Journal of Stored Products Research 72 (2017) 1–10

Contents lists available at ScienceDirect

Journal of Stored Products Research

journal homepage: www.elsevier.com/locate/jspr

Detection and prediction of *Sitophilus oryzae* infestations in triticale via visible and near-infrared spectral signatures

M.K. Khedher Agha [a, b], W.S. Lee [a, *], C. Wang [c], R.W. Mankin [d], A.R. Blount [e], R.A. Bucklin [a], N. Bliznyuk [a]

[a] Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611, United States
[b] Department of Agricultural Machinery and Equipment, University of Baghdad, Al-Jadriya, Baghdad, Iraq
[c] Department of Statistics, University of Florida, Gainesville, FL 32611, United States
[d] USDA-ARS, Center for Medical Agricultural and Veterinary Entomology, 1700 SW 23rd Drive, Gainesville, FL 32608, United States
[e] Department of Agronomy, NFREC, University of Florida, Marianna, FL 32446, United States

ARTICLE INFO

Article history:
Received 30 December 2014
Received in revised form
21 December 2016
Accepted 27 February 2017

Keywords:
Degree of infestation
Rice weevil
Spectral signature
Spectroscopy
Stepwise multiple linear regression

ABSTRACT

Triticale is a hybrid of wheat and rye grown for use as animal feed. In Florida, due to its soft coat, triticale is highly vulnerable to *Sitophilus oryzae* L. (rice weevil) and there is interest in development of methods to detect early-instar larvae so that infestations can be targeted before they become economically damaging. The objective of this study was to develop prediction models of the infestation degree for triticale seed infested with rice weevils of different growth stages. Spectral signatures were tested as a method to detect rice weevils in triticale seed. Groups of seeds at 11 different levels (degrees) of infestation, 0–62%, were obtained by combining different ratios of infested and uninfested seeds. A spectrophotometer was used to measure reflectance between 400 and 2500 nm wavelength for seeds that had been infested at different levels with six different growth stages from egg to adult. The reflectance data were analyzed by several generalized linear regression and classification methods. Different degrees of infestation were particularly well correlated with reflectances in the 400–409 nm range and other wavelengths up to 967 nm, although later growth stages could be detected more accurately than early infestation. Stepwise variable selection produced the lowest mean square differences and yielded a high R^2 value (0.988) for the 4th instars, pupae and adults inside the seed. Models were developed to predict the level of infestation in triticale by rice weevils of different growth stages. Overall, this study showed a great potential of using reflectance spectral signatures for detection of the level of infestation of triticale seed by rice weevils of different growth stages.

1. Introduction

Triticale (*X Triticosecale Wittmack*) is a robust, disease-resistant wheat-rye hybrid that is well adapted to drought and difficult soils (Salmon et al., 2004). Modern cultivars provide humans and farm animals a valuable source of essential amino acids and vitamin B (Tsen, 1974). The production of triticale around the world in 2012 was 13.7 million metric tons (FAOSTAT, 2012). In the U.S. the production from triticale in 2012 was 78 thousand metric tons (NASS, 2012).

The combination of a soft seed coat and high protein content makes stored triticale highly vulnerable to insects, including the rice weevil, *Sitophilus oryzae (L.)* (Dobie and Kilminster, 1978), especially in warm climates, like Florida, where populations can grow rapidly. These vulnerabilities could be reduced by developing a method for early detection and targeting of insect infestations in stored triticale. Early detection of rice weevil can be difficult because the larvae feed hidden inside the kernels and there are no visible external indicators of damage until the adult emerges (as shown in Fig. 1). Nonvisual methods to detect insects inside grain kernels include acoustic sensing, X-ray, nuclear magnetic resonance, and visible and infrared spectroscopy (Mankin and Hagstrum, 2011; Singh et al., 2009; Newey et al., 2008; Neethirajan et al., 2007; Tigabu et al., 2004; Dowell et al., 1998). For this research, the detection of rice weevil larvae and later stages in triticale was studied by applying visible (380–750 nm) and near-

* Corresponding author.
E-mail address: wslee@ufl.edu (W.S. Lee).

http://dx.doi.org/10.1016/j.jspr.2017.02.005
0022-474X/

a b

Fig. 1. Triticale seed samples in a sample holder, a. Zero% infestation as a control, b. 62.5% infestation of larvae 2nd instar (L2) growth stage of rice weevils.

infrared (NIR, 750–2500 nm) spectroscopy, which enables analysis of reflectance characteristics of grain kernels at different insect infestation stages. NIR spectroscopy utilizes that bonds between atoms of biological molecule functional groups (C–H, O–H, N–H, and S–H), including cuticular hydrocarbons, absorb significant energy at NIR wavelengths (Bokobza, 1998; Dowell et al., 1998; Newey et al., 2008). Consequently differences among chemical constituents of test specimens can be reflected in different spectral signatures. Williams and Norris (2001) noted, for example, that grain had peaks for starch at 970, 1200, 1450, and 1950 nm, and for protein at 2080 and 2180 nm. It was of interest to develop a spectroscopic method that predicts percentage of infestation of different instars in a triticale sample based on measurements of spectral reflectance at wavelengths where the greatest differences between undamaged and insect-damaged kernels occur. In previous studies, Ridgway and Chambers (1996) reported that NIR spectroscopy was useful to rapidly detect grain weevil infestations from two different varieties of wheat. In the standard normal variate (SNV) transformation and detrended spectra, the pupae stage showed higher absorbance increase than the larvae stage at the grain moisture band (1932 nm), when compared to uninfested samples. The same trend was observed at the grain starch band (2092 nm). A second derivative spectra at 2328 and 2062 nm was useful to identify infestation status of a single kernel. Dowell et al. (1998) investigated absorbance differences in wheat kernels infested by three different grain insects (i.e., rice weevil, the lesser grain borer, and the Angoumois moth) using an NIR spectrometer. They found that identification accuracy of infested kernels was not affected by moisture, protein contents or wheat class, however larval size affected the sensitivity of the detection system. Paliwal et al. (2004) investigated the potential of NIR spectroscopy in detecting different life stages of *Sitophilus oryzae* (rice weevil) and *Rhyzopertha dominica* (lesser grain borer) in infested wheat samples at different infestation rates, and reported that insect species could be better distinguished at higher infestation rates. However, they found that determination of infestation rates would be difficult at lower infestation rates than 25%. Starch absorption at 1884 and 2102 nm, and cuticular lipids absorption at 1132 and 1668 nm were also reported. Singh et al. (2009) found that starch molecules showed absorption between 1100 and 1300 nm in wheat kernels. They used a binary classification model to differentiate between undamaged kernels and kernels with damage caused by rice weevil and three other stored product insects. The accuracy of this method ranged from 73% to 100%, depending on the insect. Peiris et al. (2010) reported that undamaged and Fusarium head blight damaged kernels of wheat could be distinguished by comparing the second-derivative near-infrared spectra of the kernels in bands at 1160–1220 nm and 1395–1440 nm. They predicted that undamaged and damaged kernels could be distinguished for other small grains as well. Siuda et al. (2010) investigated infrared measurements of starch content for four classes of infestation of Fusarium head blight in wheat: C (control), 1 (light infestation less than 15%), 2 (strongly infested up to 50%), and 3 (very strong infestation more than 50%). They found that the thousand kernels weight (TKW), Hagberg falling number (HFN), protein and starch content, and sedimentation value (SV) decreased as the infestation percentage increased.

