Detection of aflatoxin B₁ (AFB₁) in individual maize kernels using short wave infrared (SWIR) hyperspectral imaging

Xuan Chu^a, Wei Wang^{*a}, Seung-Chul Yoon^b, Xinzhi Ni^c, Gerald W. Heitschmidt^b

^a College of Engineering, China Agriculture University, Beijing, 100083, China;

^b Quality & Safety Assessment Research Unit, USDA-ARS, Athens, GA 30605, U.S.A;

^c Crop Genetics and Breeding Research Unit, USDA-ARS, 2747 Davis Road, Tifton, GA 31793, U.S.A;

* Corresponding author. College of Engineering, China Agricultural University, No. 17 Tsinghua East Road, Beijing 100083, China. Tel.: +86 10 62737288.

E-mail address: playerwxw@cau.edu.cn (W. Wang)

Abstract

Short wave infrared hyperspectral imaging (SWIR) (1000-2500 nm) was used to detect aflatoxin B₁ (AFB₁) in single maize kernels. One hundred and twenty kernels of four varieties artificially inoculated with a toxigenic strain of *Aspergillus flavus* in the field were examined. Normalisation and principal component analysis (PCA) were applied on average spectra of each kernel to reduce dimensionality and noise. Combining with support vector machine (SVM) classification methods, the first five principal components (PCs) were used to qualitatively classify the AFB₁ contamination levels (<20 ppb, 20-100 ppb, ≥100 ppb) in single kernels without effect of maize variety. Classification accuracies were 83.75% and 82.50% for calibration and validation set, respectively. It was also noted that a general correlation exists between categorical AFB₁ content and the first three PCs. Coefficients of determination (R²) of the support vector machine regression model were 0.77 and 0.70 for calibration and validation set separately. A possible distribution map of AFB₁ was also made by applying the regression model on every pixel of the hyperspectral image. Moreover, using loading plots of the mutual first three PCs, five wavelengths (1317, 1459, 1865, 1934 and 2274 nm) were selected as characteristic wavelengths. Results indicated that hyperspectral imaging could be used to classify AFB₁ level qualitatively in individual maize kernels, however

the performance of predicting the categorical AFB₁ content still needs to be improved.

Keywords: Aflatoxin B₁; Maize kernels; Near-infrared hyperspectral Imaging; Support vector machine (SVM)

1 Introduction

Aflatoxins, a family of mycotoxin with high toxicity, are known to be potent carcinogens and teratogens of livestock and humans (Piva et al., 1995). They are produced by several species of *Aspergillus* fungi, such as toxigenic strains of *A. flavus*, *A. parasiticus* and *A. nomius* (Betina, 1989). A wide variety of food items including peanuts, maize, wheat, rice, cotton seed and chili pepper are all susceptible to contamination by those fungi (Smeesters et al., 2015). Compared to other grains, maize is easier to be infected with those toxigenic fungi because it is a rich nutrient source for the fungi. Once the environment is suitable, aflatoxins are produced by the fungi under either field or storage conditions (Hesseltine, 1974). The aflatoxins contaminating agricultural products can not only pose serious health risks to humans and animals, but also have negative effects on international trade (Filazi, 2013). Due to aflatoxin's deleteriousness and frequent occurrence in agricultural products, the research on aflatoxin reduction and elimination has gained global attention. Generally in the USA the accepted aflatoxin level in food, as established by the Food and Drug Administration (FDA), is 20 ppb.

Among the aflatoxins, aflatoxin B₁ (AFB₁) is known as the most common and the strongest carcinogen (Berthiller et al., 2014). Conventional detection methods for AFB₁ include high-performance liquid chromatography (HPLC), thin layer chromatography (TLC) and enzyme-linked immunosorbent assay (ELISA) (Stroka, 2002; Scussel, 2003; Herzallah, 2009). All these methods require specialised instrumentation and involve a tedious extraction procedure (Tripathi et al., 2009). These methods are generally time consuming, require destructive sampling and trained personnel, although they have high accuracy, sensitivity and dependability. These negative factors limit the feasibility of routine use (Christensen et al., 2008; Lee et al, 2015). Therefore, rapid and non-destructive techniques are needed for identification of AFB₁ in cereals, which may be suitable for real-time

47 detection.

48 Spectroscopic techniques, specifically near-infrared spectroscopy, are being increasingly utilised to detect
49 mycotoxins in agricultural products. Lee et al. (2015) reported that Fourier transform near-infrared reflectance
50 spectroscopy (FT-NIR) can be an alternative method for aflatoxin detection in maize. Pearson et al. (2001)
51 classified kernels containing either high (>100 ppb) or low (<10ppb) levels of aflatoxin by NIR transmittance
52 spectra (500 to 950 *nm*) and reflectance spectra (550 to 1700 *nm*), with classifying accuracy over 95%.
53 Fernández-Ibañez et al. (2009) discriminated maize contaminated with AFB₁ from the toxin-free control by NIRS
54 with coefficient of determination (R^2) of 0.80 and 0.82. All these studies indicated that NIR spectrometry has great
55 potential to be used for determining aflatoxins in maize kernels. However, this method usually gets information
56 from a limited area on the samples, which has certain drawbacks by offering no information about the spatial
57 dimension (Williams et al., 2012). In fact, Pearson et al. (2004) pointed out that the distribution of mycotoxin in a
58 grain pile is uneven, and removing a small percentage of contaminated kernels is a more reasonable method to
59 reduce the mycotoxin contamination than discarding the entire batch of kernels. Thus, it is important to be able to
60 detect mycotoxin in an individual kernel. Cucullu et al. (1966) indicated that the toxins are not uniformly
61 distributed within contaminated kernels. Thus, studying the distribution of mycotoxins has the potential to reduce
62 loss and prevent contamination from early fungal infection.

