University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

2012

Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels

Lidia Esteve Agelet
Iowa State University, lesteve@iastate.edu

David D. Ellis
National Center for Genetic Resources Preservation

Susan Duvick
Iowa State University

A. J. Susana Goggi
Iowa State University

Charles R. Hurburgh
Iowa State University

See next page for additional authors

Agelet, Lidia Esteve; Ellis, David D.; Duvick, Susan; Goggi, A. J. Susana; Hurburgh, Charles R.; and Gardner, Candice A., "Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels" (2012). *Publications from USDA-ARS / UNL Faculty*. 828.
https://digitalcommons.unl.edu/usdaarsfacpub/828

Authors

Lidia Esteve Agelet, David D. Ellis, Susan Duvick, A. J. Susana Goggi, Charles R. Hurburgh, and Candice A. Gardner

Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels

Lidia Esteve Agelet [a],[*], David D. Ellis [b], Susan Duvick [c], A. Susana Goggi [d], Charles R. Hurburgh [a], Candice A. Gardner [c]

[a] Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011, USA
[b] USDA-ARS, National Center for Genetic Resources Preservation, Fort Collins, CO 80521, USA
[c] USDA-ARS Plant Introduction Research Unit and Department of Agronomy, Iowa State University, Ames, IA 50011, USA
[d] Department of Agronomy and Seed Science Center, Iowa State University, Ames, IA 50011, USA

ABSTRACT

The current US corn grading system accounts for the portion of damaged kernels, measured by time-consuming and inaccurate visual inspection. Near infrared spectroscopy (NIRS), a non-destructive and fast analytical method, was tested as a tool for discriminating corn kernels with heat and frost damage. Four classification algorithms were utilized: Partial least squares discriminant analysis (PLS-DA), soft independent modeling of class analogy (SIMCA), k-nearest neighbors (K-NN), and least-squares support vector machines (LS-SVM). The feasibility of NIRS for discriminating normal or viable-germinating corn kernels and soybean seeds from abnormal or dead seeds was also tested. This application could be highly valuable for seed breeders and germplasm-preservation managers because current viability tests are based on a destructive method where the seed is germinated. Heat-damaged corn kernels were best discriminated by PLS-DA, with 99% accuracy. The discrimination of frost-damaged corn kernels was not possible. Discrimination of non-viable seeds from viable also was not possible. Since previous results in the literature contradict the current damage-discrimination results, the threshold of seed damage necessary for NIRS detection should be analyzed in the future. NIRS may accurately classify seeds based on changes due to damage, without any correlation with germination.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Corn (*Zea mays L.*) is the main component in feed for cattle in US, with 90% of all corn used for feed (USDA, 2010). Corn is also processed into many end products for human consumption and industrial uses. Genetic traits, compositional characteristics, and overall quality of corn are important factors which impact final product. For instance, wet milling processes for ethanol production require corn with high starch content. Quality of corn (properly dried, no cracked or broken kernels, absence of foreign material) is

more important for wet milling than for commodity grain. Environmental conditions and other post-harvest activities impact the final grain quality. High moisture, heat, freezing, and artificial drying are some of the factors that damage grain and negatively impact grain quality and economic value. Six corn grades in the US establish a measurement of grain quality based on batch test weight and percent of total damaged kernels (USDA-GIPSA, 2001). For total damaged kernels, the maximum percentage of heat-damaged (by either excessive drying or improper moisture adjustment during storage), broken, and foreign material (BCFM) in each grade is specified. The percentage of frost, sprout or mold damaged seeds is included in the total damaged seed category. Other characteristics which affect corn quality but are not reflected in the US grading system are % waxy corn, stress cracks, and insect infestation.

Both quality control of grains and determination of germination of stored seeds are carried out by qualified personnel and are time-consuming tasks. The viability of seed through germination tests which reduce the number of stored seeds in germplasm collections. Germination, expressed as the percentage of

