

## Identification of species and geographical strains of *Sitophilus oryzae* and *Sitophilus zeamais* using VIS/NIR hyperspectral imaging technique

Cao, Y.#<sup>1</sup>, Zhang, C.<sup>2</sup>, Chen, Q.S.\*<sup>2</sup>, Li, Y.Y.<sup>1</sup>, Qi, S.<sup>2</sup>, Tian, L.<sup>1</sup>, Ren, Y.L.<sup>3</sup>

<sup>1</sup>Academy of State Administration of Grain, Beijing, China, 100037

<sup>2</sup>School of Food & Biological Engineering, Jiangsu University, Zhenjiang, China, 212013

<sup>3</sup>Murdoch University, 90 South Street, Murdoch, WA, Australia, 6150

\*Corresponding author, Email: qschen@ujs.edu.cn.

#Presenting author, Email: cy@chinagrains.org.

DOI: 10.14455/DOA.res.2014.39

### Abstract

Identifying stored-product insects is essential for granary management. Automated, computer-based classification methods are rapidly developing in many areas. A hyperspectral imaging technique could potentially be developed to identify stored-product insect species and geographical strains. This study tested and adapted the technique using four geographical strains of each of two insect species, the rice weevil and maize weevil to collect and analyze the resultant hyperspectral data. **SULTS:** Three characteristic images that corresponded to the dominant wavelengths, 505, 659 and 955 nm were selected by multivariate image analysis. Each image was processed and 22 morphological and textural features from regions of interest were extracted as the inputs for an identification model. We found the back propagation neural network model to be the superior method for distinguishing between the insect species and geographical strains. The overall recognition rates of the classification model for insect species were 100 and 98.13% for the calibration and prediction sets respectively, while the rates of the model for geographical strains were 94.17 and 86.88% respectively. This study demonstrated that hyperspectral imaging, together with the appropriate recognition method, could provide a means for identifying insects and become a useful tool for identification of *Sitophilus oryzae* and *Sitophilus zeamais* to aid in the management of stored-product insects.

Keywords: stored-product insects, geographical strains, rice weevil, maize weevil, identification, hyperspectral imaging

### 1. Introduction

Cereal grains are the major source of food for humans and many domesticated animals. In both developed and developing countries, most cereal grains are stored in granaries for years before being sent to consumers. This can simply be the result of over-production or it can be caused by other factors, such as differences in the point and time of production and consumption. Consequently, populations of stored-product insects can develop causing deterioration in grain quantity and quality. A survey from the Ministry of Agriculture of the People's Republic of China shows that pests cause approximately 6.5–11.4% of the loss of rural stored-grain in China every year (Yuan, 2002). The total economic losses due to stored-product insects and microorganisms in grains and oilseeds could be in the millions of dollars annually (Neethirajan et al., 2007).

Stored-product insects have been widely researched in an effort to better manage them around the world. There are approximately 600 known species of stored-product insects in the world, 224 of which are found in China. Different species of stored-product insect can cause different levels of grain loss and each requires a different approach for control. Different insects with different behavior require different control strategies. For example, the red flour beetle, *Tribolium castaneum* (Herbst) (Coleoptera: Tenebrionidae) shows better flight

capabilities than the confused flour beetle, *Tribolium confusum* Jacquelin du Val (Coleoptera: Tenebrionidae), so relatively quicker action may be required to control it. On the other hand, the rice weevil, *Sitophilus oryzae* (Linnaeus) (Coleoptera: Curculionidae) is more resistance to phosphine than its sibling species, the maize weevil, *Sitophilus zeamais* Motschulsky (Coleoptera: Curculionidae). Furthermore, insect species that develop in different regions can live in environments that are exposed to different chemical levels and, therefore, have differing levels of resistance (Jagadeesan et al., 2012). Identification the insect species and strain can, therefore, provide an index for pesticide dosing appropriately for control. Identifying stored-product insects is an essential task for granary managers as it is necessary to determine correct control procedures and whether the captured species are invasive.

Optical microscopes have always been used to observe the external characteristics of insects for identification, including their overall color, shape, texture and modality of body parts. Sometimes dissection plays an indispensable role in finding the distinction between species. For example, the aedeagus is used to differentiate between male rice and maize weevils, being smooth and shiny on its dorsal surface in the rice weevil, but having two longitudinal grooves on its dorsal surface in the maize weevil. Although the traditional method for insect identification is accurate, it has obvious disadvantages, as it is time-consuming, labor-intensive, and requires professional knowledge and experience. In the last 30 years, electron scanning microscopy, molecular marker technology, and analytical biochemical techniques have been applied to insect taxonomy. This has resulted in the determination of classificatory characteristics at the microscopic or molecular level (Yao et al., 2012; Brown et al., 2012; Hidayat et al., 1996; Peng et al., 2003). However, these methods are complex and require sample pretreatment with chemicals.

However, automatic computer-based classification for insects has rapidly been developed in recent years as a safer alternative. Al-Saqer et al. (2011) identified pecan weevils through geometric feature extraction and image processing techniques based on template matching. Larios et al.(2010) employed Haar random forest features and a Support Vector Machine (SVM) for spatially matching kernels to efficiently identify stonefly species. Similarly, Wen et al. (2009) extracted local features as the input for an orchard insect classifier. There are also a number of reports on the classification of stored-product insects based on machine vision (Neethirajan et al., 2007; Gan and Qu, 2006). There is some research on the potential of near-infrared (NIR) spectroscopy for examining insects; for example, for the detection of insect infestation externally and internally in wheat (Ridgway and Chambers, 1996; Baker et al., 1999; Maghirang et al., 2003). Baker et al. (1999) showed that 100% of kernels containing parasitoid pupae, weevil pupae, and weevil larvae were correctly identified using a NIR system (400–1700 nm). However, for more accurate classification of parasitoid larvae, the detection threshold had to be adjusted. Maghirang et al. (2003) accurately classified sound kernels (unaffected by insects) and kernels containing live insects at various stages of growth at a rate of about 80%. Identifying stored-grain insect species by NIR spectroscopy has also been studied by Dowell et al. (1999) A classifier with Partial Least Squares (PLS) regression or neural network was used to classify 11 insect species into primary and secondary pests with 96% accuracy, whilst in an 11-way classification, the accuracy ranged from 30–100% (Dowell et al.,1999). These studies demonstrate that machine vision and near-infrared spectroscopy, combined with the appropriate pattern recognition methodology, are feasible tools for accurate insect detection. However, a conventional imaging system operating at visible wavelengths is incapable of distinguishing between specimens of similar color and shape, classifying complex objectives, and analyzing differences in specimen components. On the other hand, the spectroscopy technique can only perform at some levels, and not always accurately. Therefore, hyperspectral imaging, which integrates the two classical optical

sensing technologies of imaging and spectroscopy, are brought together to identify insects with greater accuracy.