63 Hyperspectral imaging is a combination of both conventional imaging (spatial) and spectroscopy (spectral)
64 (Senthilkumar et al., 2016). The hyperspectral images are three-dimensional arrays ($m \times n \times \lambda$), where m and n are the
65 spatial axes and λ is the spectral information (Williams et al., 2012). The spectral wavelength range can be divided
66 into visual near infrared (VIS-NIR) hyperspectral imaging (400-1000 *nm*) and short wave infrared (SWIR)
67 hyperspectral imaging (1000-2500 *nm*).

68 In the past decades, hyperspectral imaging has become an important non-destructive detection method for both

69 fungi and mycotoxins (Yao et al., 2008, Williams et al., 2012, Senthilkumar et al., 2016). Wang et al. (2014, 2015a)
70 determined the concentration of pure AFB₁ deposited on the surface of clean and healthy maize kernels. They
71 selected the white stains of AFB₁ drips on the kernels as regions of interest (ROI) from VIS-NIR and SWIR
72 hyperspectral images, and calculated their average spectra. Discriminant models were built by the VIS-NIR and
73 SWIR spectra respectively, and their accuracies were 98% and 88%. The results indicated that it is possible to
74 discriminate various concentrations of AFB₁ that have been applied to the kernel surfaces by both VIS-NIR and
75 SWIR hyperspectral images. Thus, a laboratory-based feasibility study was designed using controlled samples in a
76 controlled environment, and built relationships between pure AFB₁ on maize kernel surfaces and hyperspectral
77 images. For the kernels contaminated with AFB₁ in nature, some ingredient may be consumed by AFB₁ producing
78 fungi. Taking kernels extracted from cobs which were artificially infected with aflatoxin producing fungi in field as
79 samples is also important. Yao et al. (2006, 2010, and 2010) analysed the relationship between fluorescence
80 emissions of maize kernels inoculated with *Aspergillus flavus* and their aflatoxin content by VIS-NIR hyperspectral
81 imaging system with UV light. However, this method encounters both false-positive and false-negative detection,
82 because the measured bright greenish-yellow fluorescence (BGYF) could also be emitted by other intermediate
83 metabolites, i.e., kojic acid and oxidised form of kojic acid. Thus the BGYF approach is only used as a presumptive
84 test method (Pearson et al., 2004). Casasent et al. (2004) obtained VIS-NIR hyperspectral imaging data from maize
85 kernels and classified the aflatoxin negative and positive kernels using a feature-reduction methods. They showed
86 that hyperspectral data seem to be useful in detecting aflatoxin in whole corn kernels. Wang et al. (2015c) used
87 SWIR hyperspectral images to exhibit the distribution of AFB₁ in contaminated maize kernels in terms of pixels,
88 and analysed the of AFB₁ content level (< 10 ppb or ≥ 100 ppb) in kernels. The author mainly analysed the
89 difference between spectra of each pixel. Considering that the chemical reference value of AFB₁ concentration was
90 measured based on each whole kernel, the average spectrum of the corresponding kernel may have more direct

correlation to the AFB₁ content than the spectrum of each isolated pixel in this kernel.

The objective of the present study was to detect the AFB₁ in individual maize kernels without the effect of varieties by hyperspectral images. The specific objectives were (1) to classify the AFB₁ level in each maize kernel; (2) to obtain the correlation between hyperspectral image data and AFB₁ content, and explore the feasibility of detecting categorical AFB₁ content in every single kernel and the spatial distribution of AFB₁; (3) to find some characteristic wavelengths that can be used to distinguish AFB₁ contaminated maize kernels from healthy ones, based on the comprehensive analysis of AFB₁ level identification and content prediction model.

2 Materials and Methods

2.1 Sample preparation

In order to obtain maize kernel samples containing AFB₁, a field inoculation experiment was carried out. A conidial suspension of a toxigenic *Aspergillus flavus* strain (ATCC #200026), which only produces aflatoxin B, mainly aflatoxin B₁, was prepared and diluted to 1×10^6 spores ml⁻¹ with sterile distilled water. In order to test the applicability of the proposed method for different varieties of maize, ears of four different varieties of maize ('BH8740VTTP', 'BH9051RR', 'P31G98', and 'DK697') were inoculated during the soft dough stage with the conidial suspension, following the method described by Henry et al. (2010). Maize kernels were prepared by drying and decorticating mature ears. A 365 nm UV light source was used to identify kernels exhibiting BGYP. The fluorescing kernels and adjacent kernels were selected and labelled as infected; the remaining kernels were labelled as uninfected controls. Overall, 120 (30 per variety) kernels were obtained and arranged on a dimpled tray with the germ upward for hyperspectral image acquisition. In each variety, high AFB₁ content kernels were mostly on the left side. The kernels and their arrangements are shown in Fig. 1. After image acquisition, the actual content of AFB₁ in every individual kernel was measured.

2.2 Hyperspectral imaging acquisition and preprocessing

The SWIR hyperspectral imaging system adopted in this study was the same as that used by Wang et al.