Although there have been many studies for detection of various insect infestations in wheat, no similar studies have been conducted with triticale. The objective of this study was to investigate the difference between the infested and sound kernels using NIR spectroscopy, and to identify the best wavelengths that determine degree of infestation.

2. Materials and methods

2.1. Triticale seed samples

The triticale seed variety, Trical 342, used for this experiment was grown at the North Florida Research and Education Center (NFREC), Quincy, Florida, USA from October to May 2012. This seed was harvested, then thrashed and cleaned at NFREC. The seed was sieved for 5 min using a Ro-Tap sieve shaker (Octagon 2000, Endecotts Limited, London, UK) with a set of the U. S. standard testing sieves with 5.66, 4.75, 2.36, 1.00, 0.71, and 0.50 mm opening sizes.

2.2. Seed infestation

Infested kernels were obtained by adding 1200 adult rice weevils to 350 g triticale seed, holding them for 5 d in glass jars in a conditioning chamber at 24 °C–26 °C and 60%–65% relative humidity. The jars were covered with a fine screen and two-sheets of filter paper to allow the air to exchange without allowing insects to escape from or get in the jars. The rice weevils were obtained from a colony reared at the U. S. Department of Agriculture - Agricultural Research Service - Center for Medical, Agricultural, and Veterinary Entomology (USDA-ARS-CMAVE) in Gainesville, FL, USA. The adults were removed from the jars and the infested seed (seed with eggs only) was mixed with sound (healthy) seed to create 11 different

degrees of infestation (DI), where the mixing was on weight basis between sound and infested seeds. The actual mixing ratios were from 0% to 100% with increments of ten percent.

Then, 60 g aliquots of each mixing ratio (including 0% DI, as a control) were placed in transparent plastic containers, each with a cover containing a fine metal screen at its center. The infested seeds were kept in the containers for 40 d to create different life cycle stages of egg, larva, pupa, and pre-emerge adult (adult inside seed). The zero percentage representing the sound seed that served as the experiment control (uninfested seed). Those stages represent a storage environment similar to an actual storage situation. Six different stages were grown over time until adults emerged from the seed, as shows in Table 1.

The seed samples were infested with 11 different infestation percentages from 0% to 62% with three replications for a total of 33 samples for each growth stage. These were replicated six times to produce the six growth stages from egg to pre-emergence adult. The sound seed was tested using the same procedure as for the uninfested seed. These samples provided a wide variety of degrees of infestation that represented normal infestation conditions as shown in Table 1.

The separation between the stages was done by estimating the stages based on the mean duration of each instar by counting the days after the weevil adults were placed with the seed (Perez-Mendoza et al., 2004; Sharifi and Mills, 1971). The actual percentage of infestation for each category of DI was measured using the manual counting basis procedure (Siuda et al., 2010; Wang et al., 2011), where infested seed was observed visually through counting one by one the emerged adults and the seed that had holes produced, during the emergence of the rice weevil adults from the seed.

The counting results for both seed holes and emerged adults were compared to find if two eggs exist in one seed. Equation (1) was used to calculate the degree of infestation (DI) as percentage.

$$\mathrm{DI}\ (\%) = \frac{\text{total number of seeds with holes in the sample}}{\text{total number of seeds in the sample}} \times 100 \quad (1)$$

2.3. NIR spectral measurements

Diffuse reflectance was measured for triticale seed samples of 15 g kernels with known infestation percentage placed in a sample holder as shown in Fig. 1. The spectral measurements were conducted using a spectrophotometer (Cary 500, Varian Inc. Palo Alto, CA, USA), with a mercury lamp, which have a wide range of wavelength in the visible (VIS) and near-infrared (NIR) ranges as light source. An integrating sphere (DRA-CA-5500, Labsphere Inc. North Sutton, NH, USA) with an interior coating of polytetrafluoroethylene (PTFE) was attached to the spectrophotometer. The lamp was warmed-up for 30 min before any measurements and a reference standard spectrum was collected each testing day using a standard PTFE disk. The wavelength range used for this experiment was from 400 nm to 2500 nm with one nm increments.

The seed was placed on a sample measurement holder, which had a diameter of 38 mm with a quartz glass cover to keep the seed in a fixed position during the test without affecting the measurement, as shown in Fig. 1.

2.4. Spectral data analysis

Spectral data were smoothed using the Savitzky-Golay method (Savitzky and Golay, 1964) using 3rd order polynomial in MATLAB R2010a (Mathworks Inc. Natick, MA, USA) to remove the noise from the data. The frame size for the smoothing was set to 41 points based on preliminary testing. After smoothing, the data were analyzed using a suite of linear regression and classification methods. Continuous response (percentage of infestation), multiple linear regression (MLR) with variable selection and regression trees were applied. In the case of discrete responses using four labels for the degree of infestation after binning, where the bin was chosen according to infestation percentages, with zero% infestation as 1st bin, 6.3% to 18.7% infestation as 2nd bin, 24%–41.4% as 3rd bin and 50.3%–62.5% as 4th bin (as will be explained in a later section), ordinal logistic regression and classification trees were used to give classification results. Due to the large number of highly correlated reflectance values for neighboring wavelengths used as predictors, dimension reduction of the predictor space was conducted to aggregate the reflectance values over 10-wavelength consecutive non-overlapping bins. The original wavelengths ranged from 400 to 2500 nm. By taking the mean reflectance for each 10 wavelengths, 210 resulting reflectance values were acquired, named as W40, W41, …, and W249, representing the average reflectance of 400–409 nm, 410–419 nm, …, and 2490–2499 nm, respectively (only the last one original wavelength, 2500 nm, being discarded).

