Research Article

Received: 3 March 2019 Revised: 15 April 2019 Accepted article published: 22 April 2019 Published online in Wiley Online Library: 30 May 2019

(wileyonlinelibrary.com) DOI 10.1002/ps.5448

Assessing crop damage from dicamba on non-dicamba-tolerant soybean by hyperspectral imaging through machine learning

Jingcheng Zhang,[a] Yanbo Huang,[b*] Krishna N Reddy[b] and Bin Wang[a]

Abstract

BACKGROUND: Dicamba effectively controls several broadleaf weeds. The off-target drift of dicamba spray or vapor drift can cause severe injury to susceptible crops, including non-dicamba-tolerant crops. In a field experiment, advanced hyperspectral imaging (HSI) was used to study the spectral response of soybean plants to different dicamba rates, and appropriate spectral features and models for assessing the crop damage from dicamba were developed.

RESULTS: In an experiment with six different dicamba rates, an ordinal spectral variation pattern was observed at both 1 week after treatment (WAT) and 3 WAT. The soybean receiving a dicamba rate ≥0.2X exhibited unrecoverable damage. Two recoverability spectral indices (HDRI and HDNI) were developed based on three optimal wavebands. Based on the Jeffries–Matusita distance metric, Spearman correlation analysis and independent *t*-test for sensitivity to dicamba spray rates, a number of wavebands and classic spectral features were extracted. The models for quantifying dicamba spray levels were established using the machine learning algorithms of naive Bayes, random forest and support vector machine.

CONCLUSIONS: The spectral response of soybean injury caused by dicamba sprays can be clearly captured by HSI. The recoverability spectral indices developed were able to accurately differentiate the recoverable and unrecoverable damage, with an overall accuracy (OA) higher than 90%. The optimal spectral feature sets were identified for characterizing dicamba spray rates under recoverable and unrecoverable situations. The spectral features plus plant height can yield relatively high accuracy under the recoverable situation (OA = 94%). These results can be of practical importance in weed management.

Keywords: dicamba; non-dicamba-tolerant soybean; crop damage; hyperspectral imaging; machine learning

1 INTRODUCTION

Weed management in soybean farming is essential to maximize crop yield. As a substitute herbicide to the classic herbicide glyphosate, Dicamba (3, 6-dichloro-2-methoxybenzoic acid) is able to control several annual and perennial broadleaf weeds that have evolved resistance to the glyphosate. Although dicamba is effective on many broadleaf weeds, its severe drift damage to non-dicamba-tolerant (non-DT) crops has raised public concern about its applications. With commercial DT-trait soybean and cotton, off-target dicamba drift from routine use in DT crops onto susceptible (non-DT) crops has become a major concern, therefore the assessment of crop damage from dicamba drift is important for effective weed management.

Currently, crop damage is mainly determined by assessing physiological and biochemical changes (leaf area, leaf color, plant height, yield, etc.). However, these assessments are labor-intensive. A rapid, cost-effective method is needed for assessment on large scale farms. Hyperspectral imaging (HSI) is a technique that can rapidly scan plant samples. The significance of developing HSI relies in its ability to acquire a complete reflectance spectrum for each pixel in an image, and it has been developed for improving the identification and quantitative determination of biophysical and biochemical properties of plants.[1] HSI has been successfully used for detecting plant stresses caused by toxic metals,[2] salt,[3] diseases and pests.[4] With its ability to acquire abundant spectral information and imaging details of plants, HSI exhibits promising potential in capturing the plant response of abiotic and biotic stresses at both leaf and canopy levels.[5,6]

In detecting, monitoring and quantifying crop damage, a number of machine learning (ML) algorithms have been utilized, such as Bayes' decision,[7] maximum likelihood classification,[8] K-means clustering,[9] random forest,[10] support vector machines[11] and artificial neural networks.[12] Behmann *et al.*[13] mentioned that given that the ML algorithms are able to establish both linear and non-linear

* *Correspondence to: Y Huang, Crop Production Systems Research Unit, United States Department of Agriculture, Agricultural Research Service, PO Box 350, Stoneville, MS 38776, USA. E-mail: yanbo.huang@ars.usda.gov*

a *College of Life Information Science & Instrument Engineering, Hangzhou Dianzi University, Hangzhou, China*

b *United States Department of Agriculture, Crop Production Systems Research Unit, Agricultural Research Service, Stoneville, MS, USA*

models, they require few statistical assumptions and have flexibility on a wide range of data characteristics. As reviewed by Huang *et al.*[14] and Singh *et al.*,[15] the ML algorithms are essential in agricultural engineering and management, especially for high-throughput plant stress phenotyping. These algorithms can be deployed in crop production processes for identification, classification, quantification and prediction of various biotic and abiotic plant stress traits.

As for detecting herbicide injury in crops, Henry *et al.*[16] used hyperspectral data to differentiate soybean and corn plants that are treated with glyphosate and paraquate for the untreated ones. Yao *et al.*[17] proposed a modified method of spectral derivative analysis to improve the detection of soybean plant injury from glyphosate. Based on leaf hyperspectral reflectance spectra, Zhao *et al.*[18] achieved early detection of crop injury from the herbicide glyphosate by PROSPECT (leaf optical PROperty SPECTra model) inversion of several biophysical and biochemical parameters. With the aid of spectral observation and analysis, Huang *et al.*,[19] Ortiz *et al.*[20] and Huang *et al.*[21] illustrated the possibility of detecting and mapping the herbicide injury with aerial remote sensing images.

As an auxinic herbicide, dicamba has profound effects on the growth and structure of plants. As mentioned in Huang *et al.*[22] and Johnson *et al.*,[23] dicamba can cause cupping and crinkling of leaves, twisting and curling of stems and petioles, and disruption of phloem transport leading to chlorosis, growth inhibition, wilting and necrosis on non-DT crops. The biophysical changes in the plants can induce corresponding spectral variations, which allow detection via hyperspectral sensing. Our previous study showed that sprayed dicamba can result in detectable spectral changes as early as 24 h after treatment (HAT).[22] However, the relationship between dicamba spray rates that is critical to weed management and spectral changes over time is not yet clear. It is important to understand the spectral response of soybean plants to different dicamba rates and to develop appropriate assessing methods. With the aid of HSI, the aim of this study was (i) to understand the variation in biophysical parameters, yield and corresponding spectral response under different dicamba spray rates, (ii) to identify appropriate spectral features for assessing the impact of dicamba rates on the plants and (iii) to establish a model using machine learning to assess dicamba rates based on HSI features.