(2014). The system has a motorised linear stage (TLRS300B, Zaber Technologies, Inc., Vancouver, British Columbia, Canada), two halogen lamps with a power of 50 W and a line scan imaging spectrograph with a Mercury Cadmium Telluride (MCT) digital camera (MCT-851 XC403, XenICs, Leuven, Belgium) with pixel number of 320 (spatial) $\times$ 256 (spectral), spectrograph (Hyperspec SWIR, Headwall Photonics, Fitchburg, MA, USA) with the resolution of 7.4 nm and wavelength ranging from 850 to 2800 nm, 30.7 mm front lens (OLES30, Specim, Oulu, Finland), and a spectral flattening filter (SFSWIR/40, Autovision, Los Angeles, CA, USA). The system was corrected by the method described by Wang et al. (2014). During image acquisition, the tray with samples was placed in a custom-made transparent box, which was used to prevent releasing AFB₁. The box was attached to the motorised linear stage. Image acquisition was performed by in-house software.

The hyperspectral image data was calibrated by the reflectance value of a 75% Spectralon® reflectance panel and a dark current image, and the spectral range of the hyperspectral images was reduced to 1000-2500 nm, which was the functional range of the spectrograph and front lens. The size of each hyperspectral data cube was changed to 320 (W) $\times$ 320 (H) $\times$ 197 (wavelengths). Minor lighting differences, which were caused by turning off the light source between image captures, were compensated by the wavelength-dependent offset. With the first image collected serving as the primary, the remaining images were corrected to match the average reflectance of the primary image. After image calibration, spectral extraction and offset corrected, a mosaic containing all 4 images was constructed. Masking and principal component analysis (PCA), as described in Wang et al. (2015c), were used to remove background and reduce the irregular shadows and bad pixel lines on it. The average reflectance spectrum of each kernel on the image was calculated and shown in Fig. 2.

2.3 Chemical analysis for sample reference

After imaging, UV-illuminated photos were taken and then VICAM AflaTest® was used to measure the actual content of AFB₁ of every single kernel. AflaTest® is an AOAC (Association of Official Analytical Chemists), and FGIS (Federal Grain Inspection Service) certified method, which has been widely used in human food and animal

137 feed industries worldwide. The AflaTest® provides total readings for all aflatoxins, and its detection level is as low
138 as 0.1 ppb (VICAM, 2014; 2015).

139 **2.4 Multivariable data analysis method**

140 **2.4.1 Principal component analysis**

141 PCA is mostly used as a tool for data dimensionality reduction. It summarises data by converting the original
142 variables into a set of linearly uncorrelated new variables called principal components (PCs) (Chu, 2011). The
143 accumulative contribution rates indicate the ability of the corresponding PCs to explain the original variables. The
144 coefficient of each wavelength in the corresponding PCs can also provide additional information about the key
145 wavelengths in each of the PCs (Xing et al., 2010; Manley et al., 2009).

146 **2.4.2 Support vector machine (SVM)**

147 SVM was put forward by Vapnik (1995) based on statistical learning theory and has been widely used in
148 classification and regression problems. In SVM-classification, VIS/NIR spectra were used as X-data matrix, and
149 the response factor (Y-data matrix, label classes) were used as treatment names for categories (Abu-Khalaf et al.,
150 2014). The concept of SVM comes from an optimal hyper-plane of linear discriminant analysis, which can separate
151 two classes of samples accurately and make the largest classification interval between them (Melgani et al., 2004).
152 When the linear boundary in the low dimension input space is not enough to separate the two classes, SVM can
153 create a hyperplane that allows linear separation in the higher dimension feature space (Teye et al., 2013). Support
154 vector regression (SVR) is an extension of support vector machine classification, which aims to cast the original
155 data into a feature space of high dimensionality using mapping functions (usually nonlinear). Similar to SVM
156 classification analysis, SVR adopts the structural risk minimisation principle and finds a linear relation between the
157 independent and the dependent variables (Feng et al., 2012; Kao et al., 2013).

158 In SVM, the higher-dimensional space is generally implemented by a kernel function. There are some

common kernel functions, i.e., Linear Function, Polynomial Function (Poly), Radial Basis Function (RBF) and Sigmoid Function (Sigmoid) (Liang et al., 2015). In this study, RBF was selected as the kernel function, and the grid search technique was used to find the optimal model parameter values.

3 Results and Discussion

3.1 Chemical analysis result

The AFB₁ content of the samples ranged from 0 ppb to 3800 ppb. In the UV-illuminated photos, most of the kernels containing above 20 ppb AFB₁ exhibited BGYP. Four kernels with measured AFB₁ of 30 ppb, 41 ppb and 40 ppb exhibited no BGYP. These kernels were false negative kernels. The AFB₁ contents of these kernels were all close to 20 ppb, and their surface was intact, which may explain the false negative. There were also some kernels that contained AFB₁ lower than 20 ppb that exhibited BGYP. Zuzana et al. (2014) pointed out that there were other secondary metabolites, i.e. kojic acid and an oxidised form of kojic acid, which could cause BGYP.

Table 1 shows the AFB₁ content of individual kernels in each variety. There were 71 kernels with less than 20 ppb, 24 kernels with a range between 20 and 100 ppb, 25 kernels were over 100 ppb, within which there were 13 kernels with over 2,000 ppb. Further information on the AFB₁ content is shown in Table 2, the calculated skewness and kurtosis were 2.51 and 4.57, respectively. Compared with the parameters of Normal distribution (skewness equals to 0 and kurtosis to 3), the general distribution mode of AFB₁ content of all kernels in this work is right-skewed, excess kurtosis and heavy tails (Bai et al., 2005). There was a large standard deviation in the AFB₁ content, caused by the uncontrollability and unpredictability of AFB₁ production in nature.

