Bayesian discriminant analysis of plant leaf hyperspectral reflectance for identification of weeds from cabbages

Wei Deng1,3,4,5*, Yanbo Huang2, Chunjiang Zhao1,3,4,5, Liping Chen1,3,4,5 and Xiu Wang1,3,4,5

1Beijing Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China.
2United States Department of Agriculture, Agricultural Research Service, Crop Production Systems Research Unit, Stoneville, Mississippi, USA.
3National Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China.
4Key Laboratory of Agri-informatics, Ministry of Agriculture, Beijing 100097, China.
5Beijing Key Laboratory of Intelligent Equipment Technology for Agriculture, Beijing 100097, China.

Received 16 September, 2015; Accepted 19 January, 2016

In order to spray herbicides accurately on targets, this study focused on spectral classification of weeds and crops for potential to rapidly detect weeds in crop fields. A 350 ~ 2500 nm FieldSpec-FR spectroradiometer was used to measure spectral responses of the canopies of the seedling vegetables, cabbage ‘8398’ and cabbage ‘Zhonggan 11’, and weeds, Barnyard grass, green foxtail, goosegrass, crabgrass, and Chenopodium quinoa, at five- and seven-week growth stages (WGS). First, the characteristic wavelengths (CW) were determined using Principal Component Analysis (PCA). Then, the plants were classified using Bayesian discriminant analysis with the reflectance of the CWs. The results of spectral analysis indicated that the different growth stages of cabbages had little influence on the spectral identification of cabbages and weeds. The eight CWs determined were used as the input to the model for Bayesian discriminant analysis to classify two varieties of cabbages and five weeds with the correct classification rate of 84.3% for model testing. When the two varieties of cabbages were considered as the same category, the correct classification rate was improved to 100%. It was concluded that Bayesian discriminant analysis could be used to identify weeds from seedling cabbages using leaf hyperspectral reflectance.

Key words: Weed identification, spectrum analysis, visible and near-infrared, Bayesian discriminant, seedling weed, seedling cabbage.

INTRODUCTION

According to the research report from the United Nations Food and Agriculture Organization (FAO) in August 2009,
weeds should be regarded as farmers’ No. 1 natural enemy. It was reported that according to a leading environmental research organization, Land Care of New Zealand, weeds cause about $95 billion every year in the lost food production at global level, compared with $85 billion for pathogens, $46 billion for insects, and $2.4 billion for vertebrates (excluding humans). Of the $95 billion, $70 billion are estimated to be lost in developing countries (FAO, 2009). In China, the crop yield losses annually caused by weeds sum up to about 10% of the gross grain output (Tang, 2010). Facing the severity of the crop losses caused by weeds, it is urgent to seek highly efficient methods for effective weed control. The chemical weeding method commonly adopted at present has provoked a lot of problems, such as excessive pesticide residues, growing number of pesticide-resistant weeds, destruction of ecological environment, and lower quality and safety of agricultural products (Thompson et al., 1991). Therefore, it is critical to have a method which could not only control the growth of weeds, but also decrease the use of herbicides, and hence prevent from excessive herbicide application. In order to minimize crop damage and environmental pollution, herbicides should be sprayed accurately on targets with appropriate dose.

Rapid access to the information of spraying targets is a critical process in precision chemical application. Many approaches to weed detection and identification have been reported in literatures, such as photoelectric detection technique (Biller, 1998; Andújar et al., 2011), ultrasonic detection technique (Andújar et al., 2012), remote sensing detection technique (Thorp and Tian, 2004), and image processing detection technique (Weis and Sökefeld, 2010; Christensen et al., 2009; Burgos-Artizzu et al., 2009; Piron et al., 2011), X-ray weed detection technique (Haff and Slaughter, 2009), and spectral weed detection technique (Vrincts et al., 2002; Sui et al., 2008). The image processing technique can detect the target's profile and determine the coverage and amount of smart spray, but mostly this technique is still used in the laboratory instead of being used in field because of its poor stability, large amount of data processing, relatively slow response, and the high costs. In comparison, spectral detection has been widely used in real-time detection system because of its fast response, availability for non-contact detection, strong anti-interference, high reliability, low cost, simple and small configuration, and low power consumption. Many studies have focused on spectral classification of weeds and crops for potential to rapidly detect weeds in crop fields. Spectral sensors has been designed and widely used in real-time detection system in crop production, because they offer fast response, availability for non-contact detection, strong anti-interference, high reliability, low cost, simple and small configuration, and low power consumption (Wang et al., 2001; Rogalski, 2003). These sensors could be used to further study the spectral characteristics for classification of weeds and crops to realize fast and timely weed management in crop fields.