2.4.1. Regression

For each insect growth stage, the response variable was the percentage of infestation while the covariates were the reflectance values averaged across the 10-wavelength bins. To identify parsimonious models with good predictive performance, automatic model search was performed using five-fold cross-validation with the Statistical Analysis System (SAS) software. A heuristic (stepwise) model selection was carried out in SAS (Proc GLM SELECT) to search for subsets of predictors achieving high R^2 values and low root mean square error (RMSE) values using a five-fold cross validation method to validate the selection, represented in the results later as cross validation of the predicted residual sum of squares statistic (CV PRESS). The selection using a stepwise method was governed by two parameters: first significance level for entry (SLE), and second significance level to stay (SLS), where the selection was stopped when the significance level was reduced. The initial parameter values, SLE = 0.15 and SLS = 0.10, were selected according to values suggested by SAS as a starting point, then a trial and error method was used to obtain the best prediction. The selection results were corroborated using exhaustive model

Table 1
Spectral measurement dates of each growth stage of rice weevils inside triticale seed.

Date	Number of Days	Growth Stages
6/29/2012	0	Adult mixed with sound seed
7/4/2012	5	Adult taken out (only seed with egg left in cans)
7/7/2012	8	Egg and Larvae 1st instar
7/16/2012	17	Larvae 2nd instar
7/18/2012	19	Larvae 3rd instar
7/22/2012	23	Larvae 4th instar
7/25–26/2012	26–27	Pupae
8/1/2012	33	Adult inside seed (pre-emerge adult)
8/4/2012	36	Adult outside

Table 2
Planned and measured degree of infestation (DI) of rice weevils inside triticale seed.

Degree of infestation category (DI %)											
Planned	0	10	20	30	40	50	60	70	80	90	100
Measured	0	6.3	11.4	18.7	24.0	30.8	36.3	41.4	50.3	59.0	62.5

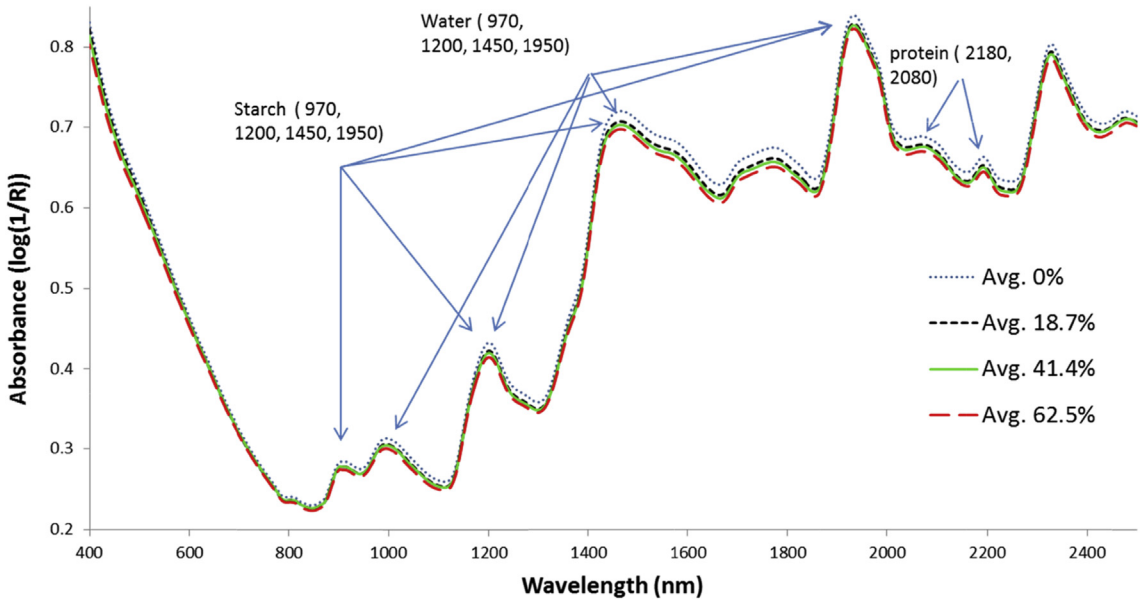


Fig. 2. Absorbance for triticale seed with different degrees of infestation percentages of rice weevils, using average spectra of six growth stages.

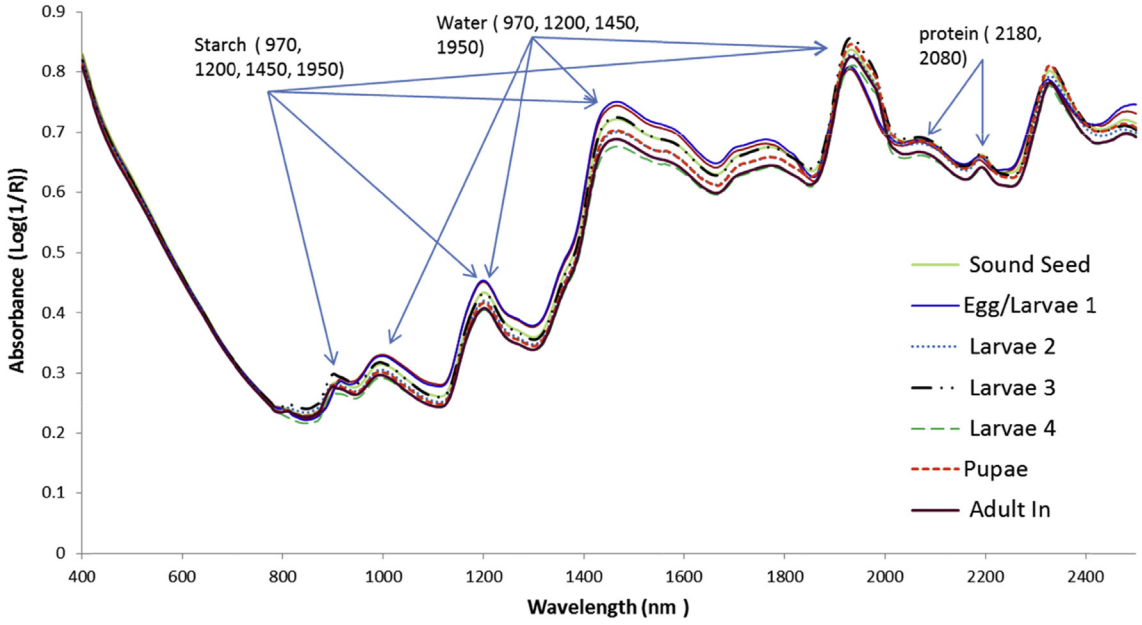


Fig. 3. Absorbance of triticale seeds with different growth stages of rice weevils, using average spectra of 33 samples for each growth stage with eleven infestation percentages.