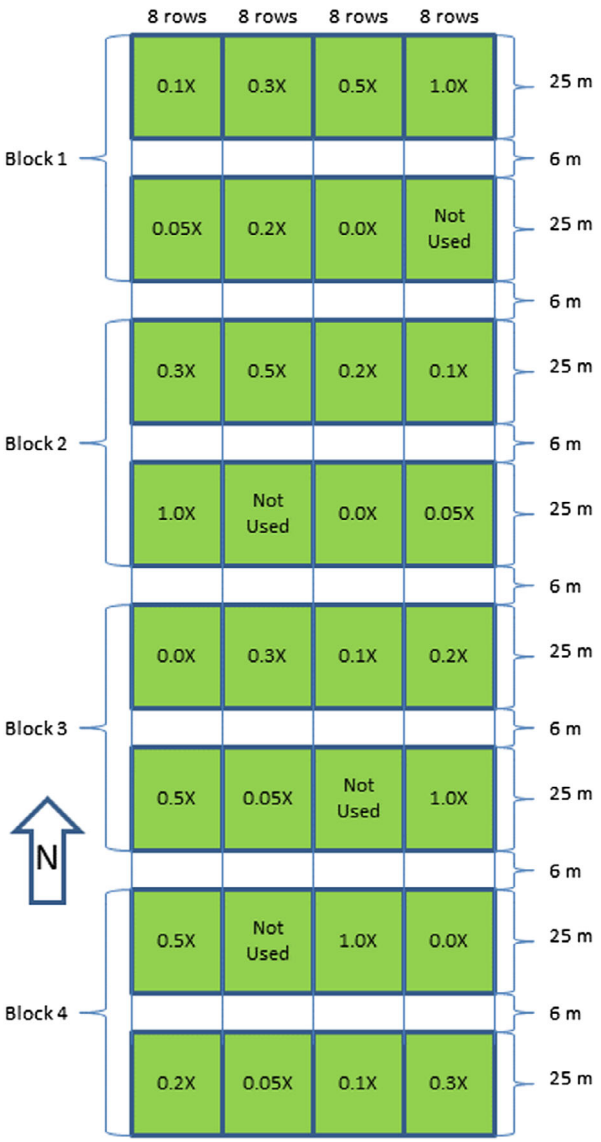

Figure 1. Experiment field layout for soybean dicamba damage study.

2 MATERIALS AND METHODS

2.1 Study area and experiment design

A field experiment was conducted in an experimental field (central latitude: 33.445062° and central longitude: 90.869967°) with a 4.5-ha area at the USDA Agricultural Research Service, Crop Production Systems Research Farm, Stoneville, MS, USA. A randomized complete block experimental design[24] was used in the setup of experimental plots, which guaranteed that the treatments were randomly assigned to the experimental units. The entire experimental field consisted of four blocks. In each block, seven plots of soybeans (Progeny P4819LL, Progeny Ag Products, Wynne, AR, USA) receiving different dicamba spray rates, 0.05X, 0.1X, 0.2X, 0.3X, 0.5X and 1X (where X = 0.56 kg ae ha^{-1}) and a dicamba-free plot (0.0X), respectively, were set, which resulted in a total of 32 plots (4 blocks × 8 plots) of soybeans. Each plot consisted of eight rows with a plot size of 7.76 × 25 m (Fig. 1). The soybeans were planted on 7 May 2014. In late March 2014, the surroundings of the experimental area were planted with maize, which was used to establish a buffer zone to minimize the spray drift of dicamba to/from the neighboring fields.

Since many broadleaf plants are sensitive to dicamba, it is recommended that dicamba is applied in the field at early leaf stage. On 17 June 2014, when the soybean was at the five to six trifoliolate leaf stage, the soybean plots were treated with the dimethylamine salt of dicamba, RIFLE® (Loveland Products, Inc., Greeley, CO, USA) using a tractor-mounted sprayer with Tee Jet 4003 standard flat-spray nozzles delivering 140 L ha^{-1} of water at 193 kPa (Fig. 2). The hyperspectral plant sensing and the measurements of plant height and plant biomass were taken 1 week after treatment (WAT) and 3 WAT. To avoid the influence of various measurements, no postemergence herbicides were applied up to 3 weeks after dicamba treatment. After 3 weeks, other postemergence herbicides were applied as needed to keep the plots weed-free. The field was furrow irrigated as needed. The soybean yield was recorded at the time of harvest on 9 September 2014.

2.2 Measurements of hyperspectral and biophysical data

In each plot, an area with representative growing conditions was chosen for hyperspectral and biophysical measurements at 1 and 3 WAT when each treated and non-treated control plot was

Figure 2. Dicamba spray system and hyperspectral imaging system.