3.2 Spectral data preprocessing

The spectral reflectance intensity was only in the range of 10 to 90, while the AFB₁ content of the maize samples ranged from 0 to 3800 ppb, which crossed several orders of magnitude. In order to minimise the data prediction error caused by the magnitude difference, normalisation was adopted to adjust all data into the interval of [0, 1]. The data were adjusted by subtracting the minimum value from the raw data and dividing this sum by

182 minimum value subtracted from maximum value (Wang, 2013). For modelling, 80 of the 120 kernels were
183 randomly selected as a calibration set and the rest was treated as a validation set. Then PCA was used to reduce the
184 dimension of spectral data. The accumulative contribution rates of the first 10 PCs are listed in Table 3.

185 In this study, different numbers of PCs were used in the classification and correlation of hyperspectral data
186 with AFB₁ content. The proper numbers of PCs were determined based on accumulative contribution rates and the
187 performance of the classification and correlation.

188 **3.3 Classification results of AFB₁ level**

189 Firstly, SVM was used to analyse the AFB₁ level in each kernel qualitatively. With the 20 ppb and 100 ppb
190 thresholds, the samples were divided into three groups; asymptomatic group (AG, ≤ 20 ppb), in which the AFB₁
191 contents were below 20ppb; contaminated group (CG), in which the AFB₁ contents were between 20ppb and
192 100ppb; heavily contaminated group (HG), in which the AFB₁ contents were above 100 ppb. The first 1 to 10 PCs
193 were taken gradually as the input of SVM classifier, and the classification accuracies were used to estimate the
194 optimal number of PCs (Fig. 3).

195 Fig. 3 shows that maximum accuracies in the calibration and validation set were achieved when the first five
196 principal components were used for modelling. After that, although the accuracies fluctuated a little with the
197 increasing number of PCs, the gap of accuracies between the calibration and validation sets became obvious. The
198 accumulative contribution rate reached 99.9% (Table 3), which mean that the first 5 PCs could represent almost all
199 information of the original spectra. Thus, the first five PCs were selected as the input of SVM classifier. The
200 classification results of AG, CG and HG groups were summarised in Table 4.

201 The overall accuracy was 83.75% and 82.5 %, respectively in calibration and validation sets. The results
202 indicated that this classifier could distinguish AFB₁ level without any interference from maize varieties.
203 Classification accuracies for AG and HG groups were much better than that for the CG group. In other words, this
204 classifier had difficulty in differentiating kernels with AFB₁ content ranging from 20 to 100 ppb, which agreed with

a previous finding (Yao et al., 2010) that it was hard to identify maize kernels with the AFB₁ content ranging from 10 to 100 ppb by fluorescence hyperspectral images. The classification accuracies of HG were lower than AG. An important factor was the uneven number of kernels in each group. There were only six kernels in HG of the validation set. Thus even one misjudgement would greatly reduce the classification accuracy. Besides, AG contained over twice the number of kernels as HG, which may cause an unbalanced training of the model in SVM classifier. When calibration sets with uneven class sizes are used, the classification error based on SVM is undesirably biased towards the class with fewer samples in the calibration set (Zhang et al., 2007). Thus, this work only showed the potential of this technology to identify the AFB₁ level in single kernel. In further experimentation, the number of samples in each AFB₁ level should be added.

This model can only classify the AFB₁ level. Further, detecting the categorical AFB₁ content in per kernels would be better for toxin detection. Thus a correlation between the spectra and the AFB₁ contents in individual maize kernels was further explored.

3.4 Regression results between hyperspectral data and AFB₁ content

A method of SVR was employed for obtaining the correlation between hyperspectral data and AFB₁ content. The way of determining the optimal number of PCs was similar to that for classification analysis. The coefficient of determination (R^2) of the calibration set (R_c^2) and the validation set (R_v^2), root mean square error (RMSE) of the calibration set (RMSEC) and the validation set (RMSEV), and relative percent deviation (RPD) were used for estimating the regression model. The R_c^2 and R_v^2 of the regression models which were built by different number of PCs are summarised in Fig. 4.

The highest R^2 for the two sets (0.79, 0.71) were obtained when the first eight principal components were taken as input of SVR (Fig.4). However, when the first three principal components were employed, the model also showed good fitness (0.77, 0.70). RMSE in the calibration and validation sets of three-PC model were lower than

eight-PC model. RPD of these two models were both in the range of 1.4-2.0 which indicated that these two models are reasonable (Williams et al., 1987). Considering these factors comprehensively, the first three PCs were selected in the regression analysis. Figs. 5a and 5b show the correlation between reference measurement and prediction in the calibration and validation sets. The regression model showed a big standard deviation for both calibration and validation sets (Fig. 5). This may be explained by the following reasons, 1) the range of AFB₁ content in all samples was grate, which was 0 ppb-3800ppb, and 2) the distribution of AFB₁ content in contaminated samples was highly skewed, that is, most of the kernels contained AFB₁ was below 20 ppb. Overall, this model can only make a general prediction. Thus, a larger sample set with a more balanced AFB₁ distribution is needed.