germinated seeds over the total tested seeds, is the estimated capability of a seed lot to produce normal plants with good vigor under favorable controlled conditions (AOSA, 2010). Dormancy (%) is the percentage of seeds which have not grown in the testing period, but are alive. This may be due to either intrinsic physical characteristics of the seed (seed coat and internal structures) or induced conditions by environmental changes. The percentage of dormant and germinating seed is required for labeling seed packages for sale. Higher germination seed rates assures farmers higher emergence in normal field conditions, and longer storage potential of the seeds. The standard germination test is the official method for testing seed viability, and methodologies are designed according to the best environmental conditions for seedling growth. The test also allows detecting abnormal, dead and dormant seeds. Tetrazolium Chloride staining of seeds (Cottrell, 1948), another destructive test, serves as a faster alternative for testing seed viability, but since it relies on the subjective evaluation of dead seed structures it is necessary to validate the results with the standard germination test (International Seed Testing Association, 1985). Some researchers have analyzed the germination process as an attempt to determine in advance which seeds produce either abnormal plants or are dead seeds. Non-destructive studies of changes in temperature profile during seed aging helped in the prediction of seed viability by Infrared thermography (Kranner et al., 2010). By repetitive comparison of the thermal profile of a given seed with previously studied seed profiles, dead seeds were discriminated from live seeds at 85% accuracy, and heat-killed seeds from viable seeds at 100%. Water binding within seeds and on seed cells seems to be the most influencing factor in seed aging and deterioration (Becker, 1998), followed by processes of protein and lipid modification by oxidation (Bernal-Lugo and Leopold, 1998). Nuclear Magnetic Resonance (NMR) was used to study the changes in tissue water and water-binding from viable and accelerated aging seeds (Krishnan et al., 2004). Differences were found in seed rehydration patterns but it was concluded that more research and advances in instrumentation are needed in order to quantify the water status and seed viability of intact seeds.

NIRS is a technology based on the absorption of near infrared light by organic compounds and water. Some of the well-known advantages of this technology are the high speed of analysis, low sample preparation requirements, and preservation of the sample. Those advantages open the possibilities of applying NIRS for whole seed batch inspection and for breeders' purposes when used in single seed analysis. The limitations, on the other hand, are the initial dependence and reliability to an alternative external reference method (i.e. HPLC, GLC, Combustion, etc.) and high detection limits, which only allows NIRS to quantify compounds above trace concentrations. However, NIRS has been used in several applications for single seed analysis with notable success. Quantitative applications in single corn kernels are mainly targeting oil content (Cogdill et al., 2004; Janni et al., 2008; Jiang et al., 2007; Orman and Schumann, 1992; Spielbauer et al., 2009; Weinstock et al., 2006), protein levels (Spielbauer et al., 2009), moisture calibrations (Armstrong, 2006; Finney and Norris, 1978), and starch (Spielbauer et al., 2009). The predictive ability of those calibrations indicated that NIRS is a suitable technology for screening organic compounds in single corn kernels. The discriminative analyses of corn kernels found in the literature are based on endosperm characteristics (vitreousness, hardness). Accurate discrimination among kernels with vitreous and floury endosperm have been reported (Campbell et al., 2000; Williams et al., 2009). A few studies attempted to discriminate corn kernels according to their toxin contamination level (Dowell et al., 2002; Pearson et al., 2001). The method could discriminate accurately between sound kernels and kernels contaminated at high toxin concentrations (>100 ppm for fumonisin, and >100 ppb for aflatoxin).

There are few studies using NIRS for discriminating sound and damaged soybeans and wheat (*Triticum aestivum* L.) kernels. Wang et al. (2001) analyzed heat-damaged kernels using NIRS and could accurately discriminate heat-damaged kernels (>95% correctly classified) with just two wavelengths and partial least squares discriminant analysis (PLS-DA). They suggested that the classification was driven by differences in light scattering and color change in heat-damaged kernels. A classification model for wheat based on vitreous and non-vitreous endosperms, including defective kernels such as bleached, cracked and sprouted, found that bleached kernels were misclassified (Wang et al., 2002a). Wang et al. (2002b) also classified soybean seeds according to the type of damage (sprout, heat, frost, mold, or weather) using artificial neural networks (ANN) with discrimination success rates over 90% for most of the damaged types. In that study heat damage classifications achieved lower accuracies compared to the ones achieved in heat-damaged wheat kernels (64%). To-date, no publications have analyzed the use of NIRS for seed viability discrimination. Kusama et al. (1997) used near infrared spectroscopy (NIRS) for analyzing aging of soybeans. They classified with 60% accuracy sound versus 3-day artificially aged soybean seeds, 80% accuracy when aged for 5 days, and 100% accuracy when aged for 8 days. However, the correlation with seed viability was not considered.

This paper analyzes the feasibility of Near Infrared Spectroscopy (NIRS) for discriminating heat and frost-damaged corn kernels from undamaged kernels, and discriminate viable and non-viable corn and soybean seeds. Soybean has the second largest acreage in the US after corn. US is also the world's top producer of this oilseed (EPA, 2009). In contrast to corn kernels, soybean seeds do not have a differentiated germ structure and have a higher protein composition, and lower starch content. Due to its relevance in US agriculture, soybean seeds are analyzed together with corn kernels in the viability study.

Four algorithms were tested and compared: Partial least squares discriminant analysis (PLS-DA), soft independent modeling of class analogy (SIMCA), k-nearest neighbors (K-NN), and least-squares support vector machines (LS-SVM).