Koger et al. (2003) analyzed the hyperspectral reflectance of soybean, Ipomoea lacunosa, and soil at the two- and four-leaf stages of weed growth using wavelet analysis. For comparison, the raw spectral bands and principal components were used as discriminating features. For the two-leaf to four-leaf weed growth stage, the two features resulted in classification accuracies of 83 and 81%, respectively. Jurado-Exposito et al. (2003) distinguished sunflowers, wheat, and seven other seedling broadleaf weeds using near-infrared spectroscopy. It was found that the near infrared spectroscopy within 750 ~ 950 nm was able to identify these plants. Slaughter et al. (2004) distinguished Solanum weeds and tomato using the spectral reflectance in visible and near-infrared wavebands, narrow-band hyperspectral modeling, and discriminant analysis. It was found that the spectral absorbance data of weeds and tomatoes at the wavelength range of 2120 to 2320 nm offered the best classification accuracy (100%), and narrow-band hyperspectral models of the data in the visible range also achieved good classification results (95%), while the broad-band models based on color information only provided 75% correct classification rate.

Thenkabail et al. (2004) studied how to select the optimum wavebands for classifying plants (shrubs and weeds) and crops (corn) in the range of 400 to 2500 nm waveband. It was found that 90% correct classification rate could be obtained by modeling using 13 to 22 wavebands selected from the original 168 wavebands using PCA and stepwise discriminant analysis, which accuracy was increased by 9 to 43% compared with modeling using ETM (Landsat Enhanced Thematic Mapper), plus brandband data. Piron et al. (2008) classified seven different weeds in carrot fields under artificial lighting conditions using a visible and near-infrared multispectral sensor, and found that overall correct classification rate was 72% when three optimal wavebands, 450, 550, and 700 nm were selected using the exhaust algorithm and used to establish the weed identification models. Mao et al. (2005) measured the spectral reflectance of wheat, shepherd's purse, and small quinoa in the wavelength range of 700 to 1100 nm using a Fourier transform infrared (FTIR) spectrometer, and extracted 7 characteristics wavelengths, 686, 708, 722, 795, 929, 956, and 1122 nm using stepwise discriminant analysis and achieved 97% correct identification rate through establishing the model of identifying wheat and weeds. Vrincts et al. (2002) measured the canopy reflectance of maize, sugarbeet, and seven weed species at 400 to 2000 nm. The spectral characteristics were also analyzed. Six wavelengths (555, 675, 815, 1265, 1455, and 1665 nm) at characteristic points in the spectrum were selected to derive the RVI. The STEPDISC and DISCRIM procedures in SAS were applied in the discrimination of crops (maize and sugarbeet) from weeds. The
classification result showed that crop and weeds could be
classified at an accuracy of higher than 97%. More than
90% of sugar beet and weeds could be identified
correctly using a line spectrophotograph (480 to 820 nm) in
classifying the plants. With the application of the spectral
technique, Sui et al. (2008) developed the Weed Seeker
sensor module to detect the presence of weeds by
measuring the reflectance of weeds and bare ground. The
module serves as a useful tool for locating weeds.

Karimi et al. (2006) used SVMs and NNs for weed and
nitrogen deficiency detection in corn with non-imaging
hyperspectral reflectance data as input. In an extensive
approach, they uniquely identified and classified four
weed management practices and three different nitrogen
rates. The classification accuracy using SVMs was higher
than with NN in this research project 69.2 versus 58.3%,
respectively. Lopez-Granados et al. (2008) used the
hyperspectral reflectance spectrum from 400 to 900 nm
for the classification of late season grass weeds and
wheat plants in a field study. Their approach using linear
and nonparametric functional discriminant analysis and
NNs has shown that, in general, a preliminary
computation of most relevant PCs improves the
classification accuracy. They concluded from their study
that analysis in real time of high spectrally resolved
images will be adequate to map grass weed patches in
wheat. For practical implementation, Moshou et al. (2002)
developed a weed species spectral detector based on
neural networks with local linear mappings. A self-
organized map (SOM) neural network achieved fast
convergence and good generalization. The proposed
method classified crops and different species of weed
with high accuracy. Chen et al. (2009) measured the
spectral reflectance of leaves of rice, cotton, Barnyard
grass and Cephalanoplos indoors in the range of 350 to
2500 nm wavelength using a spectroradiometer and
determined the characteristic wavelengths (CWs) using
the stepwise discriminant method, and then classified
these plants using the Discrim processing, a function of
discriminant analysis in SAS statistical software (SAS
Institute Inc., Carrey, North Carolina, the United States).