A possible AFB₁ distribution map for the kernels in BH8740VTTP was calculated by applying the regression model to every pixel of the hyperspectral image. This variety was chosen because their infection of AFB₁ is obvious. This map can visualise the possible distribution of AFB₁ in different kernels, even within the kernels. The kernels with high AFB₁ content were mostly on the left side (Fig. 6a), which was consistent with the actual sample arrangement during image acquisition (Fig. 1). The distribution of AFB₁ was not uniform within individual kernels, which was consistent with the report by Cucullu et al. (1966). The AFB₁ is mainly distributed on the germ part of the kernels. This was similar to the distribution map given by the multidimensional PC scatter plot (Fig. 6b). The consistency between possible distribution map and actual situation indicated that the regression model could make a general prediction. The results also suggests that visible hyperspectral imaging could become an important and useful tool to assess the quality of a sample of maize.

3.5 Identification of characteristic wavelengths

The first five and first three PCs were respectively used in the classification and correlation of individual kernel hyperspectral data with AFB₁ content. Thus, the shared first three PCs were considered as important variables, and the loading line plots of the first three PCs were adopted to identify characteristic wavelengths. Those wavelengths with larger absolute coefficients in each PC were considered to play a relatively important role

250 in representing the corresponding PC. Wavelengths of 1317 nm and 1459 nm are peaks of the loading line plot of
251 PC₂, and wavelengths of 1865 nm, 1934 nm and 2274 nm are peaks of the loading line plot of PC₃ (Fig. 7a). The
252 loading of PC₁ with its weight was shown in Fig. 7b, and the loading of PC₂ to PC₃ with their weight were shown
253 in Fig. 7c. All wavelengths had large absolute coefficients in Fig. 7b, and the wavelengths (1317, 1459, 1865, 1934
254 and 2274 nm) showed even more obvious in peaks or valleys in Fig. 7c than that in Fig. 7a. Thus, these five
255 wavelengths were chosen as characteristic wavelengths. If compared with previous research by Wang et al. (2015b;
256 2015c), except for those common wavelengths (1459 and 1865 nm), there was some differences for the
257 characteristic wavelengths selected. Near infrared spectrum technology relies on the statistical analysis of a given
258 set of data. Thus, the existence of differences may be caused by the different data selection methods in the two
259 studies. In previous research, isolated pixel spectra were used to explore characteristic wavelengths, and the first
260 two PCs which represent macroscopic characteristics of kernels such as standard deviation and average value of all
261 studied kernels were removed, while in this work, the average spectrum of each kernel was used and the whole data
262 set of the first three PCs were considered. Therefore, it is normal to have some difference about selected
263 wavelengths by different methods.

264 In the NIR region, most of the bands, especially in peaks and valleys, were attributable to functional groups of
265 the main chemical components (Bendini et al., 2007). During the growth of *A. flavus*, it consumes nutrients for its
266 metabolism, growth and development (Wei et al., 2007). Thus, the presence of the fungus and production of AFB₁
267 could affect such kernel characteristics as water, starch, lipids or protein structure. Any of this change would be
268 manifested as a difference in reflectance spectra between healthy (control) kernels and AFB₁ contaminated kernels.
269 According to the previous reports, the characteristic wavelengths all corresponded to the components or the main
270 functional groups in maize kernels that the fungus would utilise. The prominent peak around 1459 nm was
271 attributed to N-H stretching 1st OT of CONH₂ (Wang et al., 2015c), while 1865 nm was ascribed to the C-C₁ stretch

(Wang et al., 2015b). In a similar manner, the peak at 1930 *nm* that is close to 1934 *nm* was mainly associated with O-H stretch functional group of water and starch (Burns et al., 2007), and the peak at 2274 *nm* was the result of O-H deformation and stretching modes of the alcohol group (Wang et al., 2014). The regions of 1415-1512 *nm* and 1830–1880 *nm* that contain 1456 *nm* and 1865 *nm* were related to the phenolic component in kernels (Zhang et al., 2008). Thus, we conclude that the calibrations are relying on changes in kernel composition related to AFB₁ content. The correlation between chemical substances and the characteristic wavelengths is the reason why the wavelengths could play a principal role in analysis.

4 Conclusions

The current study examined the possibility of using hyperspectral imaging to detect AFB₁ produced by artificial inoculation of maize kernels with *A. flavus* conidial spore suspension. The conclusions from the study are:

(1) SWIR hyperspectral images can be used to classify the AFB₁ level in maize kernels. The overall classification accuracies of the three groups (<20 ppb, 20-100 ppb, >100 ppb) reached 83.75% and 82.50% for the calibration set and the validation set, respectively. Specifically, the classification accuracy of kernels reached 95.56%, 96.15% for low level (<20 ppb) and 82.35%, 75.00% for high levels (>100ppb).

(2) A modest correlation was also discovered between the hyperspectral image data and AFB₁ content of individual kernels. The regression model resulted in $Rc^2=0.77$ and $Rv^2=0.70$. It is more difficult to predict exact content than to categorise AFB₁ level. The possible distribution map indicated that the AFB₁ distribution were basically consistent with the actual situation, which also indicated that the hyperspectral image is a useful tool to assess the quality of maize.

(3) The first three PCs played an important role in classification and correlation of hyperspectral data with AFB₁ content. Some key wavelengths (1317, 1459, 1865, 1934 and 2274 *nm*) were selected from the loading line plots of the first three PCs, and the wavelengths all corresponded to the variation of the main nutritional

294 components of maize such as water, starch, lipids and alcohol, which may be affected by fungal activity.