2. Experimental

2.1. Seed samples

2.1.1. Heat damage

Twenty-four corn kernels each from nine different accessions (216 kernels total) were heat-killed via microwave at the USDA-ARS National Center for Genetic Resources Preservation (Fort Collins, Colorado). The varieties were: NSL 2843, NSL 6528, PI 267209, PI 515179, NSL 32736, PI 167968, PI 213766, PI 176800, and PI 483549. Individual kernels were placed on a petri dish and microwaved for 45 s. Microwaving heat damages the kernels without obvious color change for most of the cases. Three of the total microwaved kernels showed excessive heat damage and were not scanned, leaving a total of 213 microwaved kernels that could be further analyzed. An additional set of 216 corn kernels from the same accessions were used as control seeds (not microwaved).

2.1.2. Frost damage

A hundred kernels from a single hybrid (B37 × IRF311) were obtained from the department of Agronomy in Iowa State University (Ames, Iowa). Fifty of those kernels were artificially frost damaged in their early stages of seed development (moisture content from 50 to 55%) when seeds were on the cob still in the husk. Ears were frozen in a Conviron growth chamber (Controlled

Environment Limited, Winnipeg, Manitoba, Canada) during 24 h frost cycle (DeVries et al., 2007). Damaged and sound corn kernels were stored in a controlled constant temperature (10 °C) and relative humidity (50%) storage for 1 year. Kernels were allowed to equilibrate to ambient temperature before being scanned.

2.1.3. Viability

Corn kernels and soybean seeds were obtained from the National Center for Genetic Resources Preservation (Fort Collins, Colorado). Three corn accessions (NSL 2837, NSL 2838, NSL 2842) and two soybean accessions (PI 79379 and PI 132214) stored at controlled temperature and moisture conditions since the early 1960s were selected. All accessions had multiple germination test results over the period of storage and selections were made based on the most recent viability results (2010 for corn samples and 2006 for soybean samples). Samples with relatively high initial germination rates which had decreased during storage to 50% germination were desired for this study. The available germination results on record for selected corn were: NSL 2837, 54%; NSL 2838, 56%; and NSL 2842, 54%. For soybeans: PI 79379, 80%; and PI 132214, 72%.

For each corn accession, 168 kernels were randomly selected (504 corn kernels analyzed in total). Two hundred forty soybean seeds for each accession were selected, giving a total of 480 soybean seeds analyzed. No cracked, broken, decolorized, or wrinkled seeds and kernels were found in the samples. All seeds were scanned at the Grain Quality Laboratory at Iowa State University (Ames, Iowa). The kernels, individually identified in plastic multi-well plates (24 wells for corn and 48 wells for soybeans), were sent back to the National Center for Genetic Resources Preservation (Fort Collins, Colorado) to conduct the warm standard germination test (7 days at 20/30 °C day/night).

2.2. NIRS spectrometer and data collection

The NIRS instrument used was a Perten DA 7200 (Perten Instruments, Inc., Springfield, IL). It is a diode array instrument which takes measurements from 850 nm to 1,650 nm, at 5 nm intervals (141 data points). The integration time was set to 153 ms. The company provided a special single seed adapter consisting of a concave mirror surface which can be inserted in the place of the regular bulk sample cup. The instrument was set to take three scans per kernel/seed after taking two blank readings, and provide the average spectrum. The seeds were placed with tweezers on the concave mirror with the germ facing up to the light and detectors.

2.3. Data processing and discrimination models

Data was imported from Jcamp format to The Unscrambler software v.9.8 (Camo AS, Trondheim, Norway) for organization purposes. Matlab v.7.10 (Mathworks, Natick, MA) with PLS_toolbox v.5.8.3 functions (Eigenvector Research Inc., Wenatchee, WA) was utilized for data handling and developing the discrimination models. Two data points from each side of the whole spectra were removed because of the noise, leaving the working wavelength range from 860 to 1,640 nm. Assessment for detection of outliers was done by observing the spectra plot and utilizing principal component analysis (PCA). Possible outliers were expected to show high residual values and high Hotelling's T^2 statistic values. T^2 is calculated from the sample scores (projected samples to the PCA model) and gives an idea of the variation of each sample within the PCA model. No clear outliers were detected, hence no spectra were removed from the analysis.

For each of the studies, the following four classification algorithms were tested: Partial least squares discriminant analysis (PLS-DA), Soft Independent Modeling of Class Analogy (SIMCA), Least Squares Support Vector Machines (LS-SVM), and K-nearest neighbors (K-NN). The spectra were analyzed both raw (only mean-centering applied) and preprocessed with standard normal variate (SNV). This preprocessing method is known to reduce the scattering effects smoothing the noise of the signal. It has led to improvement in the results of several studies of single seed applications using NIRS because it removed differences due to sample positioning (Weinstock et al., 2006).