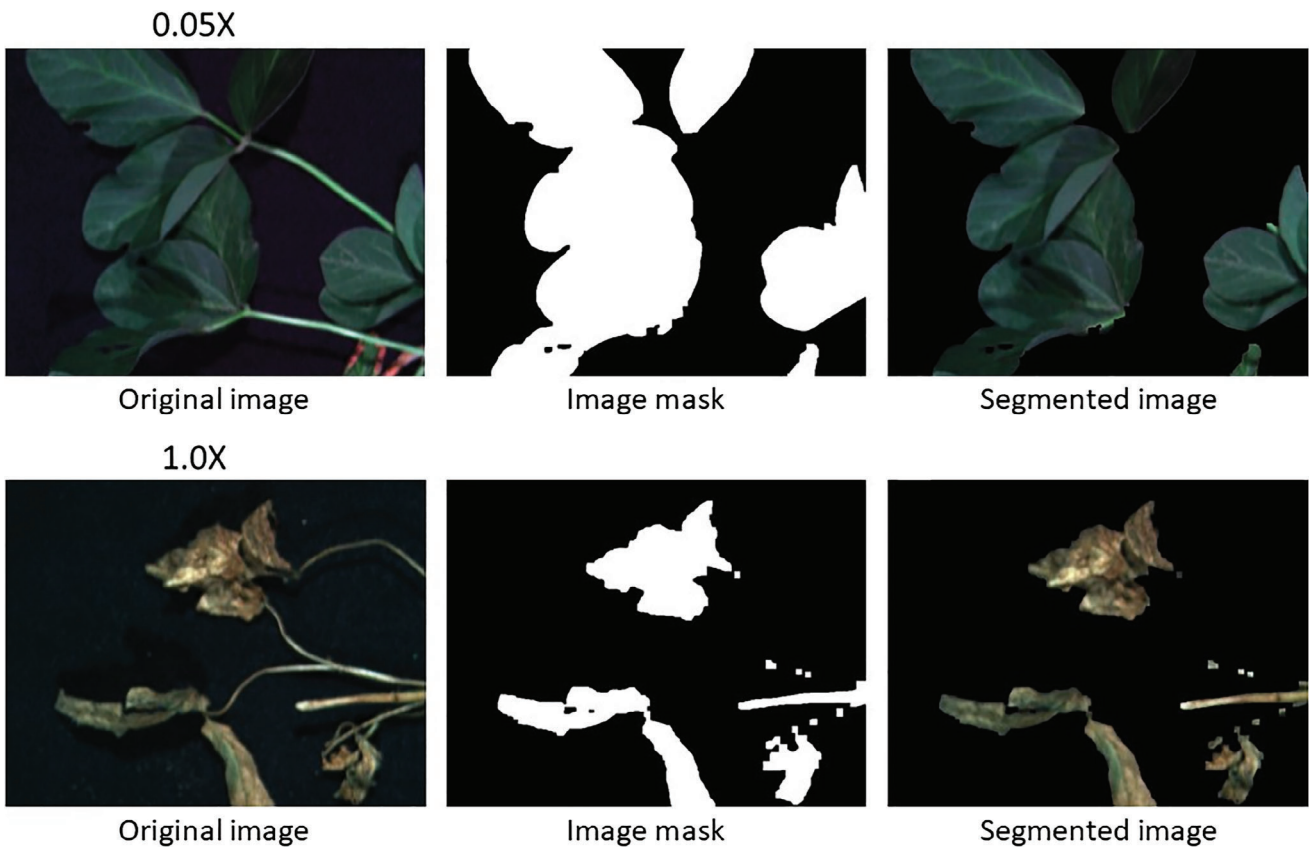

Figure 3. Hyperspectral image masking and segmentation for low-dose (0.05X) and high-dose (1.0X) images, respectively (selected images acquired at 3 WAT).

Table 1. Conventional spectral features used in this study

Spectral features	Definition	Description/formula	Literature
Derivative features			
D_b	Maximum value of first derivative within blue edge	Blue edge covers 490–530 nm D_b is a maximum value of first-order derivatives within the blue edge of 35 bands	Gong *et al.*[26]
λ_b	Wavelength at D_b	λ_b is wavelength position at D_b	Gong *et al.*[26]
SD_b	Sum of first derivative values within blue edge	Defined by sum of first-order derivative values of 35 bands within the blue edge	Gong *et al.*[26]
D_y	Maximum value of first derivative within yellow edge	Yellow edge covers 550–582 nm D_y is a maximum value of first-order derivatives within the yellow edge of 28 bands	Gong *et al.*[26]
λ_y	Wavelength at D_y	λ_y is wavelength position at D_y	Gong *et al.*[26]
SD_y	Sum of first derivative values within yellow edge	Defined by sum of first-order derivative values of 28 bands within the yellow edge	Gong *et al.*[26]
D_r	Maximum value of first derivative within red edge	Red edge covers 670–737 nm D_r is a maximum value of first-order derivatives within the red edge of 61 bands	Gong *et al.*[26]
λ_r	Wavelength at D_r	λ_r is wavelength position at D_r	Gong *et al.*[26]
SD_r	Sum of first derivative values within red edge	Defined by sum of first-order derivative values of 61 bands within the red edge	Gong *et al.*[26]
Continuous removal features			
$DEP_{550\text{-}750}$	The depth of the feature minimum relative to the hull	In the range of 550–750 nm	Pu *et al.*[27,28]
$WID_{550\text{-}750}$	The full wavelength width at half DEP (nm)	In the range of 550–750 nm	Pu *et al.*[27,28]
$AREA_{550\text{-}750}$	The area of the absorption feature that is the product of DEP and WID	In the range of 550–750 nm	Pu *et al.*[27,28]
Vegetation index			
GI	Greenness index	R_{554}/R_{677}	Zarco-Tejada *et al.*[29]
CARI	Chlorophyll absorption ratio index	$(\|(a\,670 + R_{670} + b)\|/(a^2+1)^{1/2})*(R_{700}/R_{670})$ $a = (R_{700}\text{-}R_{550})/150$, $b = R_{550} - (a*550)$	Kim *et al.*[30]
TCARI	Similar to OSAVI	$3[(R_{700} - R_{670}) - 0.2*(R_{700} - R_{550})]* R_{700}/ R_{670}$	Haboudane *et al.*[31]
MCARI	Modified chlorophyll (a and b) absorption in reflectance index (MCARI)	$[(R_{700} - R_{670}) - 0.2*(R_{700} - R_{550})]* R_{700}/ R_{670}$	Daughtry *et al.*[32]
NPCI	Normalized pigment chlorophyll ratio index	$(R_{680} - R_{430})/(R_{680} + R_{430})$	Peñuelas *et al.*[33]
PRI	Photochemical reflectance index	$(R_{531} - R_{570})/(R_{531} + R_{570})$	Gamon *et al.*[34]
RARS	Ratio analysis of reflectance spectra	R_{760}/R_{500}	Chappelle *et al.*[35]
ARI	Anthocyanin reflectance index	$ARI = (R_{550})^{-1} - (R_{700})^{-1}$	Gitelson *et al.*[36]
PSRI	Plant senescence reflectance index (PSRI)	$(R_{678} - R_{500})/R_{750}$	Merzlyak *et al.*[37]

sampled to measure biological responses: plant height (Ht), shoot wet weight (wet Wt) and dry weight (dry Wt). Soybean Ht was recorded on five randomly selected plants in each plot. Ten soybean plants from each plot were excised at the soil surface and the wet Wt values recorded. The plants were then oven dried at 60 °C for 72 h, after which dry Wt was recorded. Simultaneously with field biophysical measurements, five soybean plants from each plot were excised, stored in the cooler and transported to the laboratory for immediate hyperspectral imaging. When harvesting, the yield of each plot was recorded by the yield monitor mounted on the combine harvester.