This study found that monocotyledons, like rice and
Barnyard grass, could be accurately classified using the
five CWs, 375, 465, 585, 705, and 1035 nm, in which the
correct identification rate reached 100%; dicotyledons,
like cotton and Cephalanoplos, could also be accurately
classified using the three CWs, 383, 415, and 435 nm, in
which the correct identification rate also reached 100%.
The spectral reflectance of cabbages and weeds were
measured in the 350 to 2500 nm band and preprocessed
the data with different levels to improve the operation
efficiency. All kinds of plants were classified using the
Soft Independent Modeling of Class Analogy (SIMCA).
While the selected 23 feature wavelengths were set as
the input variables, the classification rate of the modeling
set and the predicting set were respectively 98.6 and
100% (Zu et al., 2013a, b).

The objects of most of the previous studies were specific
to crops like corn, wheat, rice, and cotton, but few to
vegetables. The vegetables, especially dicotyledonous
vegetables, are important economic crops in China,
which are widely cultivated throughout the country with
wide cultivated area and high inputs of labors. Therefore,
weed identification in vegetable fields has considerable
social and economic benefits and practical significance.
Additionally, few of the previous studies were on whether
the possible changes of spectral characteristics caused
by the changing metabolism at the different growth
stages of a crop would affect the consistency of spectral
identification of crops and weeds during different growth
stages.

For optimal results with minimal investment, weeds
should be handled at seedling stage in crop fields (Li et
al., 2007). Seedling weeds are sensitive to herbicides
because plants are small and the tissues are young.
Therefore, the seedling stage is the time for complete
weed control. In the further growth stage, weed plants
become large and their tissues are strong. Consequently,
with the thickening of the waxy coat, it will be difficult for
herbicide agents to penetrate the weed leaves and weed
herbicide tolerance will be increased accordingly. As the
result, the herbicide will be hard to take effect. This study
was conducted on seedling weeds in the field with
seedling cabbages.

In this study, two varieties of seedling cabbages, ‘No.
8398’ and ‘Zhonggan No. 11’, and five breeds of weeds,
Barnyard grass, green foxtail, goosegrass, crabgrass,
and Chenopodium quinoa, which are commonly-seen.
Annual gramineous plants in cabbage fields with strong
adaptability, wide coverage, fast multiplying, and
inestimable harm to crops, were selected as the
representatives of plants. The spectral reflectance of the
plant canopies were measured in the wavelength range
of 350 to 2500 nm at two seedling growth stages of the
35th (five-week growth stage (WGS)) and 50th days
(seven WGS), respectively. The objectives of this study
are (1) to determine CWs at which the spectral
reflectances were sensitive to plant identification using
the spectral data measured, respectively at five WGS and
second WGS; (2) to establish the Bayesian discriminant
model to classify two varieties of cabbages and five
different weeds.

MATERIALS AND METHODS
Experimental
Two varieties of cabbages used in the study were cabbage ‘No.
8398’ and cabbage ‘Zhonggan No. 11’, whose seeds were provided
by the Institute of Vegetables, the Chinese Academy of Agricultural
Sciences (Beijing, China). Five varieties of weeds were Barnyard
grass, green foxtail, goosegrass, crabgrass, C. quinoa, whose
seeds were provided by the College of Agronomy and
Biotechnology, China Agricultural University (Beijing, China). The
two varieties of cabbages and five kinds of weeds were planted in
pots in a greenhouse of the Chinese National Engineering
Research Center for Information Technology in Agriculture (Beijing, China) on 23 March, 2012. Each variety of plants was grown in 30 pots; therefore, the total number of the plant samples was 210 for the seven varieties of plants.

Data acquisition

The instrument for measuring spectral data was the ASD full range FieldSpec Pro. spectroradiometer (ASD, Inc., Boulder CO., USA). The measuring range of the spectroradiometer is 350 to 2500 nm. The spectral resolution is 1.4 nm in the range of 350 to 1000 nm and 2 nm in the range of 1000 to 2500 nm. The field of view (FOV) of the measuring probe is 25°. The spectral data of the 210 pots of plant canopies were collected in the test field of National Engineering Research Center for Information Technology in Agricultural (Beijing, China) during 10:30 am to 14:30 pm on 28 April and 13 May, 2012, respectively, corresponding to two growing stages of the plants, five WGS and seven WGS.