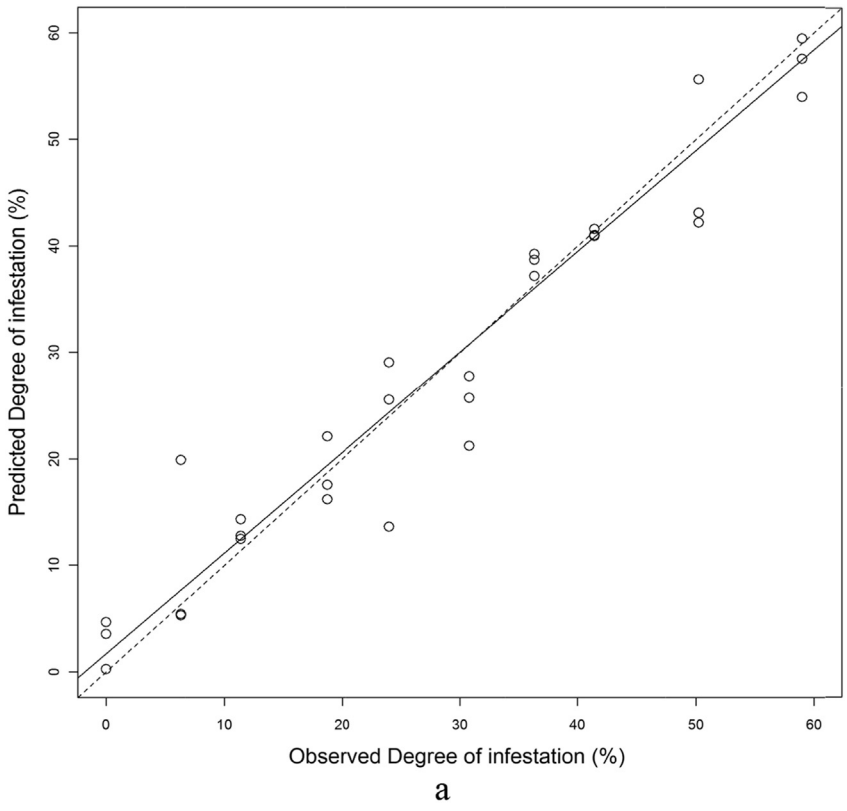
enumeration in R package 'leaps' (http://www.r-project.org). The out-of-sample model performance was assessed using five-fold cross-validation (Hastie et al., 2008) with random approximate

Table 3
Linear regression (SAS GLM Select) output evaluation results for each growth stage.

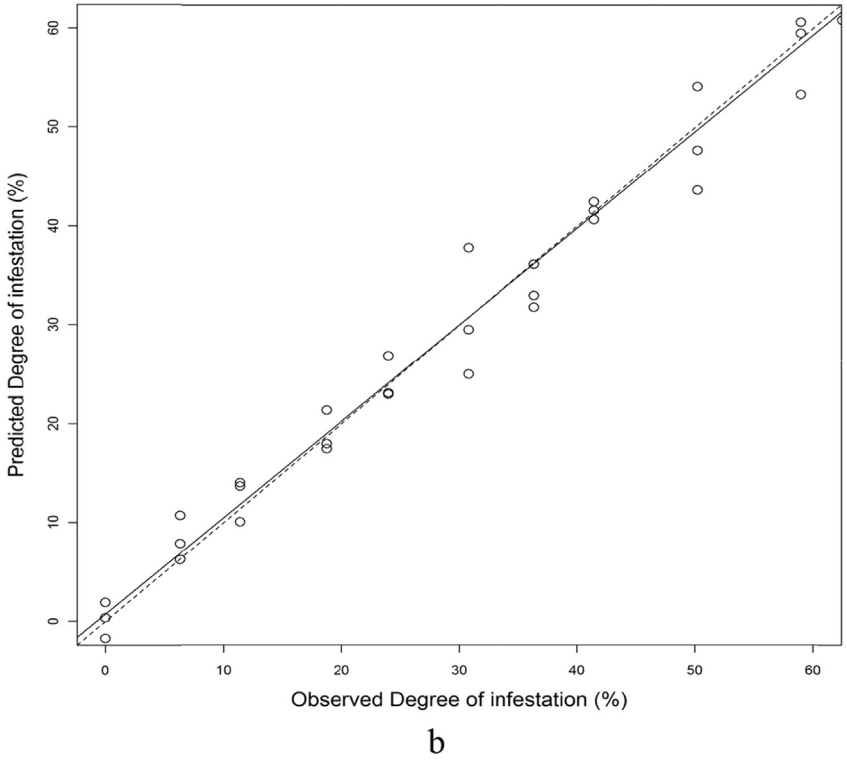
Stage	R^2	CV PRESS	RMSE (%)	Number of predictors
Egg L1	0.945	1219.7	6.4	14
L2	0.951	1010.9	6.0	14
L3	0.947	1274.1	6.1	13
L4	0.988	357.2	3.2	16
Pupa	0.957	906.8	5.1	10
Adult-In	0.976	463.3	3.7	9
All Stages	0.725	28523	10.9	16

Table 4
GLM Select chosen (selection) parameters for each growth stage.

Stages	Parameters
Egg L1	W40 W41 W65 W80 W114 W143 W145 W197 W153 W196 W199 W207 W217 W222
L2	W40 W41 W61 W62 W68 W88 W108 W109 W113 W114 W116 W133 W194 W217
L3	W40 W41 W45 W46 W71 W72 W79 W88 W212 W237 W239 W246 W247
L4	W40 W41 W43 W47 W51 W55 W97 W100 W104 W107 W113 W122 W125 W128 W137 W247
Pupa	W40 W49 W57 W59 W74 W76 W156 W189 W197 W209
Adult-In	W40 W47 W108 W116 W117 W122 W127 W132 W133
All Stages	W40 W41 W42 W54 W57 W59 W64 W137 W146 W202 W210 W211 W213 W215 W240 W244



a



b

Fig. 4. **a**. A scatterplot of predicted vs. observed degrees of infestation for Egg L1 growth stage. Predicted degrees of infestation were obtained after variable selection (row 1 of Table 4). The solid line is the simple linear regression fit to the pairs of observed and predicted infestation values ($R^2 = 0.945$). Dashed line is the 45-degree line representing the perfect fit ($R^2 = 1$). **b**. A scatterplot of predicted vs. observed degrees of infestation for Pupae growth stages. Predicted degrees of infestation were obtained after variable selection (row 5 of Table 4). The solid line is the simple linear regression fit to the pairs of observed and predicted infestation values ($R^2 = 0.957$). Dashed line is the 45-degree line representing the perfect fit ($R^2 = 1$).

Table 5
Best subsets of size less than 5 for each stage.

Subset size	1	2	3	4
Egg and Larvae 1st	W40	W46 W48	W40 W103 W113	W40 W79 W103 W113
Larvae 2nd	W40	W40 W41	W40 W74 W78	W40 W41 W75 W78
Larvae 3rd	W40	W40 W41	W40 W106 W112	W40 W86 W107 W112
Larvae 4th	W40	W40 W41	W40 W61 W89	W40 W61 W103 W113
Pupae	W40	W40 W193	W40 W157 W197	W40 W48 W159 W190
Adult-In	W40	W40 W51	W93 W107 W112	W40 W78 W107 W112

20% splits (randomly splitting the sample into 5 approximately equal size groups, then using each of the 5 groups as validation data and the other 4 as training data). Regression tree models (implemented in R package 'tree') were validated by the same cross-validation technique.