2.3.1. Training and validation sets

For the first two studies of corn damage, 75% of the spectra were utilized for developing the models and the remaining 25% (one spectrum picked every four spectra from the database) were kept for validation. Accuracy of the models were reported by the number of misclassified kernels in each study.

For the heat damage study, a total of 306 corn kernels were utilized as a training set (155 sound and 151 microwaved kernels) and 103 were utilized for validation (52 sound and 51 microwaved). Seventy five corn kernels were part of the training set for the frost damage study (37 sound and 38 frost-treated kernels) and 25 for validation (13 sound and 12 frost-treated kernels).

The percentage of dead seeds from the germination test was low, probably because the selected kernels were in good condition. The abnormal category from the germination test makes reference to seeds which show damage or shredding in cotyledons or shoots, missing shoots or cotyledons, weak roots, or show impaired structures. Although those seeds were able to germinate, they exhibited a lack of vigor and would not survive in normal field conditions. For this reason, abnormal and dead seeds are lumped together as non-viable seeds.

Because some of the classification methods tested such as PLS-DA and LS-SVM can give biased results when the training class sizes are unbalanced (different number of samples in each class) and most of the seeds showed normal germination, we selected a reduced set of seeds from each variety for the training set in order to have a similar number of samples per class (Table 1). The validation set included all samples left after creating the training set.

2.3.2. PLS-DA models

PLS-DA is a supervised classification method in NIRS applications because it allows for highly correlated NIR variables (wavelengths). Similar to partial least squares (PLS) for quantification, data reduction is conducted creating latent variables which are orthogonal with each other but at the same time describes the response variable (in this case, the class labels). The SIMPLS algorithm from the PLS_toolbox v.5.8.3 was used (Eigenvector Research Inc., Wenatchee, WA). Class labels were entered as a logical array, with each class represented by a column of zeros and ones (this last indicating the membership to one of the classes). The threshold for class separation was calculated according to the initial membership in each class. Through leave-one-out cross-validation, the optimal number of latent variables for the final model was selected by

Table 1
Final number of seeds (spectra) for each class utilized in each study. Class 1 are sound seeds, while class 2 are non-viable. N stands for normal, A stands for abnormal, and D stands for dead.

Seed	Training set			Validation set		
	Class 1 (sound)	Class 2 (non-viable)	Total	Class 1 (sound)	Class 2 (non-viable)	Total
Corn	90 N	45 D 45 A	180	242 N	39 D 43 A	324
Soybean	50 N	25 D 26 A	101	299 N	19 D 61 A	379

looking at the fractional misclassification rate of each class and the root mean squared error of cross validation (RMSECV).

2.3.3. SIMCA models

SIMCA models each class independently by principal component analysis (PCA), a method known for reducing data dimensionality, explaining the major sources of variance with fewer variables. Each class is independently modeled by PCA. For each class, an acceptance boundary defined by the maximum residuals of the samples from that group is created. Each class is expected to be modeled by a different number of principal components (PCs). We determined the number of PCs to keep in the PCA model for each class by 10-fold cross-validation, assessed by the eigenvalues, the Q and Hotelling's T^2 plot from cross-validation, and the Predicted Residual Error Sum of Squares (PRESS). The Q statistic is a measure of the residual between a sample in its initial dimension and its projection into the principal components.

When a new sample is classified, it is fitted into each model and its residual variance is calculated. If the variance is significantly higher than the average of the residuals from that class, the sample does not belong to that group. In this study, the combination of Q and Hotelling's T^2 statistics were used for determining which class a sample belonged to. The sample was assigned to the class associated to the model which obtains a low Q and T^2 value after being projected in a specific PCA model. However, it is possible that a sample does not belong to any group or could fit in more than one group. When the sample does not fit in any class (high Q and T^2 values for all classes), it is assigned to the group which results in a smaller residual and T^2. In the second case, the PLS-toolbox algorithm assigns the sample to the class model where the centroid is closest in terms of Euclidean distance.

2.3.4. LS-SVM models

Classification methods known as Support Vector Machines (SVM) were initially created for linear discrimination between two classes. The method is based on finding the widest margin of separation between classes. Using the mapping function (kernel function), this method can deal with complex classification problems which are not linear in the initial dimension but may be at high dimensional spaces. We developed LS-SVM models using the LS_SVMlab v.1.5 toolbox functions for Matlab (Lukas et al., 2003). The Guassian Radial Basis Function (RBF) was selected as the non-linear mapping function. Two parameters had to be optimized: the kernel width (σ) and the model regularization parameter (γ), which is the trade-off between the margin width and tolerance to misclassification. The values of both parameters were simultaneously optimized using grid search ten-fold cross-validation. The input spectra were previously mean-centered, and the sample classes were defined by a binary vector of 1 and -1.