295 In general, the results illustrated that hyperspectral imaging is a useful tool to detect AFB₁ in individual maize
296 kernels without being affected by variety. Furthermore, in order to further reveal the truth and improve the
297 robustness of the proposed method, more comparative experiments using maize samples of same cultivars
298 inoculated with atoxigenic strains, as well as control samples not inoculated, will be carried out gradually. In
299 addition, in vitro tests using sterilised maize kernels inoculated with producing and nonproducing strains should
300 also be conducted in our future research.

301 **Acknowledgements**

302 The authors would like to express their gratitude to Dr. Kurt C. Lawrence, Research Leader, with the Quality
303 & Safety Assessment Research Unit, USDA, ARS, for his kind help to provide experimental conditions and
304 guidance of experimental implementation. The authors also thank Penny J. Tapp and Trevor Perla in the Crop
305 Genetics and Breeding Research Unit, USDA-ARS, Tifton, GA, for their assistance in inoculation, corn sample
306 collection and aflatoxin analysis; and Vernon Savage and Jerrie Barnett in the Quality & Safety Assessment
307 Research Unit, USDA-ARS, Athens, GA, for their technical assistance in sample preparation and image
308 acquisition. This work was supported financially by the China National Science and Technology Support Program
309 (Grant no. 2012BAK08B04).

310 **Reference**

- 311 Abu-Khalaf, N., & Salman, M. (2014). Visible/Near infrared (VIS/NIR) spectroscopy and multivariate data analysis (MVDA) for
312 identification and quantification of olive leaf spot (OLS) disease. *Palestine Technical University Research Journal*, 2(1), 1-8.
- 313 Bai, J., & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*,
314 23(1), 49-60.
- 315 Bendini, A., Cerretani, L., Di Virgilio, F., Belloni, P., Bonoli-Carbognin, M., & Lercker, G., (2007). Preliminary evaluation of the

316 application of the FTIR spectroscopy to control the geographic origin and quality of virgin olive oils. *Journal of Food Quality*.
317 30(4), 424-437.

318 Berthiller, F., Burdaspal, P. A., Crews, C., Iha, M. H., Krska, R., Lattanzio, V. M. T., & Stroka, J., (2014). Developments in mycotoxin
319 analysis: An update for 2012-2013. *World Mycotoxin Journal*. 7(1), 3-33.

320 Betina, V., (1989). Mycotoxins. Chemical, *Biological and Environmental Aspects*. Elsevier, Amsterdam, 395(5), 463–503.

321 Burns, D. A., & Ciurczak, E. W. (Eds.). (2007). *Handbook of near-infrared analysis*. CRC press, 356-357.

322 Casasent, D., & Chen, X. W., (2004). Aflatoxin detection in whole corn kernels using hyperspectral methods. Optical Technologies for
323 Industrial, Environmental, and Biological Sensing. International Society for Optics and Photonics, 275-284.

324 Christensen, D., Alles, M., Rosenkrands, I., Rantanen, J., Foged, C., Agger, E. M., & Nielsen, H. M., (2008). NIR transmission
325 spectroscopy for rapid determination of lipid and lyoprotector content in liposomal vaccine adjuvant system CAF01. *European*
326 *Journal of Pharmaceutics and Biopharmaceutics*. 70(3), 914-920.

327 Chu X., (2011). Molecular spectroscopy analytical technology combined with chemometrics its applications. Chemical Industry
328 Press, 55-57.

329 Cucullu, A. F., Lee, L. S., Mayne, R. Y., Goldblatt, L. A. (1966). Determination of aflatoxins in individual peanuts and peanut sections.
330 *Journal of the American Oil Chemists Society*, 43(2), 89-92.

331 Dong, L., Yang, Q., Huang, W., Peng, D., Zhao, J., & Huang, L., et al., (2015). Estimation of leaf area index based on wavelet
332 transform and support vector machine regression in winter wheat. *Infrared & Laser Engineering*, 44(1), 335-340.

333 Dowell, F. E., Pearson, T. C., Maghirang, E. B., Xie, F., Wicklow, D. T., (2002). Reflectance and transmittance spectroscopy applied to
334 detecting fumonisin in single corn kernels infected with *Fusarium verticillioides*. *Cereal Chemistry*. 79(2), 222-226.

335 Feng, Y. Z., Sun, D. W., ElMasry, G., & Morcy, N. (2012). Direct prediction of Enterobacteriaceae loads in chicken fillets using near
336 infrared (NIR) hyperspectral imaging and support vector regression (SVR).

337 Fernández-Ibañez, V., Soldado, A., Martínez-Fernández, A., De la Roza-Delgado, B., (2009). Application of near infrared spectroscopy

338 for rapid detection of aflatoxin B1 in maize and barley as analytical quality assessment. *Food Chemistry*. 113(2), 629-634.

339 Filazi, A., & Tansel, U., (2013). Occurrence of aflatoxins in food. *Intech Open Access Publisher*, 143-170.

340 Gotsis, P. K., Chamis, C. C., & Minnetyan, L., (2004). Classification of hyperspectral remote sensing images with support vector
341 machines. *IEEE Transactions on Geoscience & Remote Sensing*. 19(3), 753-773.

342 Henry, W. B., Krakowsky, M. D., Windham, G. L., Williams, W. P., Scully, B. T., Rowe, D., & Hawkins, L. K., (2010). Comparison of
343 the side-needle and knife techniques for inducing *Aspergillus flavus* infection and aflatoxin accumulation in corn hybrids. *Toxin*
344 *Reviews*. 29(3-4), 123-129.