2.4.2. Classification

To apply classification algorithms, the group categories can be defined by binning the percentages of infestation into categories. Both the 11 original and 4 binned infestation levels were considered in ordinal logistic regression models. Variable selection methods for binned reflectances, as well as dimension reduction approaches based on principal components regression were used to define covariates.

3. Results

3.1. Absorbance spectra of different degrees of infestation and different growth stages of rice weevils

Eleven different degrees of infestation (DI) were created based on weight basis between sound and infested seeds. The planned mixing ratios were from 0% to 100% with increments of ten percent, however the actual ratios were different than those planned when the degree of infestation was measured manually, as shown in Table 2. This might be because the environmental conditions might have not been optimal.

The absorbance curves for triticale seed with different infestation percentages are shown in Fig. 2 using average of 18 spectra from six growth stages and three replications each. In Fig. 2, the water, starch and protein absorption bands (Williams and Norris, 2001) are marked. The curve shows also that there is a distinctive difference in the region between 1400 and 1900 nm, caused by the well-known water and starch absorption bands. Fig. 3 shows absorbance of triticale seeds with different growth stages of rice weevils, using average spectra of 33 samples for each growth stage with eleven infestation percentages. Spectral differences among different growth stages can be observed, especially in 1400–1900 nm range.

3.2. Multiple linear regression with stepwise model search

As described in the method section, SAS (PROC GLM SELECT) was used to search for subsets of predictors achieving high R^2 values and low root mean square error (RMSE) values (Table 3). Unfortunately, the small number of data points and replications were insufficient to train the classifier to achieve acceptable out-of-sample misclassification rate. Because of this, the focus was on regression, rather than classification-based procedures.

The Larvae 4th instar (L4) yielded a highest value of R^2 of 0.988 and a lowest RMSE value (3.2%). The L4 stage used many prediction parameters as shown in Table 4, where 16 parameters were used. Table 3 shows that the late stages of L4, Pupae, and Adult-In yielded higher R^2 and lower RMSEs than the other early stages. Similar wavelength ranges were chosen as parameters in many growth stages as shown in Table 4, where the W40 appeared in all stages and other wavelengths such as W213, W217 and W247 appeared in two stages only. It was observed that more than 10 wavelengths were needed to produce good prediction results for almost all stages, as shown in Table 4.

The relationships between predicted and observed degree of infestation for triticale seed infested with rice weevils of two different growth stages are shown in Fig. 4a and b based on Tables 3 and 4 Larvae 4 t h instar produced the best prediction with a high R^2 of 0.988, as shown in Table 3. Meanwhile, the other stages also yielded a good prediction with high R^2 values above 0.94.

3.3. Exhaustive search and predictive quality of the linear regression models

Function 'regsubsets' from 'leaps' package in R was used for variable selection (best subset size < 5) using exhaustive search with aggregated reflectances as predictors. The best models are summarized in Table 5, which shows that W40 is a significant predictor for all stages (except Adult-In, with three subset size and egg and larvae 1st with two subset size). Predictive performance on a left-out data subset was assessed using a five-fold cross validation with random approximate 20% splits. This was carried out 1000 times, and the results are summarized in Table 6. A trend of decreasing RMSEs can be observed in Table 6 as the insect stages grow from Egg to Adult-In and this trend is clearer as the selected variable numbers increases. As the subset size increased and the growth stage developed, the RMSE decreased. The lowest RMSE value was 4.9% from a larger subset and a last growth stage.

Because there were 210 binned reflectances, increasing the size of the best subset in excess of four was computationally expensive. To mitigate this problem, the number of candidate variables was reduced from 210 to 50 by prescreening the ones that had the greatest correlations (in absolute value) with the response. Again 1000 rounds of five-fold cross-validation were performed. Boxplots of the test RMSEs for each stage are shown in Fig. 5.

3.4. Regression trees

Regression tree models were fitted in R (library 'tree') with the percentage of infestation as a numerical response and aggregated

Table 6
Mean values (standard errors) of the RMSEs of the entire cross validation for each size (<5).

Subset size	1	2	3	4
Egg and Larvae 1st	15.0 (0.40)	13.0 (0.35)	11.7 (0.41)	10.4 (0.39)
Larvae 2nd	18.3 (0.65)	15.8 (0.64)	13.8 (0.51)	13.1 (0.63)
Larvae 3rd	14.3 (0.34)	13.1 (0.38)	12.0 (0.31)	10.6 (0.38)
Larvae 4th	11.9 (0.23)	10.0 (0.24)	8.7 (0.18)	7.4 (0.16)
Pupae	14.2 (0.33)	9.6 (0.26)	8.0 (0.19)	6.0 (0.10)
Adult-In	8.6 (0.15)	6.5 (0.11)	5.7 (0.09)	4.9 (0.09)