Hyperspectral imaging systems with optically tunable filters can record images at hundreds of contiguous wavelengths (narrow spectral resolution) in the form of a hypercube (three-dimensional hyperspectral data). The hyperspectral imaging technique, therefore, has the ability to rapidly and simultaneously monitor morphological characteristics. As an emerging non-destructive and reagent-less analytical technique, hyperspectral imaging has been widely employed in diverse fields, such as agriculture (Cogdill et al., 2004; Migdall et al., 2009), pharmaceuticals (Roggo et al., 2005) and the food industry (Maftoonazad et al., 2011; ElMasry et al., 2011; Zhao et al., 2009). NIR hyperspectral imaging has also been attempted for the detection of insect-damaged wheat kernels (Singh et al., 2010; Singh et al., 2009), but reports on the identification of insect species using hyperspectral imaging have not been found.

In this study, we used a visible/near infrared (VNIR) hyperspectral imaging system to identify insect species and geographical strains of those species. The sibling species, rice weevil and maize weevil, were used as the subjects of the study. Insect specimens were created first, then hyperspectral images of the specimens were collected using the adjusted apparatus and lastly, a series of image processing and model recognition methods were applied for the actual insect classification. Through the data processing, three optimal wavelength bands were extracted to construct multi-spectral imaging systems that offer a rapid, simple and convenient instrument for stored-product insect detection, which could be used to help make decisions in the management of stored products.

## **2. Materials and Methods**

### *2.1. Insects*

All insect species used in this experiment were acquired from the Academy of State Administration of Grain, China. To ensure consistency within the experiment, adult rice weevils and maize weevils of different geographical strains were kept for 5-15 days after emergence. About 200 first-generation adults were mixed with 60 g whole wheat for 4 days in a growth chamber at 30°C and 70% r.h. The adults were then sieved out to get wheat that contained only insect eggs, which was then also kept at 30°C and 70% r.h. As soon as the adult insects emerged from the wheat kernels, they were transferred to new whole wheat and kept at the same temperature and relative humidity conditions for another 5–15 days before imaging. A total of 400 weevil adults (Table 1) were selected for hyperspectral imaging. The entire process was completed at the Academy of State Administration of Grain, Beijing, China.

### *2.2. Hyperspectral imaging system*

A VNIR hyperspectral imaging system at Jiangsu University was used for this study. The system consisted of a spectral imager (Model no. V10EB1610, Spectral Imaging Ltd., Finland) with a 50 mm f/1.1 C-mount lens (Model no. 1-19179, Navitar OptiStar, New York, USA), a 150W quartz-halogen illuminator (Fiber-Lite DC-950, Dolan-Jenner Industries Inc., USA), a linear motorized slide (SC300-1A, Zolix Instruments Co. Ltd, Beijing, China), data acquisition and pre-processing software, and a personal computer. The imaging spectrograph operates in line-scanning mode, giving a 430–960 nm spectral range. It can disperse the light entering through the entrance slit (30 µm) into different wavelengths instantaneously with a spectral resolution of 2.73 nm. The area detector converts incident photons into electrons to quantify the intensity of the acquired light and adopts a charge-coupled device (CCD) with a 1628×1236 pixel spatial resolution. A lens with a longer focal length was chosen to give a smaller object area for the small target samples and the linear motorized slide was used to

move the sample perpendicular to the scanning camera. All of these parts were mounted in an enclosure protected by a duralumin shield to avoid interference from external light. Figure 1A shows a schematic of the hyperspectral imaging system.

### 2.3. Image acquisition and pre-processing

To get the static images of stored-product insects, selected specimens were prepared as samples according to the following steps:

- The insect was killed by placing it in an oven at about 80°C for 10 seconds.
- Due to rigor mortis which often resulted in distortion and fragility of the body, the insect was then made pliable. This was achieved using a simple ‘relaxing chamber’. The chamber was constructed from a conical flask and a covered dish. The flask was filled with hot water to provide a damp environment and a piece of filter paper supporting the insect was placed between the flask and the dish for half an hour.
- The insect was finally fixed on a ready-made stand (Fig. 1B) with the legs and head placed in position to complete the specimen. The stand was a glass slide with ink daubed on the surface and transparent double-sided adhesive tape placed on top of that. The ink provided a dark background that could handle higher light intensities so that the specimen did not become overexposed, resulting in clearer images. We had previously tried black cardboard; however, it failed to create a clear distinction between the object and background at the long-wavelengths. Ink was therefore chosen as the background in this experiment. According to the steps above, 50 insects per geographical strain were pasted with the same orientation onto a single stand (Fig. 1C). In total there were eight stands containing 400 adult insect specimens prepared for the experiment.

The reflectance mode was used to acquire hyperspectral images of the insect samples. The insect specimens were passed through the spectral imager’s field of view at the optimized speed of 0.17 mm/s. The light intensity was set to 60% of the maximum and the camera exposure was set to 40 ms. These parameters were optimized through trial and error to ensure the best image quality, and SpectraCube software controlled the image acquisition process. Six hundred and eighteen images were captured for each sample in the 400–960 nm wavelength range (0.858 intervals) and with 1628×618 pixels spatial resolution. Thus, the hypercube corresponding to each sample was 1628×618×618 pixels. The information was stored on the computer for subsequent analyses.

As the light intensity was not consistent across all wavelengths and there was some dark current, a reflectance calibration had to be performed to calibrate the raw spectral image, which is actually the detector signal intensity. The calibration image ( $R$ ) was obtained based on a dark image and a white image using equation (1), and implemented in a Spectral Image Analyzer (Isuzu Optics, Inc., China). The dark image was obtained when the camera lens was completely covered and the white image was a picture of a white Teflon board obtained under the same conditions as the raw image.

$$R = \frac{I - B}{W - B} \quad (1)$$

Where  $I$  is the raw spectral image,  $B$  is the dark image, and  $W$  is the white image.

Although the light intensity and exposure time were modulated, the gray-scale value of the image had a nonhomogeneous distribution. Linear enhancement was used to improve the image quality. Linear transformation is frequently used to enhance images, mainly through

contrast manipulation, by applying a linear function. Contrast stretching can make the overall image clearer and so improve the usability of the image. Insects are relatively small organisms with complex structures and their legs and head are rarely fixed in the same position across specimens. To account for this variation, a morphological open algorithm was applied to the hyperspectral images to eliminate legs and head, and focus on the tergum of the thorax and abdomen. A threshold segmentation was then applied to exclude the background and to obtain the region of interest (ROI) for the test insect's image.

#### *2.4. Characteristic wavelength selection and feature extraction*

Hyperspectral imaging provides hundreds of continuous spectral bands of images, much larger than those bands of multispectral images. In most cases, variables based on conjoint wavelengths have similar spectral information and the neighboring band's images are highly correlated, which leads to multicollinearity. Multicollinearity means that the correlations among the independent variables are strong. The optimal wavelengths that relate most to the properties of interest must be selected and used as a reference to extract images that are characteristic of each species or strain. Moreover, hyperspectral data consist of images for each of the 618 bands, so the huge data set substantially increases the computational burden and unavoidably creates redundant information. The selection of optimal wavelengths that have characteristic information can be used to design a multispectral imaging system with the advantage of a simple structure and low cost.