2.3.5. K-NN models

For this algorithm, any new sample is classified according to the majority of votes from its "k" closest neighbors. The distance between two spectra, after mean-centering, is calculated from each wavelength as the difference in intensity between the two wavelength absorbance from each spectrum. The square root of the sum of the squared distances at all wavelengths gives the Euclidean distance between the two spectra in Eq. (1), where the distance d between point i and j is calculated adding the differences in absorbance $\times$ between the two spectra at each wavelength l. When a sample is tested, its Euclidean distance for each sample is calculated and the k closest neighbors are considered for determining the sample class. The number of neighbors k is an odd number which were tested for numbers one, three, five, seven, and nine. The optimal number was selected as the one that lead to the lowest

leave-one-out cross-validation misclassification of the training set. That optimal number was utilized to classify the samples from the validation set.

$$d_{ij} = \sqrt{\sum_{l=1}^{p} \left(x_{il} - x_{jl} \right)^2} \tag{1}$$

3. Results and discussion

3.1. Heat damage

Spectra of heat-damaged kernels overall had low absorptions when compared with sound kernels. This agrees with Wang et al. (2001), where they observed differences in light reflection and scattering between sound and heat-damaged wheat kernels. However, offset and baseline differences can be easily removed by spectral preprocessing. The major absorbance differences between sound and heat-damaged kernels is observed in the carbohydrate regions (for instance at 1,150 nm or 1,190–1,225 nm). This could be due to the fact that starch constitutes more than half of the corn kernel by weight and any change in this fraction may be more easily detectable than any change in the germ. Misclassification rates are summarized in Table 2. PLS-DA achieved the lowest misclassification rates with eight latent factors (seven factors with SNV preprocessing), with >98% classification of the two classes. The regression vector plot also shows the relevance of the carbohydrate region in the classification model (Fig. 1).

SIMCA achieved the second lowest misclassification rate of 4.9% when using raw spectra. Nine PCs were used for heat-damaged kernels and eight for sound kernels. When applying SNV preprocessing the misclassification increased to 15.5%. SNV may reduce the variability between classes due to light scattering and lead to similar PCA models for both classes. This again would agree with the relevance of light scattering in heat-damaged kernel classification (Wang et al., 2001). LS-SVM features for non-linear classification led to higher misclassification rates but it was the only method which benefits from the use of SNV because the misclassification rate decreased considerably, indicating that it widened the lineal separation among classes in the projection hyperspace. K-NN models showed no difference when preprocessing with SNV and the algorithm performed the poorest. The best number of neighbors from cross-validation was initially spotted as one, three, and seven, with a total of 39 misclassified kernels. When tested on the validation set, seven neighbors gave the best classification accuracy of 17.5% (Table 2).

Our study gave better results than those achieved by Wang et al. (2002b) with soybean seeds and artificial neural networks (ANN), a method which usually leads to similar results to SVM. They could only achieve classification accuracies on the upper 60%. On the other hand, the previous study by Wang et al. (2001) with wheat kernels led to similar classification accuracy by PLS-DA (>97%).

Table 2
Misclassified corn kernels in the heat-damage study for each tested algorithm (misclassified seeds/total seeds belonging to that class).

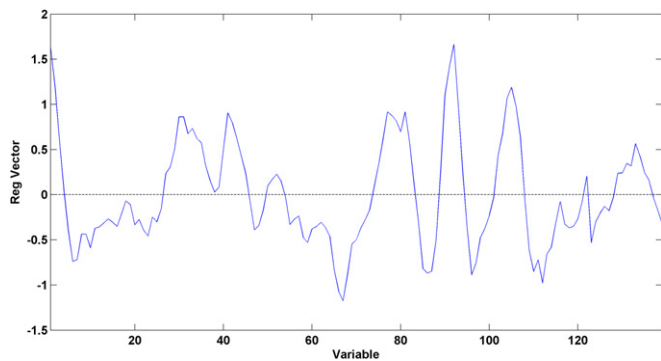| Algorithm	Raw spectra			SNV preprocessing		
	Damaged	Sound	Total (%)	Damaged	Sound	Total (%)
PLS-DA	0/51	1/52	1.0	0/51	2/52	1.9
SIMCA	3/51	2/52	4.9	3/51	13/52	15.5
LS-SVM	6/51	9/52	14.6	2/51	3/52	4.9
K-NN	10/51	8/52	17.5	10/51	8/52	17.5

Fig. 1. Regression coefficients (regression vector) from the PLS-DA model showing the largest absolute values in the carbohydrate region.