The white reference board was measured for spectral calibration each 10 to 15 min depending on weather conditions. After each white reference measurement, the fiber-optic probe of the FieldSpec spectroradiometer was placed vertically above the plant canopy and measured the data. In order not to affect the reflectivity of the plants, the operator should dress dark. The spectroradiometer was set in the condition that an output datum was obtained from the average of ten measurements. The measured spectral data were converted to reflectance and then to ASCII text format using the function in the ASD ViewSpectro Pro. Software provided by ASD Inc. After being imported to Microsoft EXCEL spreadsheet, the text files were transformed to matrices which were then transferred into the Unscrambler (CAMO software AS, Oslo, Norway) and SAS software for further data processing.

For data collection on 28 April, 2012, each pot of plants was measured for three times so that the total number of the obtained spectral data was 90 for each variety of plants (30 pots for each plant) and 630 for all the 7 varieties of plants. For the measurement on 13 May, 2012, each pot was measured for five times so that the total number of the collected spectral data was 150 for each variety of plant and 1050 for all the 7 varieties of plants.

In order to reduce the random errors which are always accompanied with the spectral signal in the process of data acquisition, the spectral data were averaged for each pot of plants, which resulted in 30 averaged spectral data for each plant respectively for the measurements on 28 April and 13 May.

Principal component clustering analysis

The clustering analysis of the spectral data of cabbages and weeds was conducted using the Principal Component Analysis (PCA) method after data preprocessing. For each of the plants, 20 sample data were randomly selected as the training sample set, the other 10 data as the testing sample set. In the Unscrambler software system, the full cross validation methods in PCA and Partial Least Square (PLS) were separately used to extract the principal components to build the plant classification models. The analysis process was started by extracting 20 principal components from the spectral reflectance data. Then, the outliers were repeatedly excluded by considering the spatial aggregation conditions and spatial position of all the sample points in scoring graphs of the results based on the principle of maximization of distance between the classes and minimization of distance within a class. The appropriate principal components were determined according to the cumulative credibility of each Principal Component (PC) and the classification model was re-built with clustering all the plants (Li, 2010).

Determination of characteristic wavelengths

In order to find out the CWs needed for identification of cabbages and weeds, the score of each PC, accumulative confidence level, and loading diagrams resulting from the former PCA and the relationship between PC and original wavebands expressed through loading graph should be analyzed. According to the loading graph of wavelength variable responding to the optimum PCs obtained from the former analysis, the wavelengths greatly (positive and negative) correlating with PCs were selected as the characteristic wavelengths sensitive to the identification of various plants and with higher correlation for establishing the identification models. The loading coefficients of the selected wavelengths were used to reflect the importance of the wavelengths to the PCs.

Bayesian classification model

Bayes’ theorem (Bayes 1764)

Mathematically, Bayes’ theorem gives the relationship between the probabilities of events $A$ and $B$, $P(A)$ and $P(B)$, and the conditional probabilities of $A$ given $B$ and $B$ given $A$, $P(A|B)$ and $P(B|A)$. The form of Bayesian inference is mostly expressed as follows, provided that $P(B) \neq 0$.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Probability measures a degree of belief. Bayes’ theorem then links the degree of belief in a proposition before and after accounting for evidence. For example, suppose somebody proposes that a biased coin is twice as likely to land heads than tails. Degree of belief in this might initially be 50%. The coin is then flipped a number of times to collect evidence. Belief may rise to 70% if the evidence supports the proposition.

For proposition $A$ and evidence $B$, $P(A)$, the prior, is the initial degree of belief in $A$. $P(A|B)$, the posterior, is the degree of belief having accounted for $B$. The quotient $P(B|A)/P(B)$ represents the support $B$ provides for $A$.

The event $B$ is fixed in the discussion, and we wish to consider the impact of its having been observed on our belief in various possible events $A$. In such a situation, the denominator of the last expression, the probability of the given evidence $B$ is fixed; what we want to vary is $A$. Bayes theorem then shows that the posterior probabilities are proportional to the numerator:

$$P(A|B) \propto P(A) \cdot P(B|A)$$

(proportionality over $A$ for given $B$).