A Multivariate Image analysis (MVI)(Singh et al., 2010) based on principle component analysis (PCA) was used to find the optimal wavelength. First, one of the hyperspectral images of an insect sample was converted to a binary image using global threshold segmentation and then labeled to create a labeled matrix, excluding the background. The three-dimensional hypercube was then reshaped into a two-dimensional array in which all the pixels reflected intensities of a sample rearranged into a column at each of the 618 wavelengths. This resulted in an array with a size of  $k \times 618$ , where  $k$  is the number of pixels enclosed in the labeled matrix. Next, PCA was applied to the two-dimensional array data. PCA is a mathematical procedure that decomposes the characteristic variables into several principal components (PCs) using an orthogonal transformation. This transformation is defined in such a way that the first PC accounts for as much of the variability in the data as possible and each subsequent component in turn has the highest variance possible under the constraint that it be orthogonal to the preceding components. The front three wavelengths corresponded to bigger coefficients for the first three principle components (PCs) and were selected as the characteristic wavelengths to be used for feature extraction.

After the characteristic wavelengths had been identified, the ROI images corresponding to the optimal wavelength for each insect sample were selected for feature extraction. Because of the biological variations among the test insects, each sample reflected and absorbed electromagnetic energy at a specific spectral band. In essence, the intensity value of each pixel of the picture corresponded to its spectral value at a specific spectral band. In other words, the intensity level of the picture can be used to characterize chemical and physical variations in the insect. There is a general belief that rice weevils are smaller than maize weevils. On close examination, there are distinguishing features on the pronotal punctures as well. On rice weevils, the pronotal punctures are separated by a flat, median, longitudinal puncture-free zone, while on maize weevils, pronotal punctures have no median puncture-free area and are almost equally spaced apart. The different distribution in punctures would certainly lead to changes in textural features (Hidayat et al., 1996). As a result, four morphological features were extracted from the ROI image for each insect sample at one optimal wavelength. These were: Area (w), Eccentricity (Ec), MajorAxisLength (Ma) and

MinorAxisLength (Mi). Six texture descriptors based on the statistical properties of the histogram of the ROI images at three optimal wavelengths were also extracted. These were: Mean (m), Standard Deviation ( $\sigma$ ), Smoothness (R), Third Moment ( $\mu_3$ ), Uniformity (U) and Entropy (e).

For the four shape features, w is the pixel number in the ROI; Ec, Ma and Mi are respectively the eccentricity, major axis length and minor axis length of the ellipse, whose secondary moment is the same size as the region. The six textural features are calculated according to the following equations:

$$m = \sum_{i=0}^{L-1} z_i p(z_i), \quad (2)$$

$$\sigma = \sum_{i=0}^{L-1} (z_i - m)^2 p^2(z_i), \quad (3)$$

$$R = 1 - \frac{1}{1 + \sigma^2}, \quad (4)$$

$$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i), \quad (5)$$

$$U = \sum_{i=0}^{L-1} P^2(z_i), \quad (6)$$

$$e = -\sum_{i=0}^{L-1} P(z_i) \log_2 p(z_i) \quad (7)$$

Where:  $p(z_i)$  is the probability of y density of a pixel's intensity distribution in a two-dimensional (2D) region that is the histogram of the intensity levels in this region

L is the number of possible intensity levels

m computes the average level of intensity in the image

$\sigma$  describes the variation of intensity around the mean

R is a measure of flatness of the histogram

$\mu_3$  is an indication of symmetry in the histogram

U is a measure of histogram uniformity and finally

e is an amount that reflects the complexity of the image texture

In addition, a ratio image was created by comparing two band ROI images from an insect sample to extract six textural features. A ratio image usually eliminates the disturbance of drop shadows and reflects the actual information of the sample. By calculating two wavelength images, some textural features changed. For example, by increasing the difference in mean gray value of two classes of images, this helped distinguish between the sample classes.

To test the usefulness of extracted features during insect species identification, an independent samples t-test was used to analyze the morphological data for all the insect samples. Group variables were defined according to insect species, in which the value 1 stood for rice weevils and 2 for maize weevils. The significance level was set to 0.05. Significant features constituted the data matrix column. Lastly, an  $n \times m$  sized array, where n is the number of

samples and  $m$  is the number of features (independent variables), was obtained as the input for the recognition model to identify the insect species and distinguish between the geographical strains.

### 2.5. Model development

This study used a Back Propagation Neural Network (BPNN) to develop identification models. BPNN is one of the most popular predictive tools at present and is widely used in various other fields (Kuo et al., 2013; Liu et al., 2013; Nagalakshmi et al., 2013). BPNN is an Artificial Neural Network (ANN) with three or more layers. Sample data imported into the network were sent from the input layer and through the intermediate layer to obtain a response signal from the output layer. Back propagation refers to the back-sending of the error between output error and correct error from output layers to the input layers to upgrade the weights and biases in the learning process of the ANN (Demuth et al., 1998). In the development of the BPNN model, a three-layer (input, hidden, and output) back propagation network with three hidden slabs was used and had a tan h transfer function. The epoch was set to 1000, a learning rate and momentum gene set to 0.1, and initial weight set to 0.3. All the data analysis, feature extraction and recognition algorithms were implemented in Matlab R2009b (Matworks Inc., Natick, MA, USA).

## 3. Results

### 3.1. Feature extraction

ROI image extraction was based on the gray level image of the test insect and the morphological opening was applied. A processing flow chart is shown in Figure 2. The structural element was a disk with a 9 pixel radius. Threshold segmentation using the Otsu method was then used to exclude the background and the ROI of the test insect image (the white area, Image D, Fig. 2) was extracted. The ROI image for feature extraction (Image F, Fig. 2) was acquired through algorithms between image D and gray scale images under the characteristic wavelength (Image E.).

The characteristic wavelengths selected in the MVI analysis were 505, 659 and 955 nm, as they have the maximal coefficient in the first three PC loadings (Fig. 3). Pictures of samples corresponding to the three characteristic wavelengths are shown in Fig. 4. Near-infrared spectroscopy is based on the molecular overtone and combination vibrations in organic compounds. In a previous study, by Ridgway et al. (1999), wavelengths 982 and 1014 nm were selected to develop a two-wavelength classification model for insect-damaged wheat kernels with near-infrared reflectance spectroscopy. Furthermore, Perez-Mendoza et al. (2003) found 890 nm to be one of the significant wavelengths for detecting insect fragments in wheat flour. It is thus clear that wavelengths around 890 and 980 nm are appropriate for insects. The difference in optimal wavelength may be due to the effect of wheat flour or kernels on the insect species. The bands of 505 and 659 nm may be sensitive to the color of the weevils, as there is a slight difference between the two species. Variance in the visible spectra was dominated by extremes in color development. The wavelengths 505 and 660 nm correspond to turquoise and red respectively, however, the insect bodies displayed brown or black, which are not caused by homogeneous light and might be the mixing of 505 and 660 nm. The ratio image was used to calculate gray values of the ROI corresponding to wavelengths of 955 and 659 nm.