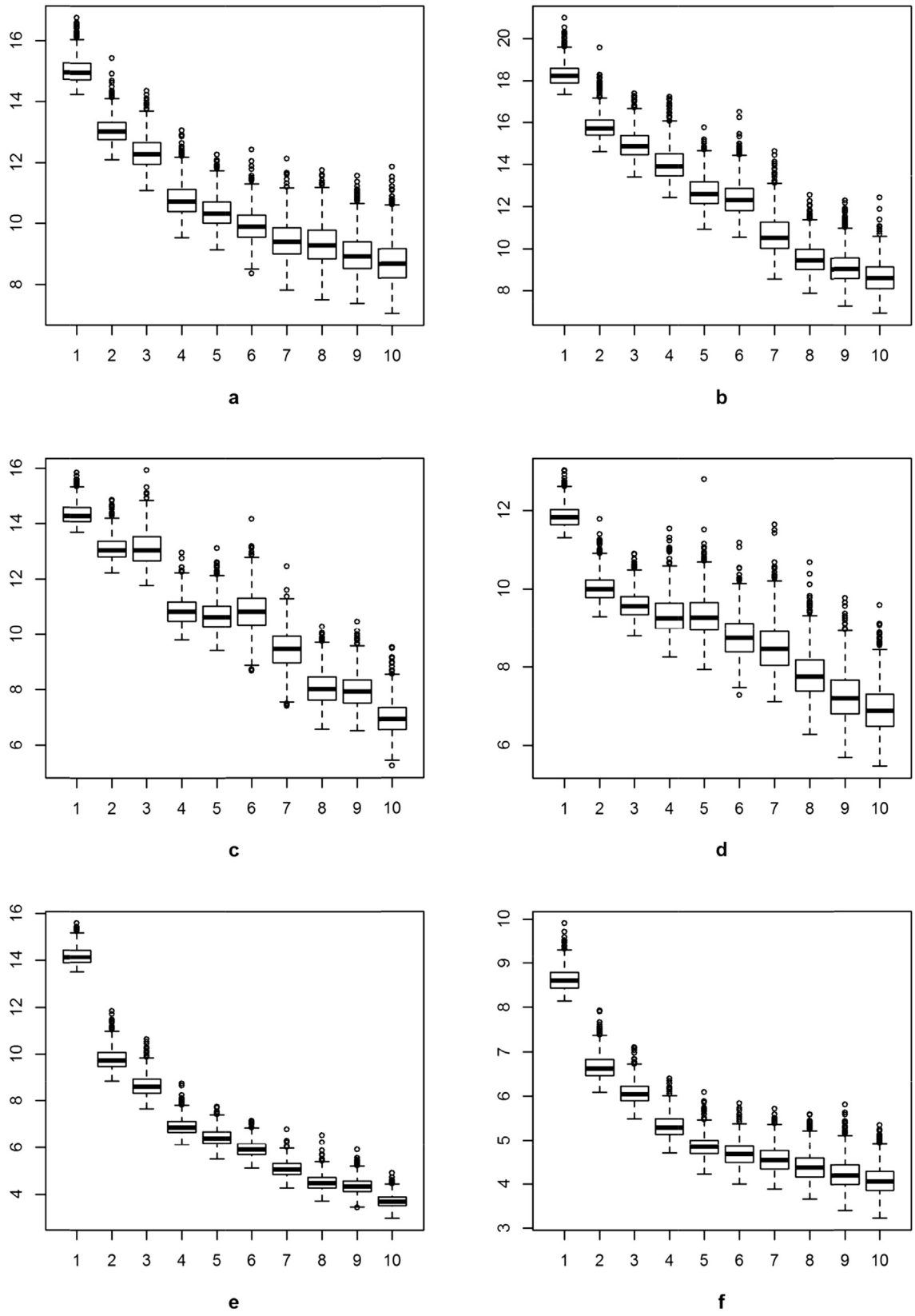


Fig. 5. Boxplots of test data RMSE across 1000 rounds of cross validation for different growth stages, where X axis is the subset size and Y axis is the RMSE: a. Egg L1, b. L2, c. L3, d. L4, e. Pupae and f. Adult-In.

Table 7
Wavelengths selected by regression tree and corresponding RMSEs.

Stage	Wavelengths selected	RMSE (%)
Egg and Larvae 1st	W40 W41W86 W70	10.5
Larvae 2nd	W44 W69 W40 W218	12.7
Larvae 3rd	W40 W78 W82 W242	10.5
Larvae 4th	W40 W197 W56 W139	7.6
Pupae	W40 W111 W233 W88	11.2
Adult-In	W40 W44 W97	7.0

Table 8
RMSE comparison of different selection methods for each growth stage.

Method of prediction	GLM Select	Subset Selection Size 4[a]	Regression Tree
Stages	RMSE	RMSE	RMSE
Egg L1	6.4	10.4	10.5
L2	6.0	13.1	12.7
L3	6.1	10.6	10.5
L4	3.2	7.4	7.6
Pupa	5.1	6.0	11.2
Adult-In	3.7	4.9	7.0

[a] See Table 6. This is where subset selection size 4 is from.

reflectances as predictors. The wavelengths used to define where splits occurred among different groups are listed in Table 7, which shows that the early growth stages had a higher RMSE than late growth stages. A shown in Fig. 6, performance of the tree models is comparable to that of linear regression models with variable selection. Fig. 6 shows selected wavelengths for two different growth stages (Egg L1 and Pupae). The predictions for the degree of infestation are given in the end of the tree, as explained with more details in Table 7.

Table 8 compares the performance of different selection methods using RMSE, and shows that the GLM SELECT yielded the lowest RMSE compared with the other methods for all the stages, where the error was as low as 3.7 RMSE value for Adult-In stage, and highest was 6.4 RMSE value for Egg L1 stages. It also shows that the second best method of selection was the exhaustive search, which produced an RMSE lower than the regression tree method for all the stages except for L2 and L3 stages. Also, the early growth stages had higher RMSE values than the late growth stages for all the methods that used in this research.

4. Discussion

The main objective of this research was to develop prediction models of the infestation degree for triticale seed infested with rice weevils of different growth stages using the visible and NIR spectral signature. A large number of reflectance values (33 spectra from each growth stage) showed differences between the growth stages of rice weevils infesting the triticale seed using an average spectrum of eleven infestation percentages. As shown in Fig. 3, the reflectance between 900 and 1150 nm shows moderate differences due to the water and starch absorption bands. In all the wavelength regions, the reflectances between 1450 to 1850 nm and 1900–1980 nm exhibit bigger differences among all the stages based on visual inspection of the spectra, particularly at 1500 and 1900 nm, due to the combined effect of the water and starch absorption bands. Based on Pena (2004), the starch and water represent about 60% and 13.5% in triticale, respectively and the later growth stages would consume more substrates. There was a smaller reflectance difference among the growth stages at protein absorption bands between 2080 and 2200 nm due to the smaller percentage of protein in the seed, approximately 12.5% compared with starch (Pena, 2004).

It was found that heuristic (stepwise) model produced the best prediction model that yielded lowest error and with a high R^2 of 0.988. This model was able to predict the infestation well for each growth stage separately. In addition, the prediction model for all the stages yielded an RMSE of 10.9%. A second method of analysis was an exhaustive search that produced RMSE lower than the