Further, if events $A_1$, $A_2$, ..., are mutually exclusive and exhaustive, that is, one of them is certain to occur but no two can occur together, and we know their probabilities up to proportionality, then we can determine the proportionality constant by using the fact that their probabilities must add up to one. For instance, suppose some evidence increases the probability of some event $A$ by a factor of two. Bayes theorem then shows that the posterior probabilities are proportional to the numerator:

$$P(A|B) \propto P(A) \cdot P(B|A)$$

Adding these two formulas, we deduce that,

$$c = \frac{1}{P(A) \cdot P(B|A) + P(-A) \cdot P(B|-A)}$$

For that the extended form, often, for some partition $\{A_i\}$ of the event space, the event space is given or conceptualized in
Figure 1. Curves of average spectral reflectance of seven plants. (a) Curves of average spectral reflectance at five WGS. (b) Curves of average spectral reflectance at seven WGS.

terms of $P(A_j)$ and $P(B|A_j)$. It is then useful to compute $P(B)$ using the law of total probability:

$$P(B) = \sum_j P(B|A_j)P(A_j)$$

$$\Rightarrow P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_j P(B|A_j)P(A_j)}$$

Bayesian modeling

Using the eight CWs determined from the data at five WGS as the input variables, the discrimination model was built based on the Bayesian criterion and used to discriminate the cabbages and the weeds. In the process, 7 different plants were separately labeled using categorical variables as Y-8398 (cabbage 8398), ZG (cabbage Zhonggan 11), BC (Barnyard grass), GW (Setaria viridis), MT (crabgrass), NJ (Eleusine indica), and XL (C. quinoa). For each plant, two-thirds of the samples were randomly selected as the training group (140 samples) so that all the samples for the 7 plants were divided into two groups for training and testing, respectively. Then, using the data of categorical variables and 8 CWs, the discrimination model was built. In order to verify the reliability and robustness of the model, the other one-third of the samples (70 samples) were used as the testing group and the input of the model to classify the cabbage and weed samples.

RESULTS AND DISCUSSION

Data observation

Further, the 30 averaged data were averaged for each variety of the plants. Figure 1 shows the averaged spectral reflectance curves for each of the plants. It can be seen that the curves are all the same as the typical healthy plant spectral reflectance curves. In the vicinity of
450 and 650 nm, the light of blue and red bands is absorbed by chlorophylls for photosynthesis, leading to two distinct absorption valleys. At 550 nm, the light of green light is partly absorbed by chlorophylls and partly reflected, forming a reflection peak. A steep slope at 700 to 800 nm demonstrates a sharp increase of reflectivity to form a high reflection platform. In the range of 800 to 1300 nm, the porous parenchyma tissue (spongy tissue) of plant leaves have always a very strong reflection to near-infrared light, creating a peak area of reflectance on spectral curves with the reflectance up to 40%. At about 1450 and 1950 nm, an apparent absorption valley is formed due to the cell sap, cell membranes, and absorbed vapor of the plant leaves.

Particularly, it is shown in Figure 3 that for the spectral reflectance curves at the first WGS, the spectral reflectivity of green foxtail in the range of 700 to 1800 nm is obviously higher than other plants, while the spectral reflectivity of cabbages is modest and the spectral curves of two cabbages are almost superimpose. For the spectral curves at the seven WGS, the spectral reflectance curve of goosegrass is obviously distinct from other plants. It is seen that the curves of cabbages are relatively stable while the curves of weeds fluctuate dramatically, which could be a characteristic used to differentiate cabbages from weeds. In overall, there are some differences between the spectral curves of cabbages and weeds, but the spatial distributions of some samples overlap, which make it difficult to exactly distinguish the variety of each sample. In order to accurately classify cabbages and weeds, quantitative discriminant models should be established.

**Clustering based on PCA**

In the case of optimum preprocessing, the score plot of PC 1 and PC 2 of the training set is shown in Figure 2, in which the horizontal axis presents the score value of the
Figure 3. Loading graph corresponding to the PC1 in modeling the data at five WGS.

first PC and the vertical axis is the score value of the second PC. It can be seen from Figure 2a that the data samples of cabbages mainly concentrate in the second and third quadrant, goosegrass samples mainly in the upper area of the second quadrant, crabgrass and green foxtail samples in upper area of the first quadrant, Barnyard grass and C. quinoa principally around the horizontal axis in the fourth quadrant. It can be found from Figure 2b that cabbage samples closely distributing in the first quadrant show a good degree of aggregation which indicates that the two varieties of cabbages can be regarded as the same category. As well in a good degree of aggregation, all the Barnyard grass samples closely gather in the fourth quadrant and all the green foxtail in third quadrant. Although crabgrass samples distribute in both the second and third quadrant, the aggregating degree is still high. The samples of C.quinoa and goosegrass loosely gather in the second quadrant. Therefore, it illustrates that PC1 and PC2 have better contribution to clustering cabbages and weeds. The synthetic method of PCA and clustering analysis can not only to a large extent reduce the data dimension but also greatly express the features of original data without losing the effective information.