Six textural features were then extracted from each characteristic wavelength ROI image (containing the ratio image). Four shape features were extracted from images corresponding to 955 nm, because the samples that were highly reflective were clearly distinguished from the background at 955 nm (Fig. 4D). Therefore, four shape features and 24 textural features

were extracted from each insect sample. This resulted in a 400×28 sample sized matrix for the subsequent analyses. In the sample matrix, there were 28 independent variables for each of the 400 insect samples. Independent samples t-test was used to analyze the matrix to test the usefulness of the 28 independent variables (Table 2). The result indicated that most of the variables, especially those based on the ratio images, were significant for the identification of the insect species, whereas five of the textural features ( $\sigma$ , R,  $\mu_3$ , U, e), based on the ROI image at 955 nm, did not reach a significant level. We attempted to exclude the six features extracted from images that corresponded to 955 nm and were left with a 400×22 sample sized matrix.

Normalization and PCA were performed in order. The score plot of the first three PCs derived from PCA is shown in Fig. 5. The first three PCs accounted for 70.41% of the overall variance (PC1=27.08%, PC2=24.90%, PC3=18.43%). As shown in Fig. 5, the two species are clearly separate in space and the majority of insects from each geographical strain congregate into separate groups in the three-dimensional score plot. It illustrates that there are obvious differences between the insects. First, the plot has only the first three PCs, which does not explain the useful information completely. Second, the plot only describes the score of PCs, and doesn't take account of the weight of each PC. Lastly, PCA is a linear transformation of the initial independent variables and does not yet consider the actual sample class. Therefore, other recognition model methods were developed for further analysis.

### 3.2. Identification of insect species

To develop the model, the sample was stochastically divided into an independent calibration set (three fifths of the insect samples) and a prediction set (two fifths of the insect samples). The calibration data set (240×22 sized) was used for model's development, while the prediction data set (160×22 sized) was used for the model's validation. The insect class (categorical variable) was labeled numerically (1 stood for rice weevils and 2 stood for maize weevils) and used as the dependent variable for the BPNN identification model.

The top PCs were extracted to be the inputs for the model and the optimal number of PCs was determined using the discriminating rate from the calibration set. The result was that the BPNN model with seven PCs performed with the highest recognition rate: 100% in the calibration set and 98.13% in the prediction set (Fig. 6). The detailed recognition result of the BPNN model with seven PCs for each class of insect in the prediction set is shown in Table 3. There was only one sample of rice weevil misclassified as maize weevil and two samples of maize weevil misclassified as rice weevil. This is a very promising result suggesting great potential in hyperspectral imaging systems combined with BPNN models for nondestructive identification and classification of insect species.

### 3.3. Classification of geographical strains of an insect species

For the classification of geographical strains of an insect species, rice and maize weevils were separately processed with each comprising four different geographical strains. As in the preceding test, BPNN was adopted to classify the insect strains. The recognition results with different numbers of PCs are shown in Figure 7. The optimal recognition models were acquired with 7 PCs for the rice and maize weevils. For the rice weevils, the optimal model classified the strains with 90.83% accuracy in the calibration set, and 86.25% in the prediction set. For the maize weevils, the accuracy was 97.50 and 90.00% in the calibration and prediction set, respectively. The overall accuracy for the entire set of insects in the experiment reached 94.17 and 86.88% in two sample sets. The detailed recognition results of the optimal BPNN model are given in Table 4 and 5. Table 4 shows the rice weevils from Australia are 100% accurately classified by the model and the misclassification of geographical strains occurred only among the three Chinese strains. The lowest recognition rate is 77.78% for SO-SC. Table 5 shows there

are several maize weevils in each strain that are misclassified by the model. Overall, the classification results for the rice weevils and maize weevils are similar. There is no doubt that the difference in characteristics between species is larger than that between strains. Additionally, the transportation and mixing of different grain loads which may carry insects from other areas is likely to reduce the occurrence of visibly distinctive strains. Consequently, the overall recognition rate of the strains would be lower than that of the species.

## **4. Discussion**

### *4.1. The data*

The major limitation of the hyperspectral imaging technique is the large size of the data set that demands proper processing to interpret the results accurately. In this study, three characteristic wavelengths were selected according to MVI, which greatly reduced the amount of data and improved the data processing speed. Features describing the shape and texture of the insects were extracted and screened for model recognition. PCA was also employed for dimension and noise reduction to simplify the BPNN model. This not only reduced the data processing time but also enhanced the model's robustness. It is hoped that a multispectral imaging system based on the three selected wavelengths will be developed for insect identification.

The identification of insect species is based on their morphology and their absorbance characteristics. That cuticular lipids have peaks in NIR regions corresponding to the C-H overtone regions has been reported by Ridgway and Chambers (1996), Dowell et al. (1998) and Baker et al. (1999). Different geographical strains of insect species may reveal some distinction that helps in the identification of insect strains. In addition, insect are complex organisms with subtle physical structures. It is assumed that the relationship between the insects and their hyperspectral data is non-linear. The overlap of samples in the score plot of PCA and the high recognition rate of the BPNN model, which is a complex nonlinear algorithm, are accordant with this hypothesis. The BPNN model used in this work has a triple-layer structure with three nonlinear hidden layers, and has weights and biases that are self-adjusted to assist analysis.

### *4.2. The technique*

As mentioned above, rice and maize weevils are both from the family Curculionidae and are very similar in morphology. Highly similar patterns of the resultant amplicons in DNA amplification fingerprinting (DAF) have confirmed that the two weevils are closely related (Brown et al., 2012). Indeed, it is difficult to identify the rice and maize weevils with an optical microscope, even for those who have worked in the field for years. Dissection and molecular methods are complex, need professional knowledge and are not online remote identification methods. In a study by Dowell et al. (1999) using NIR spectroscopy to identify stored-grain insect species, the classification accuracies for rice and maize weevil were 70% and 90%. Thus, NIR spectroscopy alone was not found able to classify closely related insects to the species level with a high degree of accuracy. However, hyperspectral imaging contains rich spectral and spatial information, and is particularly good at detecting and discriminating objects with even imperceptible differences. Our results have proven that the hyperspectral imaging technique is superior for distinguishing between very similar sibling insect species such as the rice and maize weevils. Although previous studies have shown that the most significant region for detecting insects within wheat kernels is in longer wavelengths, this work has shown the feasibility of visible and shorter wavelengths in NIR to identify insect species. This can cut costs because Vis/NIR hyperspectral systems are cheaper than NIR. In addition, hyperspectral imaging does not require as intense a source of light and has a higher level of imaging processing as machine vision (Zayas and Flinn, 1998).

