vulgaris* L.). *Research Signpost*, 37661(2), 217–236. Retrieved from http://www.trnres.com/ebook/uploads/gonzalezfinal/T_1231138943gonzalez 10.pdf
- Hentges, D. L., Weaver, C. M., and Nielsen, S. S. (1991). Changes of Selected Physical & Chemical Components the Development of the Hard-to-Cook Bean Defect. *Journal of Food Science*, 56(2), 436–442.
- Hohlberg, A. I., & Stanley, D. W. (1987). Hard-To-Cook Defect in Black Beans. Protein and Starch Considerations. *Journal of Agricultural and Food Chemistry*, 35, 571–576.

- 504 Jones, P. M. B., & Boulter, D. (1983). The Cause of Reduced Cooking Rate in *Phaseolus vulgaris*
505 Following Adverse Storage Conditions. *Journal of Food Science*, 48, 623–649.
- 506 Katungi, E., Sperling, L., Karanja, D., Farrow, A., & Beebe, S. (2011). Relative importance of
507 common bean attributes and variety demand in the drought areas of Kenya. *Journal of*
508 *Development and Agricultural Economics*, 3(8), 411–422.
- 509 Kilmer, O. L., Seib, P. A., & Hoseney, R. C. (1994). Effects of minerals and apparent phytase
510 activity in the development of the hard-to-cook state of beans. *Cereal Chemistry*, 71(5), 476–
511 482.
- 512 Kinyanjui, P. K., Njoroge, D. M., Makokha, A. O., Christiaens, S., Ndaka, D. S., & Hendrickx, M.
513 (2015). Hydration properties and texture fingerprints of easy-and hard-to-cook bean varieties.
514 In *Food Science and Nutrition* (Vol. 3). <https://doi.org/10.1002/fsn3.188>
- 515 Kinyanjui, P. K., Njoroge, D. M., Makokha, A. O., Christiaens, S., Sila, D. N., & Hendrickx, M.
516 (2017). Quantifying the Effects of Postharvest Storage and Soaking Pretreatments on the
517 Cooking Quality of Common Beans (*Phaseolus vulgaris*). *Journal of Food Processing and*
518 *Preservation*, 41, 1–10. <https://doi.org/10.1111/jfpp.13036>
- 519 Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5th
520 ed., pp 555-603). McGraw-Hill/Irwin, Boston MA.
- 521 Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive
522 analysis of biological materials. *Chemical Society Reviews*, 43(24), 8200–8214.
523 <https://doi.org/10.1039/c4cs00062e>

- 524 Mendoza, F. A., Cichy, K. A., Sprague, C., Goffnet, A., Lu, R., & Kelly, J. D. (2018a). Prediction
 525 of canned black bean texture (*Phaseolus vulgaris* L .) from intact dry seeds using visible /
 526 near infrared spectroscopy and hyperspectral imaging data. *Journal of the Science of Food*
 527 *and Agriculture*, 98, 283–290. <https://doi.org/10.1002/jsfa.8469>
- 528 Mendoza, F. A., Wiesinger, J. A., Lu, R., Nchimbi-Msolla, S., Miklas, P. N., Kelly, J. D., & Cichy,
 529 K. A. (2018b). Prediction of Cooking Time for Soaked and Unsoaked Dry Beans (L.) Using
 530 Hyperspectral Imaging Technology. *The Plant Phenome Journal*, 1–9.
 531 <https://doi.org/10.2135/tppj2018.01.0001>
- 532 Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *A user friendly guide to multivariate*
 533 *calibration and classification*. NIR Publications, Chichester, United Kingdom.
- 534 Nicolai, B., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., & Lammertyn, J. (2007).
 535 Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy:
 536 A review. *Postharvest biology and technology*, 46 (2), 99-118.
- 537 Nørgaard, L. N., Saudland, A., Wagner, J., Nielsen, J. P., Unck, L. M., & Engelsen, S. B. (2000).
 538 Interval Partial Least-Squares Regression (i PLS): A Comparative Chemometric Study with
 539 an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy*, 54(3), 413–419.
- 540 Oomah, B. D., Liang, L. S. Y., & Balasubramanian, P. (2007). Volatile compounds of dry beans
 541 (*Phaseolus vulgaris* L.). *Plant Foods for Human Nutrition*, 62(4), 177–183.
 542 <https://doi.org/10.1007/s11130-007-0059-3>
- 543 Osborne, B. G. (2000a). Near-Infrared Spectroscopy in Food Analysis. *Encyclopedia of Analytical*
 544 *Chemistry*, 1–14. <https://doi.org/10.1002/9780470027318.a1018>