Determining CWs

The loading diagrams corresponding to the optimum PCs obtained from spectral data processing at five and seven WGS are respectively shown in Figures 3 and 4. In the diagram, only the first PC is shown. In the loading diagrams, the horizontal coordinate represents the wavelength and the vertical coordinate is the load factor (that is, the correlation between wavelength and plant species) of each wavelength, wherein, the larger the absolute value of the corresponding load factor of a wavelength variable is, the stronger the correlation between the PC and the corresponding load factor is, and the more sensitive to the discrimination of the plant species.

It could be found from the loading diagrams that obvious crests and troughs present at some wavelengths and the rates of change of corresponding load factors appear as local maximum/minimum. These wavelengths are likely to play a decisive role in the identification of cabbages and weeds (Piron et al., 2011). Whereby, the corresponding CWs selected from the loading diagrams were 552, 567, 602, 607, 667, 715, 725, 1345, 1402, 1447, 1725, 1925, 1945, 1955, 2015, and 2072 for the five WGS plants and 425, 567, 667, 685, 745, 755, 1095, 1135, 1155, 1235, 1315, 1345, 1385, 1402, 1435, 1535, 1545, 1625, 1725, 1805, 1815, 1925, and 2030 for the seven WGS plants. The number of the CWs selected from the spectral data in the five and seven WGS was respectively 16 and 23.

Although the dimension of the data was already greatly reduced relative to the original data, the number of the band data is still relatively large for a practical instrumental design and development for agricultural use. Therefore, the selected CWs needed to be further optimized. The optimization process was to start with the
first PC by sorting of wavelengths in terms of the absolute value of the corresponding load factors. Then, the wavelengths were further selected at which the absolute value of load was large and obvious crests and troughs were present in the loading diagrams. As the result, the selected CWs were 567, 667, 715, 1345, 1402, 1725, 1925, and 2015 nm for five WGS and 567, 667, 745, 1345, 1402, 1545, 1725, and 1925 nm at seven WGS (Table 1). In order to evaluate the effect of the selected CWs, the identification models were built using these CWs.

Among the each 8-CW set further determined respectively at two growth stages, just two of them were different as highlighted in bold in Table 1, which indicated that the change of the growth stage of cabbages and weeds had limited influence on the spectral features.

**Bayesian classification**

Based on the method mentioned in earlier, the discriminant functions of the classification models were obtained as follows in Equation 1, in which, the input variables ($x_1$, $x_2$, ..., $x_8$) are the eight CWs extracted from the data at the five WGS through PCA method.

$$
BC = -49.89 + 75.11x_1 + 976.62x_2 - 247.88x_3 + 8.52x_4 + 148.77x_5 - 358.52x_6 + 87.37x_7 + 218.22x_8
$$

$$
GW = -68.91 + 1116x_1 + 745.44x_2 - 1112x_3 + 677.47x_4 - 436.09x_5 + 222.38x_6 + 44.97x_7 - 216.67x_8
$$

$$
MT = -45.98 + 213.48x_1 + 668.17x_2 - 309.73x_3 + 300.99x_4 + 149.24x_5 - 24.16x_6 + 114.15x_7 - 285.32x_8
$$

$$
NJ = -41.12 + 187.37x_1 + 725.07x_2 - 422.14x_3 + 663.92x_4 - 98.88x_5 - 338.86x_6 + 127.71x_7 - 76.97x_8
$$

(1)

$$
XL = -93.59 - 3026x_1 + 1839x_2 + 1459x_3 - 329.18x_4 + 78.99x_5 + 600.50x_6 - 3.27x_7 - 331.96x_8
$$

$$
Y-8398 = -83.34 - 1803x_1 + 1380x_2 + 1344x_3 + 39.46x_4 - 148.77x_5 - 358.52x_6 + 87.37x_7 + 218.22x_8
$$

$$
ZG = -88.29 - 1874x_1 + 1398x_2 + 1432x_3 - 118.03x_4 - 87.21x_5 - 211.36x_6 + 54.09x_7 + 172.79x_8
$$

The frequency numbers of each training sample and the misclassified rates of the plants for 7 different plants discriminated into various categories are exhibited in Table 1.