### 4.3. Future work

In this study, the tergum of the thorax and abdomen was selected as the region of interest (ROI), and a morphological method was used to automatically extract the ROI. This is enough for insect species and strain detection, but is far from satisfactory for insect gender determination. For identifying differences between male and female insects, the sternum of the abdomen is recommended as the ROI, since genitalia are the primary difference between male and female specimens.

Further research is needed on other stored-product insects or granary weevils to develop a profile database of species and of variation within species for convenient identification. In the short-term, the insect sample preprocessing procedure must be simplified. It is possible to develop a multi-spectral imaging system coupled with an image processing method in the granary to detect and identify the insect species and number of individuals to facilitate the automatic selection of management strategies as needed. Alternatively, a portable insect identification tool could be designed to be integrated with the daily work of the granary.

## 5. Conclusion

The results show that hyperspectral imaging coupled with a BPNN technique is able to classify two species of weevil with an overall discriminating rate of 98%, and to distinguish each geographical strain at an accuracy greater than 77.78%. Three characteristic wavelengths, 505, 659 and 955 nm, were selected by MVI analysis for data reduction, and can be considered as the base bands for multispectral imaging. The development of a multispectral imaging system with the appropriate image processing and model recognition method will sufficiently cut down the detection time and improve the accuracy and stability of the system. There is scope for more work, such as simplifying the processing of insect samples, obtaining further evidence to elucidate the characteristic wavelengths and genetic analysis of geographical strains.

## Acknowledgements

We thank the Academy of State Administration of Grain for financial support. We would like to thank Dr. Wu Yi for her comments on the manuscript of this paper. We would also like to thank Mrs. Liu for her technical support.

## References

- Al-Saqer, S.M., Weckler, P., Solie, J., Stone, M., Wayadande, A., 2011. Identification of pecan weevils through image processing. *American Journal of Agricultural and Biological Sciences* 6, 69-79.
- Baker, J.E., Dowell, F.E., Throne, J.E., 1999. Detection of parasitized rice weevils in wheat kernels with near-infrared spectroscopy. *Biological Control* 16, 88-90.
- Brown, S., Armstrong, K.F., Cruickshank, R.H., 2012. Molecular phylogenetics of a South Pacific sap beetle species complex (*Carpophilus* spp., Coleoptera: Nitidulidae). *Molecular Phylogenetics and Evolution* 64, 428-440.
- Cogdill, R.P., Hurburgh, C.R., Rippke, G.R., 2004. Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE* 47, 311-320.
- Demuth, H., Beale, M., Hagan, M., 1998. *Neural Network Toolbox User's Guide*, The MathWorks. Inc., Natick, MA.

- 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. *Journal of Economic Entomology* 91, 899-904.
- Dowell, F.E., Throne, J.E., Wang, D., Baker, J.E., 1999. Identifying stored-grain insects using near-infrared spectroscopy. *Journal of Economic Entomology* 92, 165-169.
- ElMasry, G., Iqbal, A., Sun, D., Allen, P., Ward, P., 2011. Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. *Journal of Food Engineering* 103, 333-344.
- Gan, F., Qu, H., 2006. Investigation on image's classified identification of pests in stored grain. *Microcomputer Information* 22, 248-250.
- Hidayat, P., Phillips, T.W., French-Constant, R.H., 1996. Molecular and morphological characters discriminate *Sitophilus oryzae* and *S. zeamais* (Coleoptera: Curculionidae) and confirm reproductive isolation. *Annals of the Entomological Society of America* 89, 645-652.
- Jagadeesan, R., Collins, P., Daglish, G., Ebert, P., Schlipalius, D., 2012. Phosphine resistance in the rust red flour beetle, *Tribolium castaneum* (Coleoptera: Tenebrionidae): Inheritance, Gene Interactions and Fitness Costs. *PLOS ONE* 7: e315822.
- Kuo, C.F., Hsu, C., Fang, C.H., Chao, S.M., Lin, Y.D., 2013. Automatic defect inspection system of colour filters using Taguchi-based neural network. *International Journal of Production Research* 51, 1464-1476.
- Larios, N., Soran, B., Shapiro, L.G., Martinez-Munoz, G., Lin, J., 2010. Haar random forest features and SVM spatial matching kernel for stonefly species identification; Istanbul. *IEEE*. pp. 2624-2627
- Liu, M., Wang, M.J., Wang, J., Li, D., 2013. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical* 177, 970-980.
- Maftoonazad, N., Karimi, Y., Ramaswamy, H.S., Prasher, S.O., 2011. Artificial neural network modeling of hyperspectral radiometric data for quality changes associated with avocados during storage. *Journal of Food Processing and Preservation* 35, 432-446.
- Maghirang, E.B., Dowell, F.E., Baker, J.E., Throne, J.E., 2003. Automated detection of single wheat kernels containing live or dead insects using near-infrared reflectance spectroscopy. *Transactions of the ASAE* 46, 1277-1282.
- Migdall, S., Bach, H., Bobert, J., Wehrhan, M., Mauser, W., 2009. Inversion of a canopy reflectance model using hyperspectral imagery for monitoring wheat growth and estimating yield. *Precision Agriculture* 10, 508-524.
- Nagalakshmi, S., Kamaraj, N., 2013. On-line evaluation of loadability limit for pool model with TCSC using back propagation neural network. *International Journal of Electrical Power & Energy Systems* 47, 52-60
- Neethirajan, S., Karunakaran, C., Jayas, D.S., White, N., 2007. Detection techniques for stored-product insects in grain. *Food Control* 18, 157-162.
- Peng, W., Lin, H.C., Chen, C.N., Wang, C.H., 2003. DNA identification of two laboratory colonies of the weevils, *Sitophilus oryzae* (L.) and *S. zeamais* Motschulsky (Coleoptera: Curculionidae) in Taiwan. *Journal of Stored Products Research* 39, 225-235.

- Perez-Mendoza, J., Throne, J.E., Dowell, F.E., Baker, J.E., 2003. Detection of insect fragments in wheat flour by near-infrared spectroscopy. *Journal of Stored Products Research* 39, 305-312.
- Ridgway, C., Chambers, J., 1996. Detection of external and internal insect infestation in wheat by near infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture* 71, 251-264.
- Ridgway, C., Chambers, J., Cowe, I.A., 1999. Detection of grain weevils inside single wheat kernels by a very near infrared two-wavelength model. *Journal of Near Infrared Spectroscopy* 7, 213-221.
- Roggo, Y., Edmond, A., Chalus, P., Ulmschneider, M., 2005. Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms. *Analytica Chimica Acta* 535, 79-87.
- Singh, C.B., Jayas, D.S., Paliwal, J., White, N.D.G., 2009. Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging. *Journal of Stored Products Research* 45, 151-158.
- Singh, C.B., Jayas, D.S., Paliwal, J., White, N.D.G., 2010. Identification of insect-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. *Computers and Electronics in Agriculture* 73, 118-125.
- Wen, C., Guyer, D.E., Li, W., 2009. Local feature-based identification and classification for orchard insects. *Biosystems Engineering* 104, 299-307.
- Yao, M.C., Chang, S.C., Lee, C.Y., Lu, K.H., 2012. A SCAR-based method for rapid identification of four major lepidopterous stored-product pests. *Journal of Economic Entomology* 105, 1100-1106.
- Yuan, F., 2002. *Agricultural entomology*. Beijing: China Agriculture Press. 350 p.
- Zayas, I.Y., Flinn, P.W., 1998. Detection of insects in bulk wheat samples with machine vision. *Transactions of the ASAE* 41, 883-888.
- Zhao, J., Chen, Q., Cai, J., Ouyang, Q., 2009. Automated tea quality classification by hyperspectral imaging. *Appl Opt* 48, 3557-3564.