They could obtain classification accuracies above 96% using just two wavelengths. After examination of Fig. 1, we could suggest that working with wavelengths from ranges of 1250–1400 nm (carbohydrate regions with high regression coefficient values) would also provide satisfactory classification accuracies. On the other hand, the best two-wavelength models from Wang et al. (2001) involved the wavelengths 985, 1,050, 1,550, and 1,575 nm. Those wavelengths are mainly absorbed by protein. The differences in the relevant wavelengths between Wang et al. (2001) and our regression model may be due to the heating method (hot air drying vs microwaving).

3.2. Frost damage

Differences between frost damaged and sound corn kernels were not discernible by examining the spectra and this was confirmed in the discrimination results (Table 3). Classification by NIRS was unsuccessful for all the discrimination methods. Sound and frost-damaged kernels could not be differentiated and SNV led to almost all the damaged kernels to be classified as sound for SIMCA and LS-SVM. The optimal PLS-DA model used five latent variables with raw data and three latent variables with SNV preprocessing. Three and four PCs modeled frost and sound kernels respectively for SIMCA modeling. Both classes had very similar PRESS and eigenvalues, which can be anna indicator of the similarity of both classes. SNV preprocessing reduced the required number of PCs (Six for frost damaged and five for sound), but the misclassification rate remained the same. SNV preprocessing brought the two classes closer, as all the misclassified kernels were damaged classified as sound. K-NN with both five and seven neighbors gave the results shown in Table 3 ($K = 7$ when preprocessing with SNV).

As maize seeds mature and dehydrate, they undergo a series of physiological changes called acquisition of desiccation tolerance (Bewley and Black, 1994). The lack of acquisition of desiccation tolerance and thus, seed immaturity, are critical factor in determining the extent of freezing injury in corn (Hartwigsen, 1999; Kiesselbach and Ratcliff, 1920). In this study seeds were subjected

Table 3
Misclassified corn kernels in the frost-damage study for each tested algorithm (misclassified seeds/total seeds belonging to that class).

Algorithm	Raw spectra			SNV preprocessing		
	Damaged	Sound	Total (%)	Damaged	Sound	Total (%)
PLS-DA	6/12	4/13	40.0	6/12	2/13	32.0
SIMCA	6/12	4/13	40.0	10/12	0/13	40.0
LS-SVM	6/12	4/13	40.0	11/12	0/13	44.0
K-NN	5/12	5/13	40.0	3/12	7/13	40.0

Table 4
Misclassified corn kernels in the viability study for each tested algorithm (misclassified seeds/total seeds belonging to that class).

Algorithm	Raw spectra			SNV preprocessing		
	Non-viable	Viable	Total (%)	Non-viable	Viable	Total (%)
PLS-DA	125/242	30/82	47.8	80/242	44/82	38.3
SIMCA	112/242	37/82	46.0	126/242	32/82	48.8
LS-SVM	133/242	28/82	49.7	124/242	43/82	51.5
K-NN	116/242	37/82	47.2	119/242	42/82	49.7

Table 5
Misclassified soybean seeds in the viability study for each tested algorithm (misclassified seeds/total seeds belonging to that class).

Algorithm	Raw spectra			SNV preprocessing		
	Non-viable	Viable	Total (%)	Non-viable	Viable	Total (%)
PLS-DA	150/299	32/80	48.0	142/299	37/80	47.2
SIMCA	157/299	29/80	49.1	247/299	16/80	63.4
LS-SVM	166/299	37/80	53.6	133/299	35/80	44.3
K-NN	178/299	30/80	54.9	120/299	44/80	43.3

to an artificial frost when they were immature and, consequently, damage should have been extensive. However, recent studies assign less importance to the characteristics of the endosperm in preventing or reducing the severity of frost damage in corn (Woltz et al., 2006). The embryo and embryo axis should have the most severe symptoms of freezing, when compared to the endosperm, and detection could be difficult.

3.3. Viability

Discrimination of seed viability was not successful for either corn or soybeans, for all algorithms used (Tables 4 and 5). The elevated misclassification rates showed that discriminations were more random and no difference was detected by NIRS. Both PLS-DA and SIMCA models required four latent variables or PCs. With use of SNV one latent variable more was utilized for PLS-DA (5 latent variables) and up to seven and eight PCs for SIMCA models, but without enough improvement to lead to a successful application. Preprocessed spectra brought the two classes (viable and non-viable) together for SIMCA models, similar to the case of frost-damaged and sound spectra. Reported K-NN results were achieved using a single neighbor.