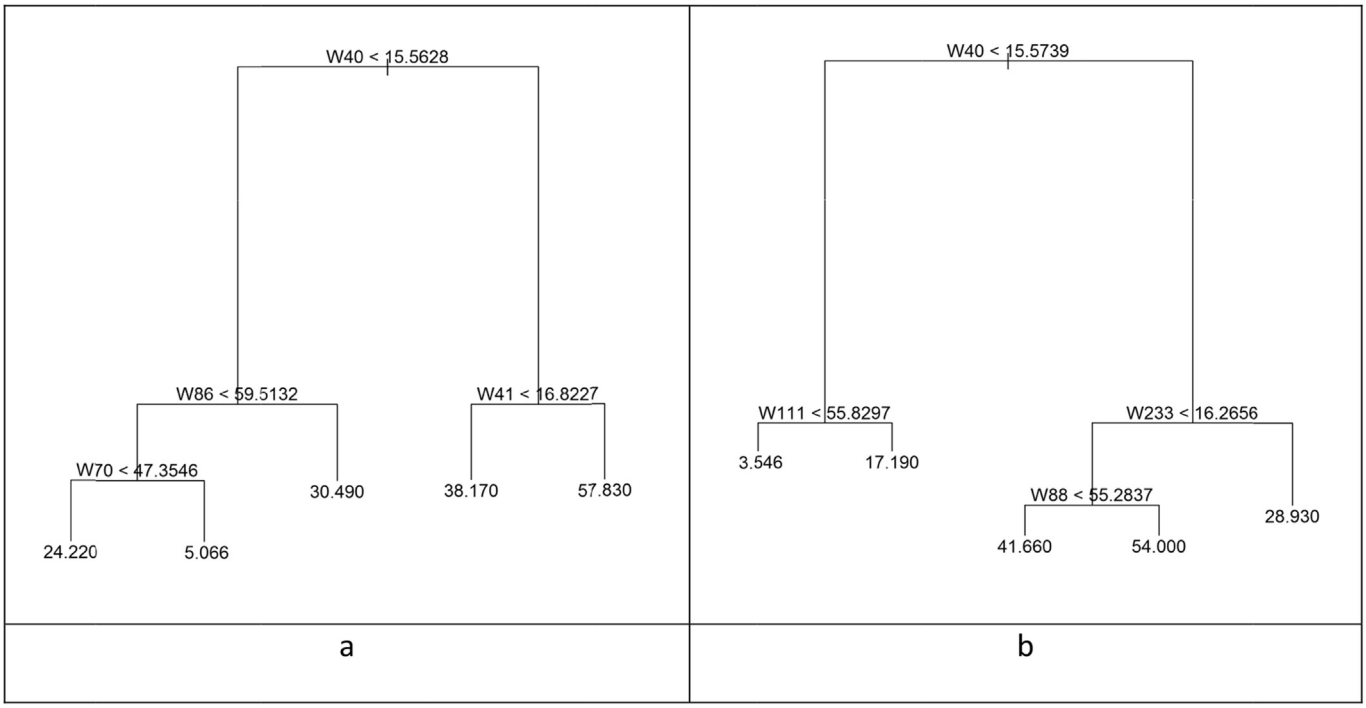


Fig. 6. Tree structures for two different growth stages: a. Egg L1, b. Pupae.

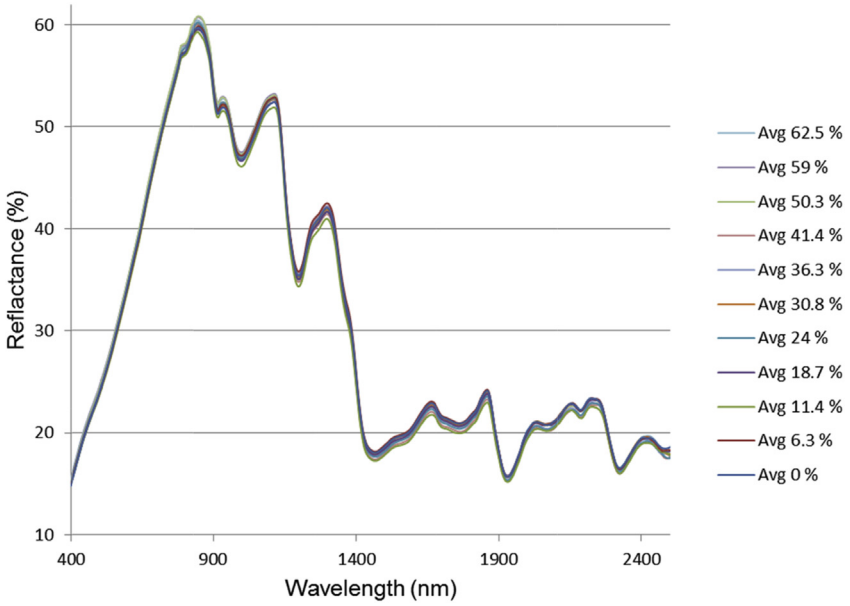


Fig. 7. Reflectance of Egg-larvae 1st instar after smoothing.

regression tree method. Different growth stages affected the selection accuracy, where the late growth stages yielded a lower RMSE than the early growth stages for all the methods. This suggested that there was a relationship between the size of the insect inside the seed and the chemical composition of the seed. As the size of the insects increased, they ate more from the seed, which caused a reduction of the seed mass and changed the relative chemical composition of the seed. This change was detected using the spectral signature. The prediction was more accurate compared with the early growth stages such as egg L1, L2 and L3 that consumed less from the seed. This led to a conclusion that the insect size affected the prediction accuracy, which means that the accuracy was increased in prediction models with later growth stages.

In Table 6, it was observed that the RMSE for all stages dropped considerably as the subset size increased from 1 to 4. At higher values of subset size, the RMSE for most stages stabilized. This corroborates the earlier finding from the stepwise search that the models achieving better out-of-sample predictions were larger models; that is, the models with ten or so carefully chosen predictors were not overfitting the data.

In Table 7 of the regression tree results, it is likely that this was due to a decrease in kernel mass relative to an increase in larval mass, as explained earlier. Also, there were some exceptions as the pupa stage had a higher RMSE than the other early stages. This could be due to characteristics of the pupa stage produced by the cocoon materials that affected the reflectance results.

Also, Table 7 shows that the wavelength W40 (an average of

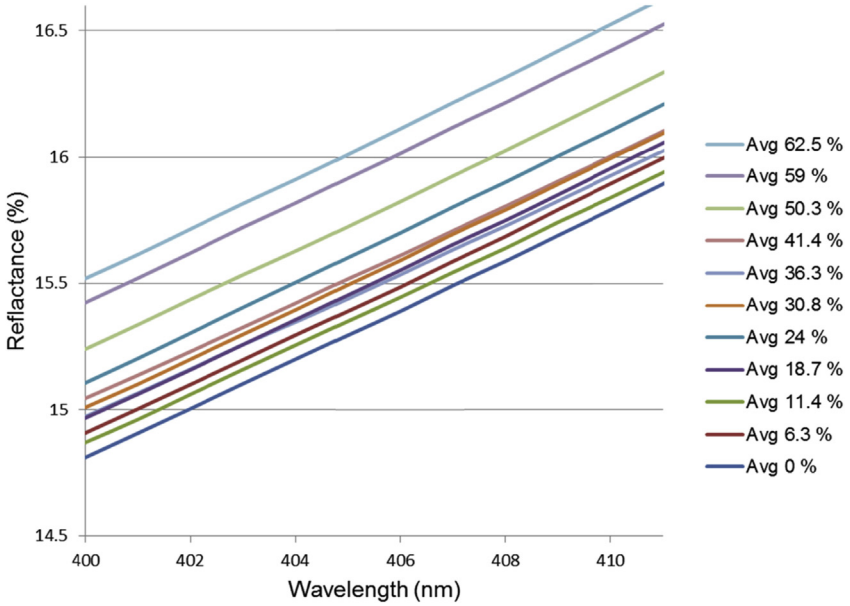


Fig. 8. Zoomed in Egg-larvae 1st reflectance graph (reflectance of wavelength 400–411 nm).