**Table 1** Contents of each insect sample.

Insect species	Geographical strains	Number of insects
Rice weevil (SO)	Tongzhou (SO-TZ)	50
	Guangdong (SO-GD)	50
	Sichuan (SO-SC)	50
	Australia (SO-Aus)	50
Maize weevil (SZ)	Miyun (SZ-MY)	50
	Guangzhou (SZ-GZ)	50
	Wuhan (SZ-WH)	50
	Zibo (SZ-ZB)	50

**Table 2** Results of independent samples t-test for 28 features.

Wave length	Feature <sup>a</sup>	Insect species <sup>b</sup>	Mean	SD <sup>c</sup>	Significance
955 nm	w	SO	7255.600	947.199	0.000
		SZ	9068.400	994.560	
	Ec	SO	0.897	0.010	0.000
		SZ	0.891	0.010	
	Ma	SO	145.350	10.750	0.000
		SZ	160.530	10.558	
	Mi	SO	64.149	4.242	0.000
		SZ	72.616	3.947	
	m	SO	122.710	18.189	0.000
		SZ	109.370	18.456	
	$\sigma$	SO	40.710	8.560	0.274
		SZ	39.792	8.197	
	R	SO	0.026	0.011	0.260
		SZ	0.025	0.009	
	$\mu_3$	SO	0.207	0.365	0.068
		SZ	0.279	0.416	
	U	SO	0.009	0.005	0.051
		SZ	0.008	0.002	
	e	SO	7.192	0.250	0.966
		SZ	7.191	0.297	
659 nm	m	SO	0.996	0.512	0.000
		SZ	0.405	0.483	
	$\sigma$	SO	0.412	0.099	0.000
		SZ	0.554	0.108	
	R	SO	3.279	1.572	0.000
		SZ	4.904	1.860	
	$\mu_3$	SO	3.173	3.140	0.000
		SZ	5.204	2.799	
	U	SO	0.688	0.145	0.874
		SZ	0.686	0.101	
	e	SO	0.723	0.257	0.000
		SZ	0.853	0.218	

**Table 2** (Con.).

Wave length	Feature <sup>a</sup>	Insect species <sup>b</sup>	Mean	SD <sup>c</sup>	Significance	
505 nm	m	SO	155.600	12.921	0.000	
		SZ	148.440	12.993		
	$\sigma$	SO	38.569	6.347	0.001	
		SZ	36.548	5.136		
	R	SO	0.023	0.007	0.000	
		SZ	0.021	0.005		
	$\mu_3$	SO	-0.315	0.209	0.004	
		SZ	-0.255	0.210		
	U	SO	0.009	0.003	0.208	
		SZ	0.008	0.001		
	e	SO	7.100	0.182	0.152	
		SZ	7.074	0.184		
	Ratio band	m	SO	74.662	9.254	0.000
			SZ	84.973	18.881	
		$\sigma$	SO	25.433	4.684	0.000
			SZ	32.362	8.502	
R		SO	0.010	0.004	0.000	
		SZ	0.017	0.008		
$\mu_3$		SO	0.220	0.728	0.000	
		SZ	0.324	0.312		
U		SO	0.012	0.002	0.000	
		SZ	0.010	0.003		
e		SO	6.594	0.238	0.000	
		SZ	6.902	0.385		

<sup>a</sup>Four shape features and 24 textural features<sup>b</sup>See Table 1 for species codes.<sup>c</sup>SD is standard deviation**Table 3** The recognition of insect species samples in the prediction set with the optimal BPNN model.

Actual species <sup>a</sup>	Recognized species		Recognition rate%
	SO	SZ	
SO	79	1	98.75
SZ	2	78	97.50

<sup>a</sup>See Table 1 for species codes.

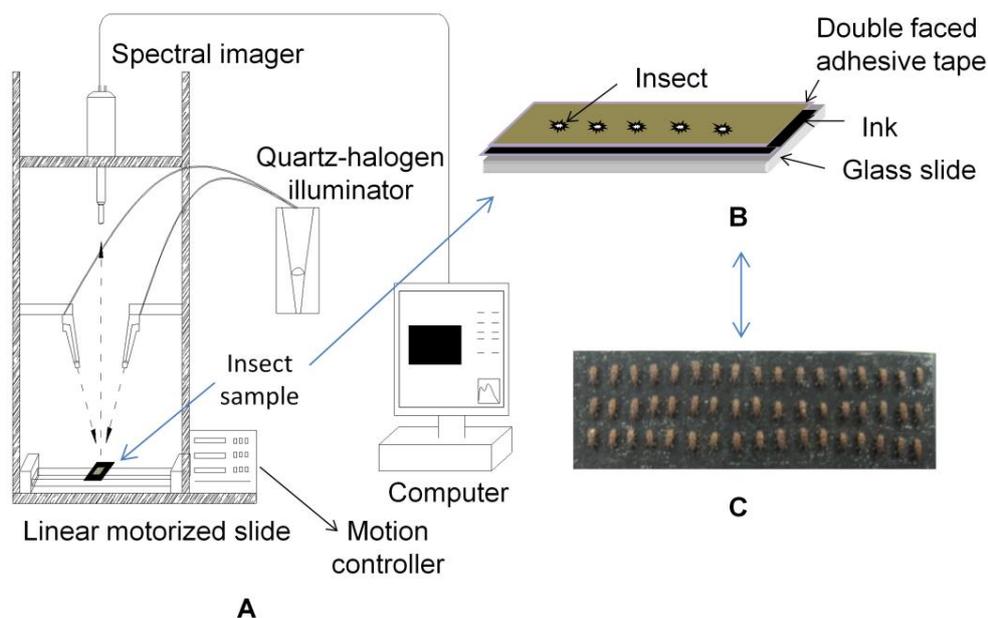
**Table 4** The recognition of geographical strains of rice weevil samples with the optimal BPNN model.