In training data classification, five Cabbage ‘8398’ samples were misclassified as Cabbage ‘Zhonggan No.11’ and four ‘Zhonggan No.11’ as ‘8398’, which was because they are all cabbages with the identical internal structure and the similar pigment and appearance. Therefore, it is apparent that different varieties of cabbages can be considered in the same category. In addition, one Barnyard grass was misclassified as crabgrass, one crabgrass was misclassified as goosegrass, and two crabgrasses were misclassified as crabgrass, which is perhaps because they are all monocotyledonous weeds with similar internal structure and composition. Total of 13 samples in the training set were falsely classified and the misclassified rate is 0.0929, which is translated to 90.7% of the overall correct classification rate.

The classification results of the testing set showed that three Cabbage ‘8398’ were misclassified as Cabbage ‘Zhonggan No.11’ and six Cabbage ‘Zhonggan No.11’ were misclassified as ‘8398’. Moreover, one crabgrass was misclassified as goosegrass and one crabgrass was misclassified as Barnyard grass. Total of 11 samples were misclassified, therefore the misclassified rate is 0.1571. Then, the correct classification rate is 84.3%.

In order to verify the similarity of the spectral characteristics of different varieties of cabbages, cabbage ‘8398’ and ‘Zhonggan No.11’ were combined into one category. The 8 CWs which had been used earlier were still used as the input variables. All varieties of plants
Table 1. The CWs selected from the loading diagrams of PC at the first and second selection.

<table>
<thead>
<tr>
<th>Selection</th>
<th>Testing stage</th>
<th>Characteristic wavelengths (CWs) / (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First selection</td>
<td>Five WGS</td>
<td>552, 567, 602, 607, 667, 715, 725, 1345, 1402, 1447, 1725, 1925, 1945, 1955, 2015, and 2072</td>
</tr>
<tr>
<td></td>
<td>Seven WGS</td>
<td>425, 567, 667, 685, 745, 755, 1095, 1135, 1155, 1235, 1315, 1345, 1385, 1402, 1435, 1535, 1545, 1625, 1725, 1805, 1815, 1925, and 2030</td>
</tr>
<tr>
<td>Second selection</td>
<td>Five WGS</td>
<td>567, 667, 715, 1345, 1402, 1725, 1925, 2015</td>
</tr>
<tr>
<td></td>
<td>Seven WGS</td>
<td>567, 667, 745, 1345, 1402, 1545, 1725, 1925</td>
</tr>
</tbody>
</table>