The plants were imaged using a Resonon Pika II VNIR (Visible + Near Infrared) hyperspectral imaging system (Resonon, Bozeman, MT, USA). The Pika II camera was mounted approximately 1 m above the sample stage, which (when combined with the lens magnification) resulted in sub-millimeter pixel sizes (Fig. 2). The camera had 2.1 nm spectral resolution and 12-bit dynamic range. The spectral range of the camera was 400–900 nm, with 240 narrow wavelength bands evenly spaced in the range. The high spatial resolution of the imaging settings ensured that a very large number of pure plant pixels were present in the images. During imaging, the sample stage was covered by black felt cloth, which has almost no spectral reflection, to provide good contrast between the plants and the background in the acquired images. The good contrast also simplified segmentation of the plant pixels from the non-plant pixels in the images. This collected hyperspectral image data permitted the extraction of the plant portion within the field of view, which thus avoid background interference (i.e. soil, canopy gaps). To achieve this, mask layer of soybean plants was generated from the hyperspectral images (Fig. 3). An average spectrum for soybean plants was then derived for each hyperspectral image with this mask. The averaged spectra that were purely for plant information were used to analyze the impact

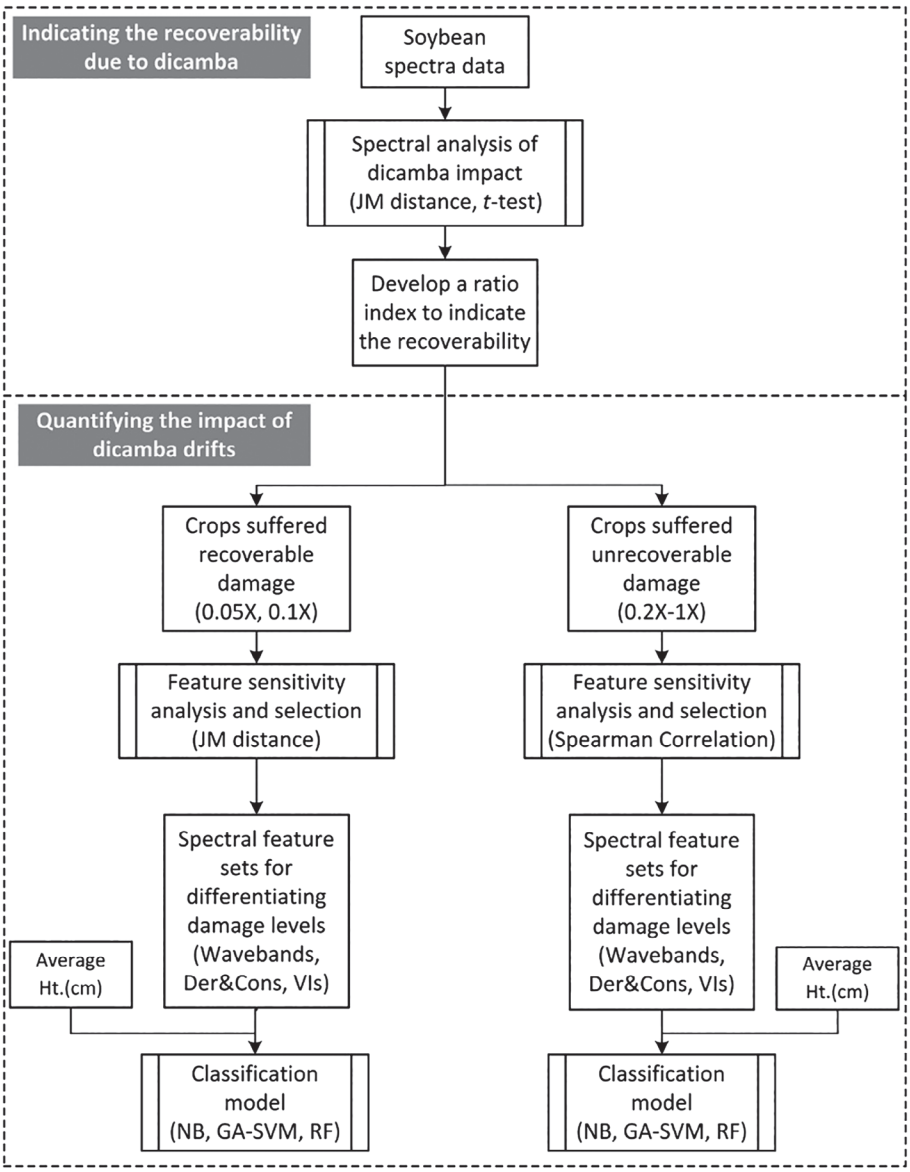

Figure 4. A framework of the data analysis in this study.

of dicamba treatments. Before plant measurement, the hyperspectral camera measured the dark current with the lens cap on and the white reference with a 0.3 × 0.3 m Spectrolon® reference target with 99% nominal reflectance (Labsphere, North Sutton, NH, USA). The dark current and white reference data were used to radiometrically calibrate the plant hyperspectral data to generate percent reflectance. The algorithm of hyperspectral image processing was programmed using Python, a widely used high-level programming language for general-purpose programming.

2.3 Data analysis

2.3.1 Spectral response analysis and features extraction

First, the spectral response to dicamba rates was analyzed. The spectra that were extracted from the hyperspectral images within one plot were averaged to represent the plot. The plot-level spectra were then further averaged in line with the dicamba rates to facilitate the comparison. In addition, to highlight the spectral difference between dicamba affected soybeans and herbicide-free soybeans, a spectral ratio analysis was performed that took the averaged spectra of the reference plots as the denominator. From the spectral pattern at 1 and 3 WAT, it was clear that the soybeans receiving a dicamba dosage ≥0.2X suffered an unrecoverable injury (for detail please see Section 3.1). Therefore, two analysis steps were designed to construct and identify effective spectral features for (i) differentiating the recoverable soybean plants from the unrecoverable ones and (ii) indicating dicamba rates under both recoverable and unrecoverable situations. In analytical modeling, 60% of the spectral and biophysical data at each dicamba rate was randomly selected as training data, whereas the other 40% of the data was used for validation.