Actual strain <sup>a</sup>		Recognized strain				Recognition rate%
		SO-TZ	SO-GD	SO-SC	SO-Aus	
Calibration set	SO-TZ	19	3	0	0	86.36
	SO-GD	1	29	2	0	90.63
	SO-SC	0	5	27	0	84.38
	SO-Aus	0	0	0	34	100.00
Prediction set	SO-TZ	22	6	0	0	78.57
	SO-GD	2	15	1	0	83.33
	SO-SC	0	4	14	0	77.78
	SO-Aus	0	0	0	16	100.00

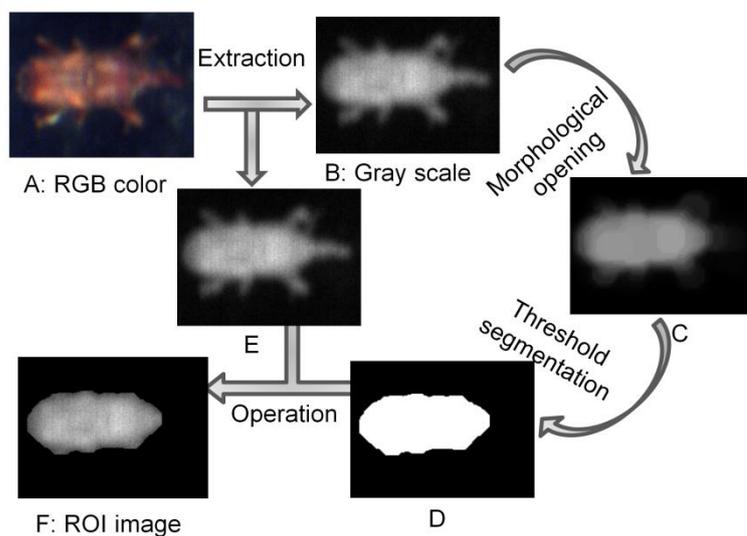
<sup>a</sup>See Table 1 for strains codes.**Table 5** The recognition of geographical strains of maize weevil samples with the optimal BPNN model.

Actual strain <sup>a</sup>		Recognized strain				Recognition rate%
		SZ-MY	SZ-GZ	SZ-WH	SZ-ZB	
Calibration set	SZ-MY	22	0	0	0	100.00
	SZ-GZ	0	31	1	0	96.88
	SZ-WH	0	0	32	0	100.00
	SZ-SD	0	0	2	32	94.12
Prediction set	SZ-MY	25	3	0	0	89.29
	SZ-GZ	0	16	2	0	88.89
	SZ-WH	0	0	17	1	94.44
	SZ-ZB	0	0	2	14	87.50

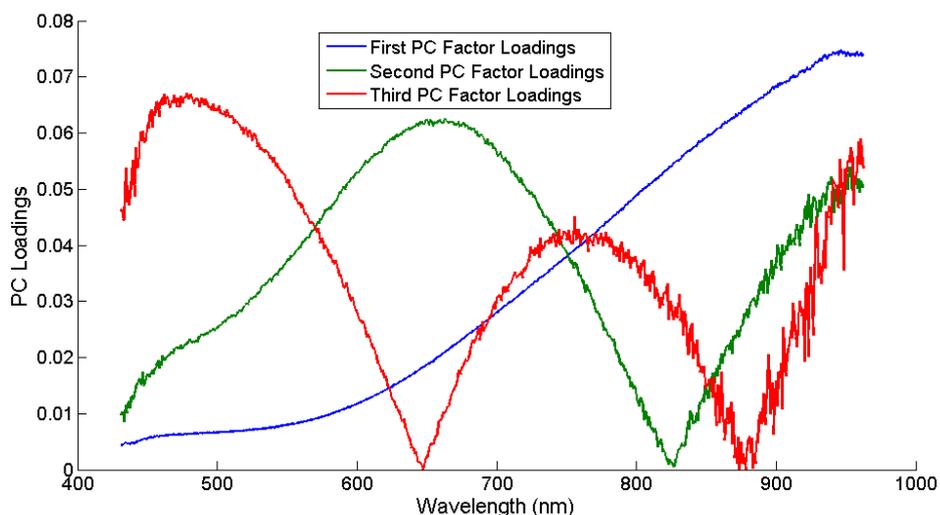
<sup>a</sup>See Table 1 for strain codes.



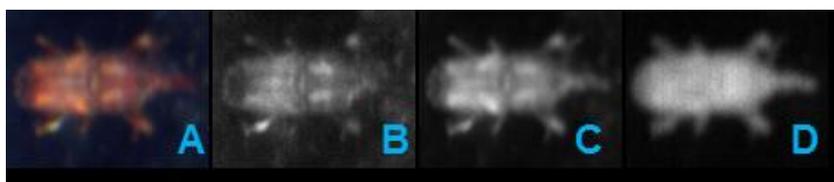
**Figure 1** Schematic of the hyperspectral imaging system with detail of the specimen stand (the hyperspectral imaging system (A), sketch map of the stand (B) and a photograph of insect specimens on the stand (C)).



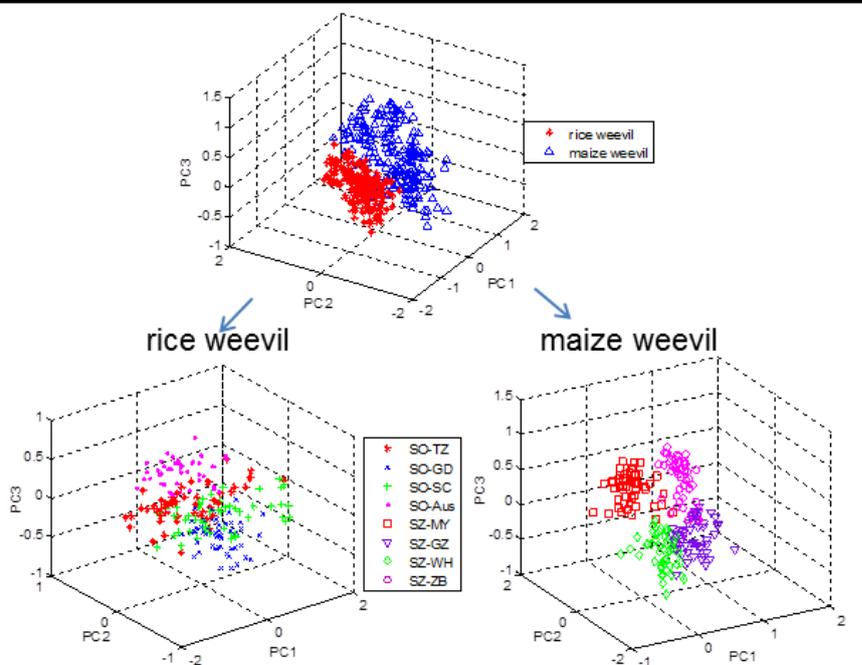
**Figure 2** Identified maize weevil through image processing and ROI extraction techniques based on template matching.