- 545 Osborne, B. G. (2000b). Near-Infrared Spectroscopy in Food Analysis. *Encyclopedia of Analytical*
546 *Chemistry*, 1–14. <https://doi.org/10.1002/9780470027318.a1018>
- 547 Parmar, N., Singh, N., Kaur, A., & Thakur, S. (2017). Comparison of color , anti-nutritional factors
548 , minerals , phenolic profile and protein digestibility between hard-to-cook and easy-to-cook
549 grains from different kidney bean (*Phaseolus vulgaris*) accessions. *Journal of Food Science*
550 *and Technology*, 54(4), 1023–1034. <https://doi.org/10.1007/s13197-017-2538-3>
- 551 Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new
552 perspectives – A review. *Analytica Chimica Acta*, 1026, 8–36.
553 <https://doi.org/10.1016/j.aca.2018.04.004>
- 554 Peng, C. J., Lee, K. U. K. L., & Ingersoll, G. M. (2002). An Introduction to Logistic Regression
555 Analysis and Reporting. *The Journal of Educational Research*, 96(1), 3–14.
- 556 Plans, M., Simó, J., Casañas, F., del Castillo, R. R., Rodriguez-Saona, L. E., & Sabaté, J. (2014).
557 Estimating sensory properties of common beans (*Phaseolus vulgaris* L.) by near infrared
558 spectroscopy. *Food Research International*, 56(2014), 55–62.
559 <https://doi.org/10.1016/j.foodres.2013.12.003>
- 560 Plans, M., Simó, J., Casañas, F., & Sabaté, J. (2012). Near-infrared spectroscopy analysis of seed
561 coats of common beans (*Phaseolus vulgaris* L.): A potential tool for breeding and quality
562 evaluation. *Journal of Agricultural and Food Chemistry*, 60(3), 706–712.
563 <https://doi.org/10.1021/jf204110k>

- Plans, M., Simó, J., Casañas, F., Sabaté, J., & Rodriguez-Saona, L. (2013). Characterization of common beans (*Phaseolus vulgaris* L.) by infrared spectroscopy: Comparison of MIR, FT-NIR and dispersive NIR using portable and benchtop instruments. *Food Research International*, 54(2), 1643–1651. <https://doi.org/10.1016/j.foodres.2013.09.003>
- Pojić, M., Mastilović, J., Palić, D., & Pestorić, M. (2010). The development of near-infrared spectroscopy (NIRS) calibration for prediction of ash content in legumes on the basis of two different reference methods. *Food Chemistry*, 123(3), 800–805. <https://doi.org/10.1016/j.foodchem.2010.05.013>
- Rebello, C. J., Greenway, F. L., & Finley, J. W. (2014). Whole grains and pulses: A comparison of the nutritional and health benefits. *Journal of Agricultural and Food Chemistry*, 62(29), 7029–7049. <https://doi.org/10.1021/jf500932z>
- Rinnan, Å., Berg, F. van den, & 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>
- Rui, X., Boye, J. I., Ribereau, S., Simpson, B. K., & Prasher, S. O. (2011). Comparative study of the composition and thermal properties of protein isolates prepared from nine *Phaseolus vulgaris* legume varieties. *Food Research International*, 44(8), 2497–2504. <https://doi.org/10.1016/j.foodres.2011.01.008>
- Saeys, W., Nguyen, D., Van Beers, R., & Nicolai, B. (2019). Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. *Postharvest Biology and Technology*, 158. doi:10.1016/j.postharvbio.2019.110981

- Shiga, T. M., & Lajolo, F. M. (2006). Cell wall polysaccharides of common beans (*Phaseolus vulgaris* L.) - Composition and structure. *Carbohydrate Polymers*, 63(1), 1–12.
<https://doi.org/10.1016/j.carbpol.2005.06.025>
- Shiga, T. M., Lajolo, F. M., & Filisetti, T. M. C. C. (2004). Changes in the cell wall polysaccharides during storage and hardening of beans. *Food Chemistry*, 84(1), 53–64.
[https://doi.org/10.1016/S0308-8146\(03\)00166-3](https://doi.org/10.1016/S0308-8146(03)00166-3)
- Stanley, D. W. (1992). A possible role for condensed tannins in bean hardening. *Food Research International*, 25(3), 187–192. [https://doi.org/10.1016/0963-9969\(92\)90136-S](https://doi.org/10.1016/0963-9969(92)90136-S)
- Stordrange, L. (2003). *Multivariate analysis of near-infrared spectroscopic data from reaction monitoring*.
- Wang, J., Liu, H., & Ren, G. (2014). Near-infrared spectroscopy (NIRS) evaluation and regional analysis of Chinese faba bean (*Vicia faba* L.). *Crop Journal*, 2(1), 28–37.
<https://doi.org/10.1016/j.cj.2013.10.001>
- Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., & Declerck, F. (2019). The Lancet Commissions Food in the Anthropocene : the EAT – Lancet Commission on healthy diets from sustainable food systems. *Lancet*, February 2(10170), 447–492.
[https://doi.org/10.1016/S0140-6736\(18\)31788-4](https://doi.org/10.1016/S0140-6736(18)31788-4)
- Williams, P. (2004). Near-Infrared technology: Getting the best out of light. *Near Infrared Technology*, 132.
- Wood, J. A. (2017). Evaluation of cooking time in pulses: A review. *Cereal Chemistry*, 94(1), 32–48. <https://doi.org/10.1094/CCHEM-05-16-0127-FI>

606 Workmann, J.J., & Weyer, L. (2012). *Practical guide and spectral atlas for interpretive near-*
607 *infrared spectroscopy* (2nd ed., pp. 239-274). Boca Roca, USA.