2.3.1.1 Development of recoverability index. To develop an index for differentiating the recoverable or unrecoverable dicamba injured samples, a sensitivity analysis was performed on the reflectance spectra to identify the wavebands that had the greatest potential in indicating recoverability. To achieve this, a Jeffries–Matusita (JM) distance[25] was used to evaluate the spectral difference between samples receiving 0.05–0.1X

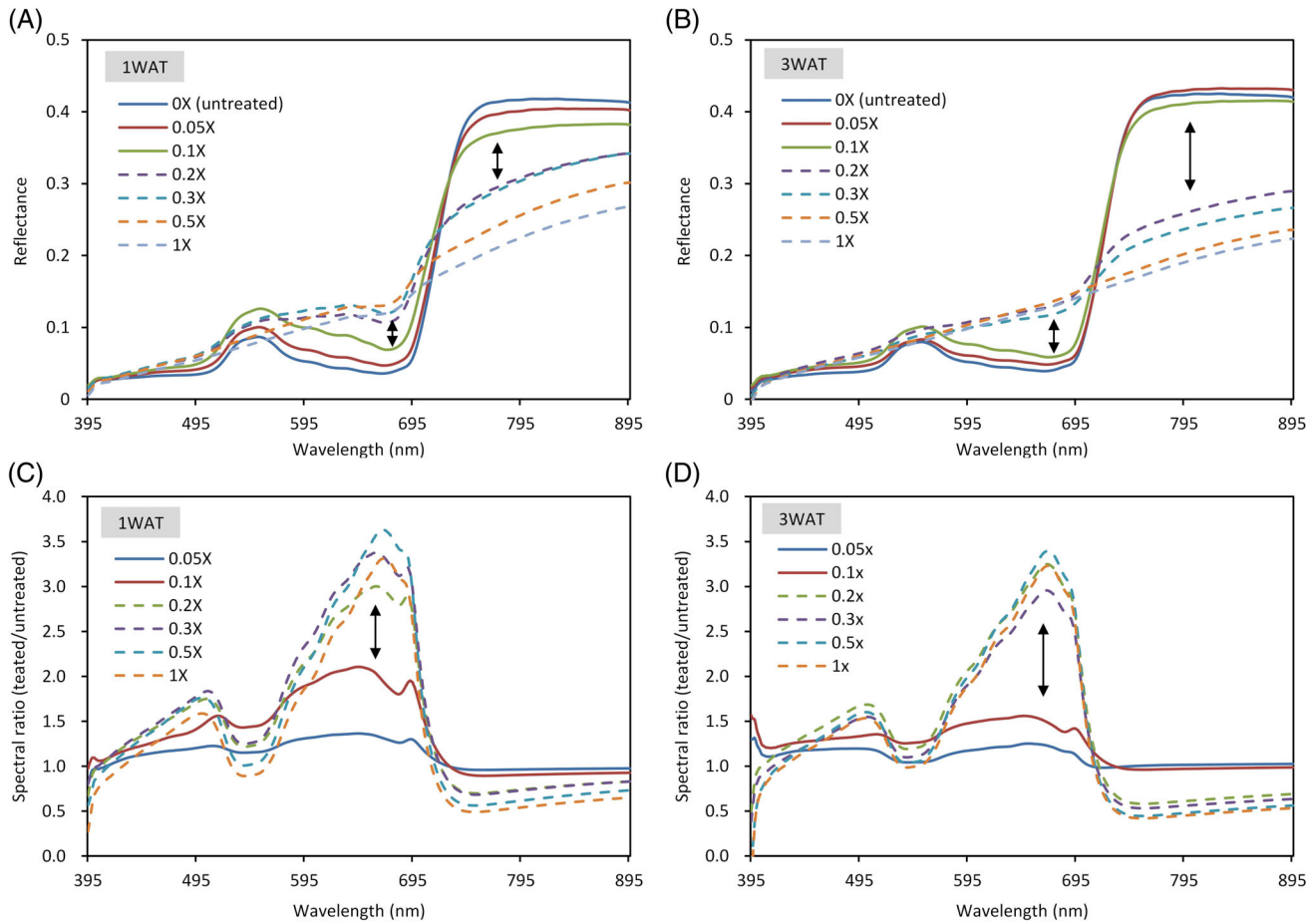

Figure 5. Averaged reflectance spectra and spectral ratio curves (treated/untreated) for different dicamba spray rates at 1 and 3 WAT (0.05 – 1.0X indicate dicamba spray rates).

dicamba dosage (recoverable) and 0.2 – 1.0X dicamba dosage (unrecoverable) for each waveband. The JM distance takes into account the distance between class means and the distribution of values from the means by involving the covariance matrices of the classes. It was applied as an efficient feature separability criterion in band selection for hyperspectral image analysis.[25] In addition, a Student's *t*-test was used to examine the statistical significance of the difference for each waveband. The wavebands corresponding to the peak values of the JM distance curves were identified as sensitive wavebands. Based on these sensitive wavebands, a recoverability index was developed by taking the ratio or normalization form as a basic structure. The index construction process is detailed in Section 3.2.