345 Herzallah, S. M., (2009). Determination of aflatoxins in eggs, milk, meat and meat products using HPLC fluorescent and UV
346 detectors. *Food Chemistry*. 114(3), 1141-1146.

347 Hesseltine, C. W., (1974). Natural occurrence of mycotoxins in cereals. *Mycopathologia et Mycologia Applicat*, 53(1-4), 141-153.

348 Hruska, Z., Yao, H., Kincaid, R., Brown, R., Cleveland, T., & Bhatnagar, D. (2014). Fluorescence Excitation–Emission Features of
349 Aflatoxin and Related Secondary Metabolites and Their Application for Rapid Detection of Mycotoxins. *Food and bioprocess*
350 *technology*, 7(4), 1195-1201.

351 Kao, L. J., Chiu, C. C., Lu, C. J., & Yang, J. L. (2013). Integration of nonlinear independent component analysis and support vector
352 regression for stock price forecasting. *Neurocomputing*, 99, 534-542.

353 Karimi, Y., Prasher, S. O., Patel, R. M., & Kim, S. H., (2006). Application of support vector machine technology for weed and nitrogen
354 stress detection in corn. *Computers and Electronics in Agriculture*. 51(1), 99-109.

355 Lasaponara, R., 2006. On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from
356 SPOT/VEGETATION NDVI temporal series. *Ecological Modelling*. 194(4), 429-434.

357 Lee, K. M., Davis, J., Herrman, T. J., Murray, S. C., Deng, Y., (2015). An empirical evaluation of three vibrational spectroscopic
358 methods for detection of aflatoxins in maize. *Food Chemistry*. 173, 629-639.

359 Pearson, T. C., Wicklow, D. T., Maghirang, E. B., Xie, F., Dowell, F. E., (2001). Detecting aflatoxin in single corn kernels by

transmittance and reflectance spectroscopy. *Transactions of the ASABE*. 44(5), 1247.

Pearson, T. C., Wicklow, D. T., Pasikatan, M. C., (2004). Reduction of aflatoxin and fumonisin contamination in yellow corn by high-speed dual-wavelength sorting. *Cereal Chemistry*. 81(4), 490-498.

Piva G, Galvano F, Pietri A, et al., (1995). Detoxification methods of aflatoxins. A review. *Nutrition Research*. 15(5): 767-776.

Scussel, V. M., (2003). Comparison of methods by tlc and hptlc for determination of aflatoxin m1 in milk and b1 in eggs. *Ciência E Tecnologia de Alimentos*. 23, 46-52.

Senthilkumar, T., Jayas, D. S., White, N. D. G., (2015). Detection of different stages of fungal infection in stored canola using near-infrared hyperspectral imaging. *Journal of Stored Products Research*. 63, 80-88.

Smeesters, L., Meulebroeck, W., Raeymaekers, S., Thienpont, H., (2015). Optical detection of aflatoxins in maize using one-and two-photon induced fluorescence spectroscopy. *Food Control*. 51, 408-416.

Stroka, J., Anklam, E., (2002). New strategies for the screening and determination of aflatoxins and the detection of aflatoxin-producing moulds in food and feed. *TrAC Trends in Analytical Chemistry*. 21(2), 90-95.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 114, 183-189.

Tripathi, S., Mishra, H. N., (2009). A rapid FT-NIR method for estimation of aflatoxin B 1 in red chili powder. *Food Control*. 20(9), 840-846.

Vapnik, V., (2013) . *The nature of statistical learning theory*. Springer Science & Business Media.

VICAM (2014) Aflatest Fluorometer Instruction Manual, Waters Corporation, Milford, MA. 67.

VICAM (2015) Aflatest-Mycotoxin Testing System, Waters Corporation, Milford, MA, 2.

Wang, W., Heitschmidt, G. W., Ni, X., Windham, W. R., Hawkins, S., Chu, X., (2014). Identification of aflatoxin B 1 on maize kernel surfaces using hyperspectral imaging. *Food Control*. 42, 78-86.

Wang, W., Heitschmidt, G. W., Windham, W. R., Feldner, P., Ni, X., Chu, X., (2015a). Feasibility of detecting aflatoxin B1 on

382 inoculated maize kernels surface using Vis/NIR hyperspectral imaging. *Journal of Food Science*. 80(1), M116-M122.

383 Wang, W., Lawrence, K. C., Ni, X., Yoon, S. C., Heitschmidt, G. W., Feldner, P., (2015b). Near-infrared hyperspectral imaging for

384 detecting Aflatoxin B1 of maize kernels. *Food Control*. 51, 347-355.

385 Wang, W., Ni, X., Lawrence, K. C., Yoon, S. C., Heitschmidt, G. W., Feldner, P., (2015c). Feasibility of detecting Aflatoxin B 1 in

386 single maize kernels using hyperspectral imaging. *Journal of Food Engineering*. 166, 182-192.

387 Wei J., Zhang N., Qi D., (2007). Laws of quality change for four moldy feedstuffs. *Journal of the Chinese Cereals and Oils*, 22 (5):

388 119-124.

389 Williams, P. J., Geladi, P., Britz, T. J., Manley, M., (2012). Growth characteristics of three *Fusarium* species evaluated by near-infrared

390 hyperspectral imaging and multivariate image analysis. *Applied Microbiology and Biotechnology*. 96(3), 803-813.