Table 2. Frequency numbers of samples classified into seven categories and rates of misclassification in Bayesian analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Categories</th>
<th>Barnyard grass</th>
<th>Green foxtail</th>
<th>Crabgrass</th>
<th>Goosegrass</th>
<th>Chenopodium quinoa</th>
<th>Cabbage '8398'</th>
<th>Cabbage 'zhonggan 11'</th>
<th>Overall samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training samples</td>
<td>Barnyard grass</td>
<td>19</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Green foxtail</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Crabgrass</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Goosegrass</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Chenopodium quinoa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cabbage '8398'</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cabbage 'zhonggan 11'</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Overall samples</td>
<td>19</td>
<td>20</td>
<td>22</td>
<td>19</td>
<td>20</td>
<td>19</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>Testing samples</td>
<td>Barnyard grass</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Green foxtail</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Crabgrass</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Goosegrass</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Chenopodium quinoa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cabbage '8398'</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cabbage 'zhonggan 11'</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Overall samples</td>
<td>11</td>
<td>10</td>
<td>8</td>
<td>11</td>
<td>10</td>
<td>13</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>Rates of misclassification</td>
<td>Prior probability</td>
<td>0.1429</td>
<td>0.1429</td>
<td>0.1429</td>
<td>0.1429</td>
<td>0.1429</td>
<td>0.1429</td>
<td>0.1429</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>Rates of misclassification of the training sample set</td>
<td>0.0500</td>
<td>0.0000</td>
<td>0.0500</td>
<td>0.1000</td>
<td>0.0000</td>
<td>0.2500</td>
<td>0.2000</td>
<td>0.0929</td>
</tr>
<tr>
<td></td>
<td>Rates of misclassification of the testing sample set</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.3000</td>
<td>0.6000</td>
<td>0.1571</td>
</tr>
</tbody>
</table>
Table 3. Frequency numbers of samples classified into six categories and rates of misclassification in Bayesian analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Barnyard grass</td>
</tr>
<tr>
<td>Barnyard grass</td>
<td>20</td>
</tr>
<tr>
<td>Green foxtail</td>
<td>0</td>
</tr>
<tr>
<td>Crabgrass</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3. Frequency numbers of samples classified into six categories and rates of misclassification in Bayesian analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Barnyard grass</td>
</tr>
<tr>
<td>Barnyard grass</td>
<td>20</td>
</tr>
<tr>
<td>Green foxtail</td>
<td>0</td>
</tr>
<tr>
<td>Crabgrass</td>
<td>1</td>
</tr>
<tr>
<td>Goosegrass</td>
<td>1</td>
</tr>
<tr>
<td>Chenopodium quinoa</td>
<td>0</td>
</tr>
<tr>
<td>Cabbage</td>
<td>0</td>
</tr>
<tr>
<td>Overall samples</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 3. Frequency numbers of samples classified into six categories and rates of misclassification in Bayesian analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Barnyard grass</td>
</tr>
<tr>
<td>Barnyard grass</td>
<td>10</td>
</tr>
<tr>
<td>Green foxtail</td>
<td>0</td>
</tr>
<tr>
<td>Crabgrass</td>
<td>0</td>
</tr>
<tr>
<td>Goosegrass</td>
<td>0</td>
</tr>
<tr>
<td>Chenopodium quinoa</td>
<td>0</td>
</tr>
<tr>
<td>Cabbage</td>
<td>0</td>
</tr>
<tr>
<td>Overall samples</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3. Frequency numbers of samples classified into six categories and rates of misclassification in Bayesian analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Barnyard grass</td>
</tr>
<tr>
<td>Barnyard grass</td>
<td>20</td>
</tr>
<tr>
<td>Green foxtail</td>
<td>0</td>
</tr>
<tr>
<td>Crabgrass</td>
<td>1</td>
</tr>
<tr>
<td>Goosegrass</td>
<td>1</td>
</tr>
<tr>
<td>Chenopodium quinoa</td>
<td>0</td>
</tr>
<tr>
<td>Cabbage</td>
<td>0</td>
</tr>
<tr>
<td>Overall samples</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 3. Frequency numbers of samples classified into six categories and rates of misclassification in Bayesian analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Barnyard grass</td>
</tr>
<tr>
<td>Barnyard grass</td>
<td>10</td>
</tr>
<tr>
<td>Green foxtail</td>
<td>0</td>
</tr>
<tr>
<td>Crabgrass</td>
<td>0</td>
</tr>
<tr>
<td>Goosegrass</td>
<td>0</td>
</tr>
<tr>
<td>Chenopodium quinoa</td>
<td>0</td>
</tr>
<tr>
<td>Cabbage</td>
<td>0</td>
</tr>
<tr>
<td>Overall samples</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3. Frequency numbers of samples classified into six categories and rates of misclassification in Bayesian analysis.

From the results of the testing set, all the samples were correctly classified therefore its correct classification rate was 100%. Compared with the previous correct classification rate from which the two varieties of cabbages were considered as two separate categories, the current correct classification rate had been greatly raised.

Conclusion

Using the ASD 350 to 2500 nm FieldSpec-FR spectroradiometer, the canopies of the seedling plants, cabbage ‘8398’, cabbage ‘zhonggan’,
Barnyard grass, green foxtail, goosegrass, crabgrass, and small quinoa, at five- and seven-week growth stages were measured. The results were concluded as follows:

(1) In terms of the load factors and the changing rate of the PCs, eight CWs which were sensitive to plant identification, were determined, respectively for the first growth stage of cabbages (five WGS) and the second growth stage (seven WGS). Among the 8 CWs for each growth stage, only two of them were different, which indicates that different growth stages of the cabbages have limited impact on the plant spectral characteristics for identification of cabbages and weeds.

(2) The corresponding spectral data of the 8 CWs determined from the data at the five WGS were used as the input variables of the Bayesian discriminant model to classify two varieties of cabbages and five different weeds. The correct classification rates for the training and testing sets were 90.7 and 84.3%, respectively. When the two varieties of cabbages were combined into the same category, the correct classification rates of the training and testing sets were improved to 95 and 100%, respectively, which indicates that different varieties of cabbages have similar spectral features to limit weed classification. Therefore, combining different varieties of cabbages as the same category could be effective to greatly improve the correct classification rates of weeds compared with the condition in which two varieties of cabbages were treated as different categories.

(3) The study results showed that Bayesian discriminant analysis could be used to identify weeds from seeding cabbages using leaf hyperspectral reflectance.

Conflict of Interests

The authors have not declared any conflict of interests

ACKNOWLEDGEMENTS

This research was financially supported by the National High Technology Research and Development Program of China (863 Program) (No. 2013AA102406) and the National Sci-Tech Support Plan Projects of China for the 12th five-year plan (2011BAD20B07). The authors acknowledge the National Experimental Station of Precision Agriculture of China, Beijing, China.

REFERENCES