2.3.1.2 Extraction of sensitive spectral features indicating dicamba rates. For both the recoverable and unrecoverable situations, sensitive spectral features (SFs) were identified as being associated with the rates of dicamba, respectively. As only subtle spectral differences were expected between different dicamba rates, in addition to the original reflectance wavebands, a total of 21 conventional SFs that are potentially sensitive to plant biochemical and biophysical responses were included to form candidate feature sets (Table. 1). In these sets, 12 derivative and continuous removal features (Der & Con features) were able to capture plant absorption features around some classic wavelength ranges (i.e. blue edge, yellow edge and red edge) of plants,[26–28] and were able to be correlated with plant biophysical status and stresses.[1] In addition, nine classic narrow-band hyperspectral vegetation indices (VIs) were considered that can be associated with plant greenness,[29] pigments such as chlorophyll,[30–33] carotenoid,[34,35] anthocyanin[36] and plant stress.[37]

For the recoverable situation, a JM distance and Student's *t*-test exam were used to identify sensitive features in differentiating samples receiving 0.05X and 0.1X dicamba rates (called 0.05X samples and 0.1X samples hereafter for short). For the unrecoverable situation, given that the 0.2X, 0.3X, 0.5X, 1.0X samples exhibited increasing dicamba damage, an ordinal Spearman correlation analysis was used to identify sensitive features in differentiating these rates. For both recoverable and unrecoverable situations, a cross-correlation check was performed on the identified sensitive SFs to eliminate features that had high cross-correlation to other features, but relatively lower sensitivity, until there were no features with cross-correlation of $R^2 > 0.8$. Such a procedure guarantees that the spectral feature sets have relatively low levels of information redundancy.[38]

2.3.2 Differentiation of dicamba rates

To differentiate the recoverable and unrecoverable situations for soybean plants that received dicamba treatments, a thresholding method was applied to the developed novel index. A stepwise method was applied to determine the optimal threshold. One hundred evenly spaced intervals were set within the data range (i.e.

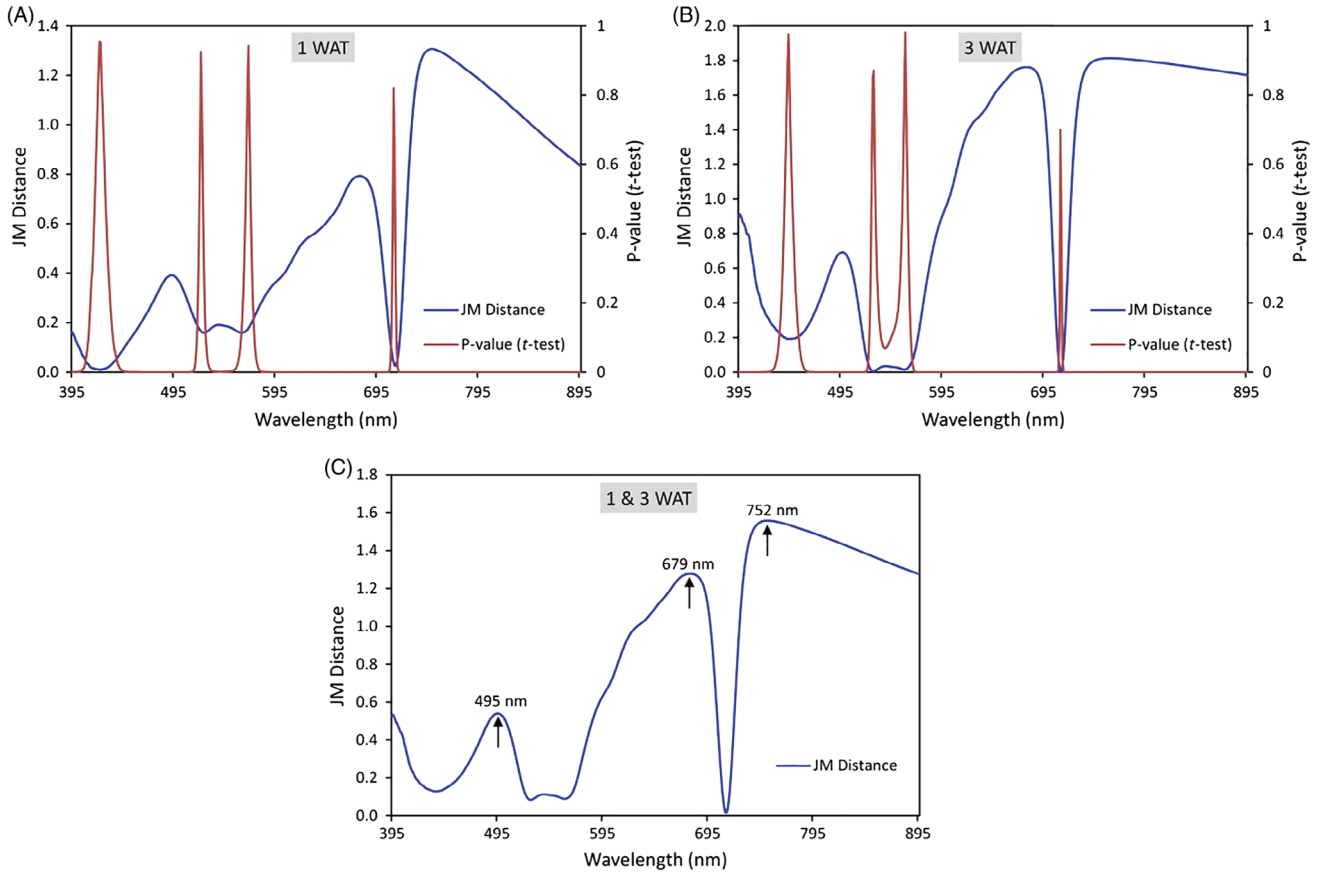

Figure 6. Sensitive analysis of wavebands to the recoverability of soybean plants due to dicamba treatment (JM distance and *t*-test were performed on the wavebands' spectral ratio between 0.05–0.1X samples and 0.2–1.0X samples, which represent the recoverable and unrecoverable situations of dicamba-treated soybean plants).

from minimum to maximum) of the index. Based on the training data, the overall accuracy was calculated by traversing all intervals. The cut-off value was defined as the point when the highest accuracy was reached.