4. Conclusions

Among the three tested applications, only the discrimination of heat-damaged corn kernels was feasible. The achieved results were similar to that of Wang et al. (2001) for wheat kernels, although in our study the carbohydrate region was the most relevant for the classification according to the regression coefficients from the PLS-DA model, perhaps due to the difference in methods used to induce heat damage. Developing classification models using fewer wavelengths from the carbohydrate region (i.e. stepwise or genetic algorithm wavelength selection) may be feasible, especially using filter instruments. However, since the regression vector coefficients for the other wavelengths are not zero, classification accuracies could decrease. Furthermore, most of current monochromators or diode array instruments work on proprietary software which deal with full wavelength range models. For this application, PLS-DA performed the best, and more complex methods such as LS-SVM with RBF kernel mapping performed worse. This indicated the

most powerful methods may not always yield the best results. SNV preprocessing, which has been useful for quantitative applications, did not lead to improve the accuracies overall. SNV preprocessing only benefited the LS-SVM algorithm, increasing the separation of classes in the higher dimension space where the data are mapped by the RBF kernel. K-NN, although performing the worst, resulted in overall discrimination accuracies very close to LS-SVM. Although the method is rarely used in NIRS studies, it may have potential to be used in more applications because of the convenience of optimizing a single parameter (the number of neighbors).

Discrimination of frost-damaged corn kernels was not possible with NIRS, even with the use of non-linear methods such as LS-SVM. Frost-damaged soybean discrimination by NIRS has been reported to be successful (accuracy over 90%) using ANN (Wang et al., 2002b). However, the embryonic tissue in corn is only a relatively small volume of the seed while in soybeans it occupies the entire seed. This raises the question of how sensitive NIRS is to detect damage in single seed or how extensive the damage must be in order to be detected.

Because extensive damage may affect seed viability, the test of viability discrimination served to test the hypotheses that NIRS accuracy in discriminating damaged seeds depends on seed viability (the higher the damage, the higher the possibility the seed is not viable). No differences were detected between sound corn and soybean seeds (viable) and naturally dead or abnormal (non-viable) with any of the tested algorithms. Because the misclassified seeds from each class were random and no differences were observed between viable, abnormal, or dead seeds, the idea of carrying out a three-way classification need not to be considered for future research. This indicates that NIRS can discriminate significant damage in seeds in terms of physical and chemical changes induced by the damage, but cannot detect changes merely caused by the death of the seed (i.e. changes in water binding, oxidation of lipids, changes in protein, lack of cell metabolic activity). Seed aging has been shown to be tracked by NIRS (Kusama et al., 1997), but the threshold that separates aged and non-viable seeds is not differentiable by conventional NIRS. However, seeds selected for the study were visibly healthy. A large proportion of seeds from the accessions used in this study had an unhealthy appearance to the naked eye and were therefore not selected for inclusion in the study. It is unknown if those were not viable and if they could be easily discriminated by NIRS. If that was the case, NIRS could still be used for gross screening. Further research could be done using chemical imaging technologies on corn kernels and analyzing pixels from the germ.

References

Armstrong, P.R., 2006. Rapid single-kernel nir measurement of grain and oil-seed attributes. Applied Engineering in Agriculture 22 (5), 767–772.

Association of Official Seed Analysts, 2010. Rules for Testing Seeds. AOSA, Ithaca, NY.

Becker, H., 1998. Saving seeds for the long term. Agricultural Research, 12–13.

Bernal-Lugo, I., Leopold, A.C., 1998. The dynamics of seed mortality. Journal of Experimental Botany 49 (326), 1455–1461.

Bewley, J.D., Black, M., 1994. Seeds. Physiology of Development and Germination, second ed. Plenum Press, New York.

Campbell, M.R., Sykes, J., Glover, D.V., 2000. Classification of single- and double-mutant corn endosperm genotypes by near-infrared transmittance spectroscopy. Cereal Chemistry 77 (6), 774–778.

Cogdill, R.P., Hurburgh Jr., C.R., Rippke, G.R., 2004. Single-kernel maize analysis by near-infrared hyperspectral imaging. Transactions of the ASABE 47 (1), 311–320.

Cottrell, H.J., 1948. Tetrazolium salt as a seed germination indicator. Annals of Applied Biology 35 (1), 123–131.

DeVries, M., Goggi, A.S., Moore, K.J., 2007. Determining seed performance of frost-damaged maize seed lots. Crop Science 47, 2089–2097.

Dowell, F.E., Pearson, T.C., Maghirang, E.B., Xie, F., Wicklow, D.T., 2002. Reflectance and transmittance spectroscopy applied to detecting fumonisin in single corn kernels infected with *Fusarium verticillioides*. Cereal Chemistry 92 (2), 222–226.