391 Xing, J., Symons, S., Shahin, M., & Hatcher, D., (2010). Detection of sprout damage in Canada Western Red Spring wheat with

392 multiple wavebands using visible/near-infrared hyperspectral imaging. *Biosystems Engineering*. 106(2), 188-194.

393 Yao, H., Hruska, Z., Brown, R. L., & Cleveland, T. E., (2006). Hyperspectral bright greenish-yellow fluorescence (bgyf) imaging of

394 aflatoxin contaminated corn kernels. *Proceedings of SPIE - The International Society for Optical Engineering*. 6381,

395 63810B-63810B.

396 Yao, H., Hruska, Z., Kincaid, R., Brown, R. L., Cleveland, T. E., (2008). Differentiation of toxigenic fungi using hyperspectral

397 imagery. *Sensing and Instrumentation for Food Quality and Safety*. 2(3), 215-224.

398 Yao, H., Hruska, Z., Kincaid, R., Brown, R., Cleveland, T., Bhatnagar, D., (2010). Correlation and classification of single kernel

399 fluorescence hyperspectral data with aflatoxin concentration in corn kernels inoculated with *Aspergillus flavus* spores. *Food*

400 *Additives and Contaminants*. 27(5), 701-709.

401 Yao, H., Hruska, Z., Kincaid, R., & Ononye, A., (2010). Spectral Angle Mapper classification of fluorescence hyperspectral image for

402 aflatoxin contaminated corn. *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 1 - 4

403 Zhang, C., Shen, Y., Chen, J., Xiao, P., Bao, J., (2008). Nondestructive prediction of total phenolics, flavonoid contents, and

404 antioxidant capacity of rice grain using near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*. 56(18),
405 8268-8272.

406 Zhang, Q. Y., Zhao, F. Q., Wang, J., YU, D. M. J. J., & Zhang, R. H. (2007). Optimization of C-SVM in case of samples with unequal
407 numbers in their different varieties [J]. *Journal of Lanzhou University of Technology*, 33(4), 90-92.

Figure Captions

Fig. 1 Maize kernel samples from the four commercial varieties and their arrangements used in the experiment: (a)BH8740VTTP
(b)DK697 (c)P31G98 (d)BH9051.

Fig. 2 Reflectance spectra (1000-2500 nm) with noise reduction.

Fig. 3 Classification accuracy of SVM classifier with different PCs. As indicated by dotted line, the highest accuracy in
calibration and validation set was obtained when the first five PCs were selected as a compromise for calibration and
validation set

Fig. 4 Correlation coefficients of regression model with different PCs. As indicated by dotted line, the highest R^2 in calibration
and validation set was obtained when the first eight PCs were selected.

Fig. 5 Scatter plot of predicted AFB₁ by regression model and actual AFB₁ content: (a) Calibration set (b) Validation set

Fig. 6 Possible distribution maps of AFB₁ contents in maize kernels: (a) Possible distribution maps of AFB₁ on maize kernels in
the varieties of BH8740VTTP. The blue indicates low AFB₁ content, and the red designates a very high AFB₁ content. (b)
Classification result image obtained by multidimensional PC scatter plot, which was made by Wang et al. (2015c). The
yellow indicates low AFB₁ content and the red designates a very high AFB₁ content

Fig. 7 Loading line plots of the first three PCs: (a) Loading line plots of the first three PCs without weight (b) Loading line plots
of PC₁ with weight (c) Loading line plots of PC₂ and PC₃ with weight.

425 **Table Captions**

426 Table 1 AFB₁ content of individual kernels in each variety

427 Table 2 Statistical information of AFB₁ content of the 120 kernels

428 Table 3 Cumulative contribution rates of the first 10 PCs

429 Table 4 Classification result of SVM classifier

BH8740					DK697					PG98					BH9051				
3400	440	12	5	1	4	69	9	15	24	880	110	4	1	4	180	2	110	59	9
3300	3200	###	5	2	2	##	11	85	2	22	3	6	1	1	150	10	14	0	0
3100	3200	770	4	4	12	10	2	2	3800	26	93	0	40	1	37	25	320	30	1
3100	3100	28	4	1	100	55	2	7	5	42	34	3	1	1	11	97	340	1	0
2600	67	9	#	2	8	4	1	1	73	23	1	2	8	1	28	53	3100	##	4
3200	3400	3	1	3	6	86	7	14	3	640	2	10	1	2	18	5	17	1	41

Parameter	Observation count	Average (ppb)	Maximu m (ppb)	Minimum (ppb)	Median (ppb)	Standard deviation (ppb)	Kurtosis	Skewness
AFB ₁ content	120	398.7	3800	0	10	998.65	4.57	2.51

PCs	PC ₁	PC ₂	PC ₃	PC ₄	PC ₅
Cumulative Contribution Rate (%)	95.026	98.663	99.634	99.880	99.929
PCs	PC ₆	PC ₇	PC ₈	PC ₉	PC ₁₀
Cumulative Contribution Rate (%)	99.957	99.971	99.980	99.986	99.990

		From classification				
From chemical analysis		0-20ppb (AG)	20-100ppb (CG)	Over 100ppb (HG)	Accuracy	Overall accuracy
Calibration set	0-20 ppb (AG)	43	2	0	95.56%	83.75%
	20-100 ppb (CG)	8	10	0	55.56%	
	Over 100 ppb (HG)	1	2	14	82.35%	
Validation set	0-20 ppb (AG)	25	1	0	96.15%	82.50%
	20-100 ppb (CG)	1	2	3	33.33%	
	Over 100 ppb (HG)	2	0	6	75.00%