ACCEPTED MANUSCRIPT

Table 1: Sampling of stored beans

Storage Conditions		Sampling Periods (months)			
Relative Humidity (% RH)	Temperature (°C)	t ₁	t ₂	t ₃	t ₄
75	25	0	4	8	12
75	35	-	4	8	12
75	45	-	2	4	6
83	35	-	2	4	6

Table 2: Results obtained after performing logistic regression

Variety	Cooking times of Aged Beans (minutes)		
	Minimum	Maximum	Number of samples
Canadian wonder	95	279	12
Pinto	79	306	11
Red haricot	62	233	12
Rose coco	75	235	12

Table 3: Prediction results of the PLS-R models

PLS-R models	LV	<u>Calibration</u>		<u>Cross Validation</u>		<u>Test Set Validation</u>	
		R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP
Four-variety DH	12	0.87	19	0.71	28	0.97	21
Four-variety WSC	8	0.72	27	0.52	36	0.77	41
CW and Pinto DH	10	0.94	13	0.77	26	0.99	26
CW and Pinto WSC	4	0.85	21	0.78	25	0.96	25
RH and RC DH	9	0.90	12	0.78	18	0.90	18
RH and RC WSC	11	0.94	11	0.75	18	0.89	18
CW DH	6	0.99	10	0.93	14		
CW WSC	8	0.97	9	0.92	15		
Pinto DH	7	0.99	8	0.96	12		
Pinto WSC	7	0.99	5	0.93	17		
RH DH	8	0.98	4	0.93	8		
RH WSC	6	0.96	7	0.93	9		
RC DH	9	0.97	8	0.93	13		
RC WSC	5	0.93	11	0.86	16		

Table 4: Results of the PLS-R models after optimization of the variable selection and pre-processing

PLS-R models	Variable Selection Method	Wavelengths selected	Pre-processing*
Four-variety DH	Reverse iPLS Width of interval: 100 variables	1000-1301, 1446-1860, 2174-2603	MC
Four-variety WSC	Reverse iPLS Width of interval: 100 variables	1000-1301, 1446-2603	SNV, MC
CW and Pinto DH	Reverse iPLS Width of interval: 50 variables	1000-1083, 1446-2004, 2174-2603	SNV, MC
CW and Pinto WSC	Forward iPLS Width of interval: 50 variables	1000-1040, 1301-1446, 1628-1860, 2006-2172	SNV, MC
RH and RC DH	Reverse iPLS Width of interval: 100 variables	1000-1083, 1628-2603	OSC, MC
RH and RC WSC	Forward iPLS Width of interval: 50 variables	1040-1083, 1532-1736, 1862-2172	MC
CW DH	Reverse iPLS Width of interval: 100 variables	1182-1446, 1628-2603	SNV, MC
CW WSC	Forward iPLS Width of interval: 100 variables	1000-1083, 1862-2172	MC
Pinto DH	Reverse iPLS Width of interval: 50 variables	1040-1083, 1182-1239, 1301-1369, 1446-1627, 1862-2603	SNV, MC
Pinto WSC	Forward iPLS Width of interval: 100 variables	1000-1301, 1446-1860	OSC, MC
RH DH	Reverse iPLS Width of interval: 50 variables	1000-1083, 1182-1301, 1628-2603	SNV, MC
RH WSC	Forward iPLS Width of interval: 50 variables	1000-1083, 1239-1369, 1532-1627, 1737-1860, 2006-2370	SNV, MC
RC DH	Reverse iPLS Width of interval: 50 variables	1000-1040, 1083-1131, 1239-1301, 1370-1446, 1532-1627, 1862-2004, 2174-2603	MC
RC WSC	Forward iPLS Width of interval: 50 variables	1370-1446, 2006-2172	SNV, MC

*Pre-processing techniques: MC (mean center), SNV (standard normal variate), OSC (orthogonal signal correction)

Figure 1. Schematic flow of the experimental steps

Figure 2: The raw NIR spectroscopy data spectra of the DH bean sample (A) and WSC bean sample (B). The spectra is color coded according to cooking times in minutes as indicated in the horizontal bar on the right.

Figure 3: Output of PLS-R models for four-variety models DH beans samples (A-C) and WSC beans samples (D-F). A. Calibration plot, B. Test set validation plot, C. Score plot, D. Calibration plot, E. Test set validation plot, F. Score plot.

Figure 4: Output of PLS-R models for two-variety (RH and RC) models DH beans samples (A and B) and WSC beans samples (C and D). A. Calibration plot, B. Test set validation plot, C. Calibration, D. Test set validation plot

Figure 5: Output of PLS-R models for one-variety (RH) models. DH beans samples A. Calibration plot of DH beans samples, B. Calibration of plot WSC beans samples.

Figure 6: An illustration of RH DH beans regression vector after variable selection and PLS-R model development.