Environmental Protection Agency (EPA), 2009. Major Crops Grown in the United States. Website information retrieved September 2011 from: http://www.epa.gov/agriculture/ag101/cropmajor.html.

Finney, E.E., Norris, K.H., 1978. Determination of moisture in corn kernels by near-infrared transmittance measurements. Transactions of the American Society of Agricultural Engineers 21, 581–584.

Hartwigsen, J.A. 1999. Changes in physiology, molecular biology, and biochemistry associated with maize seedlings germinated from freeze-damaged seed. Ph.D. dissertation. Iowa State University, Ames.

International Seed Testing Association, 1985. International rules for seed testing. Seed Science and Technology 13, 300–520.

Janni, J., Weinstock, B.A., Hagen, L., Wright, S., 2008. Novel near-infrared sampling apparatus for single kernel analysis of oil content in maize. Applied Spectroscopy 62 (4), 423–426.

Jiang, H.Y., Zhu, Y.J., Wei, L.M., Dai, J.R., Song, T.M., Yan, Y.L., Chen, S.J., 2007. Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance spectroscopy (NIRS) in maize (*Zea mays* L.). Plant Breeding 126, 492–497.

Kiesselbach, T.A., Ratcliff, J.A., 1920. Freezing Injury of Seed Corn. Research Bulletin 16. University of Nebraska Agricultural Experiment Station, Lincoln.

Kranner, I., Kastberger, G., Hartbauer, M., Pritchard, H.W., 2010. Noninvasive diagnosis of seed viability using infrared thermography. Proceedings of the National Academy of Sciences 107 (8), 3912–3917.

Krishnan, P., Joshi, D.K., Nagarajan, S., Moharir, A.V., 2004. Characterization of germinating and non-viable soybean seeds by nuclear magnetic resonance (NMR) spectroscopy. Seed Science Research 14, 355–362.

Kusama, T., Abe, H., Kawano, S., Iwamoto, M., 1997. Classification of normal and aged soybean seeds by discriminant analysis using principal component scores of near infrared spectra. Nippon Shokuhin Kogyo Gakkai-Shi 44 (8), 569–578.

Lukas, L., Hamers, B., De Moor, B., Vandewalle, J., 2003. LS-SVMlab Toolbox version 1.5.

Orman, B.A., Schumann Jr., R.A., 1992. Nondestructive single-kernel oil determination of maize by near-infrared transmission spectroscopy. JAOCS 69 (10), 1036–1038.

Pearson, T.C., Wicklow, D.T., Maghirang, E.B., Xie, F., Dowell, F.E., 2001. Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. Transactions of the American Society of Agricultural Engineers 44 (5), 1247–1254.

Spielbauer, G., Amstrong, P., Baier, J.W., Allen, W.B., Richradson, K., Shen, B., Settles, M., 2009. High-throughput near − infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. Cereal Chemistry 86 (5), 556–564.

United States Department of Agriculture (USDA), 2010. Website information retrieved January 2011 from: http://www.ers.usda.gov/Briefing/Corn/.

United States Department of Agriculture (USDA)- Grain Inspection, Packers, and Stockyards Administration (GIPSA), 2001. U.S. Corn Inspection. Document retrieved January 2011 from: http://www.gipsa.usda.gov/GIPSA/documents/GIPSA_Documents/corninspection.pdf.

Wang, D., Dowell, F.E., and Chung, D.S., 2001. Assessment of heat-damaged wheat kernels using near-infrared spectroscopy. Presentation at the 2001 ASAE annual international meeting, Sacramento, Ca, July 30-August 1.paper Num. 01−6006.

Wang, D., Dowell, F.E., Dempster, R., 2002a. Determining vitreous subclasses of hard red spring wheat using visible/near-infrared spectroscopy. Cereal Chemistry 79 (3), 418–422.

Wang, D., Ram, M.S., Dowell, F.E., 2002b. Classification of damaged soybean seeds using near-infrared spectroscopy. Transactions of the American Society of Agriculture Engineers 45 (6), 1943–1948.

Weinstock, B.A., Janni, J., Hagen, L., Wright, S., 2006. Prediction of oil and oleic acid concentrations in individual corn (zea mays L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. Applied Spectroscopy 60 (1), 9–16.

Williams, P., Geladi, P., Fox, G., Manley, M., 2009. Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. Analytica Chimica Acta 653, 121–130.

Woltz, J.M., TeKrony, D.M., Egli, D.B., 2006. Corn seed germination and vigor following freezing during seed development. Crop Science 46, 1526–1535.