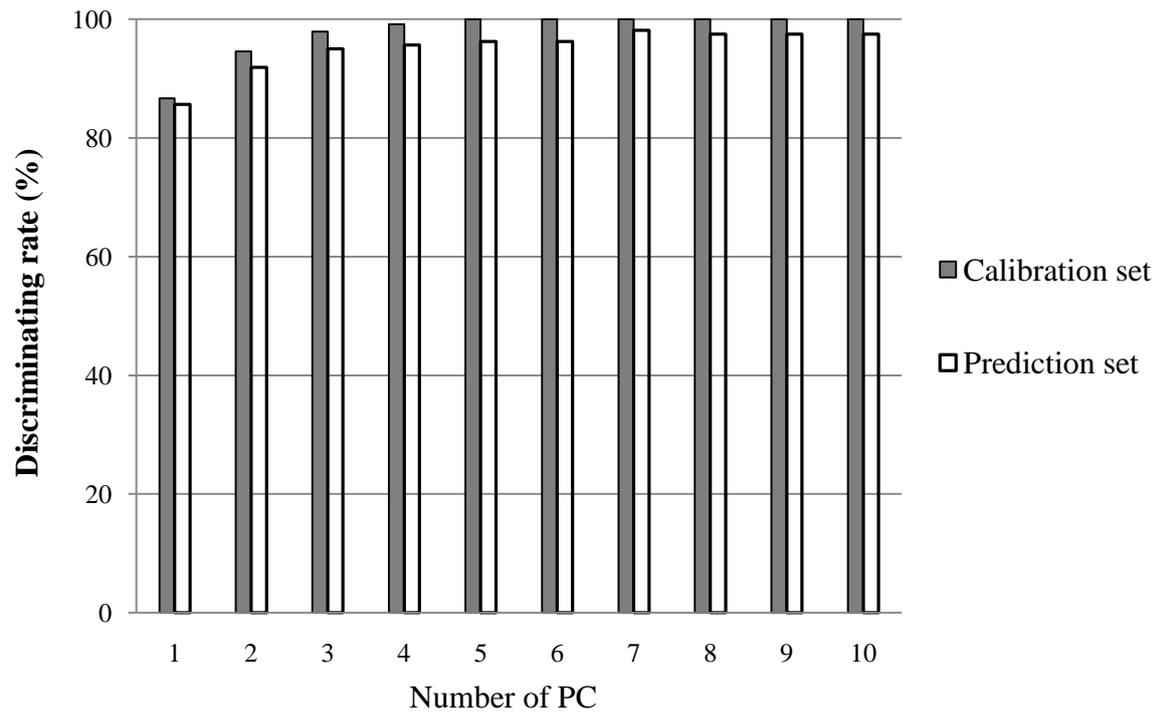
**Figure 3** The first three PC mean factor loadings (absolute values) for insect samples.



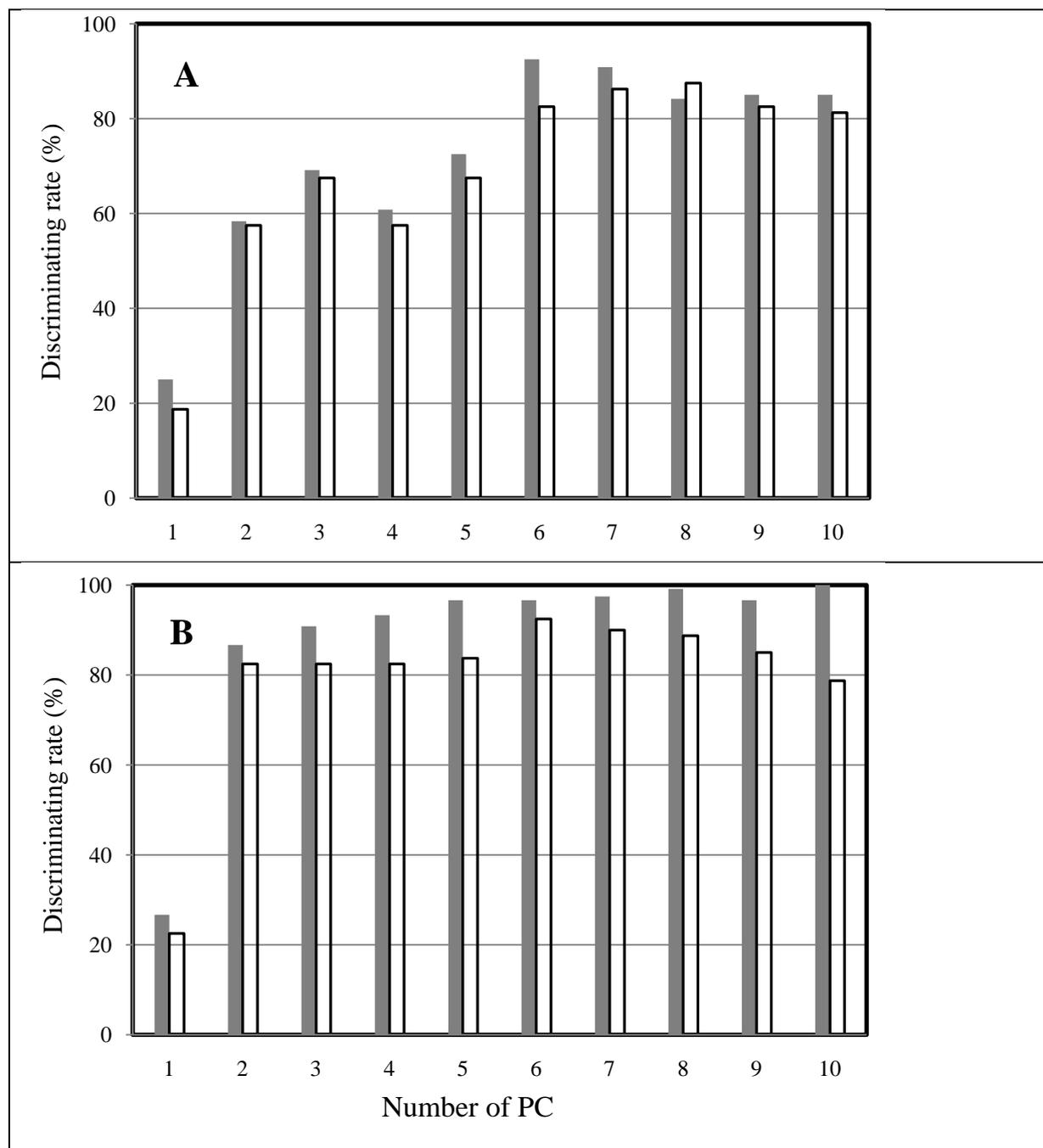
**Figure 4** Pictures of insect samples in RGB color and three characteristic wavelengths (RGB color (A), 505.21 nm (B), 658.56 nm (C), 955.42 nm (D)).



**Figure 5** Score plot of the first three PCs of insect samples.



**Figure 6** Insect species recognition result of the BPNN model with a different number of PCs.



**Figure 7** Insect geographical strains recognition result of the BPNN model with a different number of PCs for rice weevils (A) and maize weevils (B).