Based on the identified sensitive spectral feature sets corresponding to both recoverable and unrecoverable situations, three machine learning classification algorithms, naive Bayes (NB), random forest (RF) and support vector machine coupled with a genetic algorithm (GA-SVM), that have different principles were used to differentiate the effects from dicamba rates. NB is a simple probabilistic classifier based on the Bayes' theorem. The RF classifier is an ensemble classifier that produces multiple decision trees using a randomly selected subset of training variables. The SVM classifier addresses the small-size training set problem and has a good generalization capacity. In GA-SVM, the GA was implemented to optimize parameters for the SVM. These three algorithms have shown excellent classification capabilities in remote sensing applications, which are therefore adopted in this study for comparison.[10,39,40] Apart from the SFs, given that plant height is a measurable biophysical parameter that also exhibited a response to the herbicide effect, it was also included to generate a synthetic model (SFs + plant Ht), which was further compared with the model based on solely SFs. The models were trained by the training samples and validated against the validation samples. The overall accuracy (OA) and Kappa coefficients were calculated from the confusion matrices to evaluate the classification accuracies.[41] All statistical analysis and modeling was conducted using MATLAB software (MathWorks Inc., Natick, MA, USA). Figure 4 illustrates the framework of the data analysis in this study.

3 RESULTS AND ANALYSIS

3.1 Spectral response and biophysical changes due to dicamba

In comparing the averaged spectra of soybean under different dicamba rates, an orderly variation pattern can be clearly observed from the results for both 1 and 3 WAT (Fig. 5(A,B)). With the increase in dicamba rate, several typical spectral traits of green vegetation, such as green peak, red valley and red edge, gradually become vague. Comparing with the untreated control samples, the spectral ratio curves on 1 and 3 WAT provide a clear spectral variation pattern, which exhibited two peaks at 480–510 nm and 640–690 nm, and a low platform beyond 730 nm (Fig. 5(C,D)). In addition, it can be seen that the 0.2X dicamba rate seems to be a turning point of the soybeans' fate, after which the plants receiving lower dosages would recover, whereas the plants receiving higher dosages (include 0.2X) would become feeble. The spectral difference between the 0.05–0.1X samples and the 0.2–1.0X samples can be observed at 1 WAT (Fig. 5(A)), and this becomes clearer at 3 WAT (Fig. 5(B)).

3.2 Differentiating recoverable and unrecoverable situations

The curves of JM distance and the *P* value of the *t*-test between 0.05–0.1X samples and 0.2–1.0X samples (Fig. 6(A,B)) show the

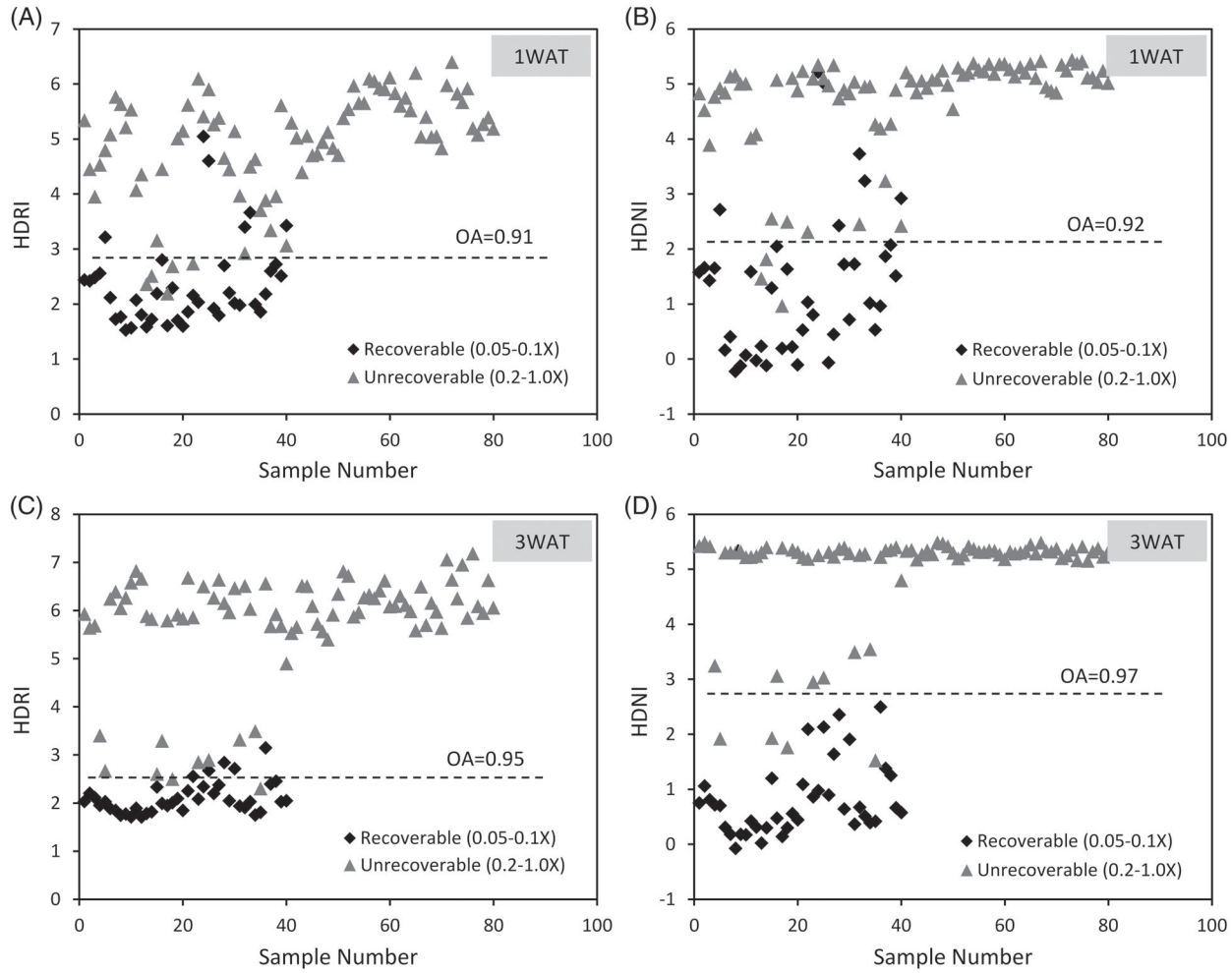

Figure 7. Differentiation between recoverable and unrecoverable samples using two newly developed indices, HDRI and HDNI. The dashed lines indicate the cut values that applied in the differentiation; 0.05–1.0X indicate dicamba rates.