400–409 nm) was selected for all the stages. To take an in-depth look at the variables being selected, Fig. 7 shows the overall smoothed reflectance of egg-larvae 1st instar, the lines are average of 3 replications. The curve in Fig. 7 shows 11 percentages of infestation from 0%, until 62.5%. Fig. 8 shows the zoomed in graph of wavelength 400–411 nm, which is approximately in correspondence to the W40 in the model selection, regression tree, and best subset selection. It can be seen that curves show a positive relationship between reflectance and infestation percentages. The wavelengths between 400 and 409 nm (W40) were selected as the common wavelengths for each growth stage and for all the methods, since the curves tend to be well-separated at this wavelength range as shown in Fig. 8. Close inspection of the other stages gives the same conclusion and this might also be the reason of W40 being the most significant predictor for all stages.

In comparing the performance of different selection methods using RMSE as shown in Table 8, the error would be related to the size of the insect and the size of the seed materials that had been eaten by the late-growth larvae. This led to the late growth stages predictions having less error than the early stages.

Acknowledgements

The authors would like to thank Mr. James Colee, Department of Statistics at the Institute of Food and Agricultural Sciences (IFAS), University of Florida, Ms. Betty Weaver, Biological Science Lab technician in ARS, CMAVE, USDA, Gainesville, FL, Dr. Ce Yang and Dr. Han Li in the Agricultural and Biological Engineering Department, University of Florida for their help in this research.

References

Bokobza, L., 1998. Near infrared spectroscopy. J. Near Infrared Spectrosc. 6 (1), 3–17.

Dobie, P., Kilminster, A.M., 1978. The susceptibility of triticale to post-harvest infestation by *Sitophilus zeamais* Motschulsky, *Sitophilus oryzae* (L.) and *Sitophilus granarius* (L.). J. Stored Prod. Res. 14, 87–93. http://dx.doi.org/10.1016/0022-474X(78)90003-6.

Dowell, F.E., Throne, J.E., Baker, J.E., 1998. Automated nondestructive detection of internal insect infestation of wheat kernels by using near-infrared reflectance spectroscopy. J. Econ. Entomology. 91, 899–904.

FAOSTAT, 2012. Food and Agriculture Organization of the United Nations Statistics Division. http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E (Accessed 29 August 2014).

Hastie, T., Tibshirani, R., Friedman, J., 2008. The Elements of Statistical Learning Data Mining, Inference, and Prediction, second ed. Springer.

Mankin, R., Hagstrum, D., 2011. Acoustic monitoring of insects. In: Hagstrum, T.W.P.D.W., Cuperus, G. (Eds.), Stored Product Protection, Publication. Kansas State Univ. Press, Manhattan, KS, 1566–S222.

NASS, 2012. The 2012 Census of Agriculture. http://quickstats.nass.usda.gov/results/744C6455-A9E7-3573-8462-57EB04DD7F57 (Accessed 29 August 2014).

Neethirajan, S., Karunakaran, C., Jayas, D.S., White, N.D.G., 2007. Detection techniques for stored product insects in grain. Food control. 18, 157–162.

Newey, P.S., Robson, S.K.A., Crozier, R.H., 2008. Near-infrared spectroscopy identifies the colony and nest of origin of weaver ants, *Oecophylla smaragdina*. Insectes Sociaux 55, 171–175.

Paliwal, J., Wang, W., Symons, S.J., Karunakaran, C., 2004. Insect species and infestation level determination in stored wheat using near-infrared spectroscopy. Can. Biosyst. Eng. 46, 7.17–7.24.

Peiris, K.H.S., Pumphrey, M.O., Dong, Y., Maghirang, E.B., Berzonsky, W., Dowell, F.E., 2010. Near-infrared spectroscopic method for identification of Fusarium head blight damage and prediction of deoxynivalenol in single wheat kernels. Cereal Chem. 87, 511–517.

Pena, R.J., 2004. Food uses of triticale. In: Mergoum, M., Gomez-Macpherson, H. (Eds.), Triticale Improvement and Production. FAO, Rome, pp. 37–44.

Perez-Mendoza, J., Throne, J.E., Baker, J.E., 2004. Ovarian physiology and age-grading in the rice weevil, *Sitophilus oryzae* (Coleoptera: Curculionidae). J. Stored Prod. Res. 40, 179–196. http://dx.doi.org/10.1016/S0022-474X(02)00096-6.

Ridgway, C., Chambers, J., 1996. Detection of external and internal insect infestation in wheat by near-infrared reflectance spectroscopy. J. Sci. Food Agric. 71, 251–264.

Salmon, D.F., Mergoum, M., Gomez-Macpherson, H., 2004. Triticale production and management. In: Mergoum, M., Gómez-Macpherson, H. (Eds.), Triticale Improvement and Production. FAO, Rome, pp. 27–34.

Savitzky, A., Golay, M.J.E., 1964. Smooting and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.

Sharifi, S., Mills, R.B., 1971. Developmental activities and behaviour of the rice weevil inside wheat kernels. J. Econ. Entomology. 64, 1114–1118.

Singh, C.B., Jayas, D.S., Paliwal, J., White, N.D.G., 2009. Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging. J. Stored Prod. Res. 45, 151–158. http://dx.doi.org/10.1016/j.jspr.2008.12.002.

Siuda, R., Grabowski, A., Lenc, L., Ralcewicz, M., Spychaj-Fabisiak, E., 2010. Influence of the degree of fusariosis on technological traits of wheat grain. Int. J. Food Sci. Technol. 45, 2596–2604. http://dx.doi.org/10.1111/j.1365-2621.2010.02438.x.

Tigabu, M., Oden, P.C., Shen, T.Y., 2004. Application of near-infrared spectroscopy for the detection of internal insect infestation in *Picea abies* seed lots. Can. J. For. Res. 34, 76–84.

Tsen, C.C.H., 1974. Triticale: first Man-made Cereal. The American Association of Cereal Chemists, St. Paul, Minnesota.

Wang, J., Nakano, K., Ohashi, S., 2011. Nondestructive detection of internal insect infestation in jujubes using visible and near-infrared spectroscopy. Postharvest Biol. Technol. 59, 272–279.

Williams, P., Norris, K., 2001. Near Infrared Technology in the Agricultural and Food Industries, second ed., vol.40. The American Association of Cereal Chemists, Inc, St. Paul, Minnesota, pp. 239–280.