sensitivity of wavebands to the recoverability of soybean plants from dicamba treatment. Fig. 6(C) shows the JM distance based on the pooled data from 1 and 3 WAT. From the data, the three best sensitive wavebands were determined at 495, 679 and 752 nm, which correspond the spectral peaks of the JM distance curve. These spectral values were used to construct a recoverability index. The results of the *t*-test for the three wavebands at 1 and 3 WAT also confirmed that the difference between 0.05–0.1X samples and 0.2–1.0X samples was statistically significant. As indicated in Fig. 6(C), the reflectance tended to significantly increase at 495 and 679 nm for dicamba-treated samples, whereas it decreased at 752 nm. Considering that adverse spectral variation can enhance the sensitivity of the recoverability index, two types of indices were constructed following the basic ratio and normalization structures. Instead of using the original band reflectance, the spectral ratio of a band (i.e. the band reflectance of a dicamba-treated sample divided by the averaged band reflectance of untreated reference samples) was used to create a new index to improve the generality of the index. Two recoverability indices, the Herbicide Damage Ratio Index (HDRI) and the Herbicide Damage Normalized Index (HDNI), were constructed:

$$\mathrm{HDRI} = a * \frac{RatioRef_{495}}{RatioRef_{752}} + b * \frac{RatioRef_{679}}{RatioRef_{752}}, a = 1.53, b = 0.095 \quad (1)$$

$$\mathrm{HDNI} = a * \frac{RatioRef_{679} - RatioRef_{752}}{RatioRef_{679} + RatioRef_{752}} - b * \frac{RatioRef_{495} - RatioRef_{752}}{RatioRef_{495} + RatioRef_{752}}, a = 2.15, b = 1.01 \quad (2)$$

Given the wavebands of 495 and 679 nm located at two separate peaks of the spectral ratio curves (Fig. 6(C)), the combination of the two components in both the HDRI and HDNI thus provides complementary information in capturing the spectral response caused by dicamba treatment. In this study, the coefficients *a* and *b* in HDRI and HDNI were determined through fitting by a Fisher linear discrimination analysis based on the training dataset. In this process, the two fraction terms in the HDRI and HDNI were treated as the independent variables (x_1 and x_2) whereas the recoverability was treated as the dependent variable (y) with recoverable and unrecoverable situations indicating 0 and 1, respectively. The fitted coefficients for x_1 and x_2 were determined as the coefficients *a* and *b* in HDRI and HDNI.

A stepwise thresholding method was applied on HDRI and HDNI to determine the optimal cut-off values for differentiating the recoverable and unrecoverable situations. As shown in Fig. 7, the cut-off values were 2.89 (1 WAT) and 2.58 (3 WAT) for HDRI and 2.15 (1 WAT) and 2.82 (3 WAT) for HDNI. According to the validation samples, the OA of determining the recoverable and

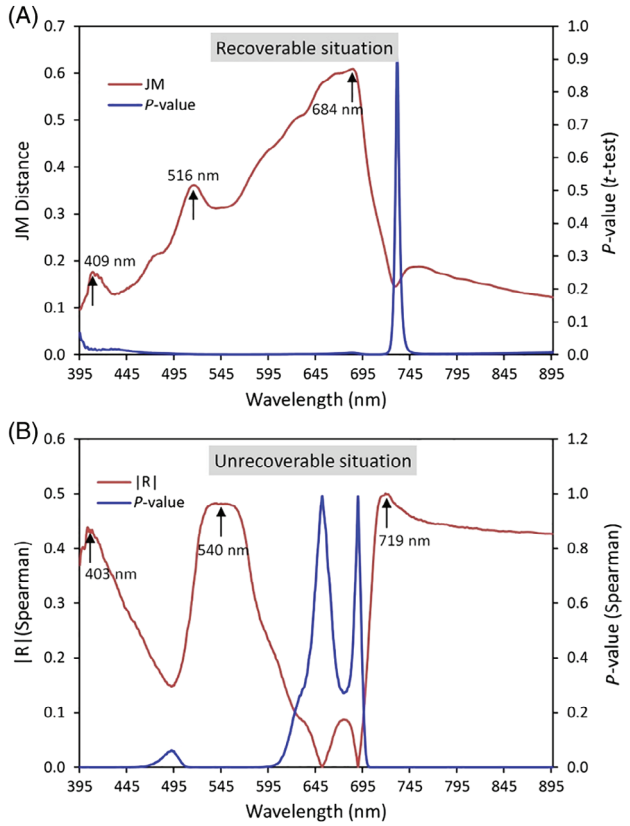

Figure 8. Sensitive analysis of wavebands to dicamba rates under recoverable and unrecoverable situations (for the recoverable situation, the JM distance and the *P* value of the *t*-test were used to indicate the sensitivity of wavebands to dicamba rates. For the unrecoverable situation, the Spearman correlation coefficient $|R|$ and the *P* value of the correlation analysis were used as the selection criteria. The analysis was performed based on the pooled data at 1 and 3 WAT.

unrecoverable samples reached 0.91 (1 WAT) and 0.95 (3 WAT) for HDRI and was slightly better at 0.92 (1 WAT) and 0.97 (3 WAT) for HDNI. It is therefore feasible to differentiate the recoverable soybean plants from the unrecoverable ones through HSI analysis.

3.3 Quantifying dicamba rates by spectral features

The sensitivity of spectral wavebands and other types of spectral features was evaluated to model dicamba rates under both recoverable and unrecoverable situations. For the recoverable situation, based on the pooled data from 1 and 3 WAT, the JM distance and *P* value of the *t*-test between 0.05X samples and 0.1X samples are shown in Fig. 8(A). From this, three sensitive wavebands at 409, 516, 684 nm can be identified at the peaks of the JM distance curve. The *P* values of these wavebands were less than 0.01, which ensures the differences are statistically significant. Moreover, the sensitivity analysis was also performed on Der & Con features and VIs. A cross-correlation check with a criteria of $R^2 < 0.8$ was implemented on those sensitive SFs, which retained a total of eight sensitive SFs as the input variables of the model: 409 and 684 nm, λ_b, λ_r, $WID_{550\text{-}750}$, GI, TCARI and PSRI.

For the unrecoverable situation, the absolute value of the Spearman correlation coefficient $|R|$ and the *P* value of the correlation analysis are shown in Fig. 8(B). Three sensitive wavebands at 403, 540 and 719 nm were identified at the peaks of the $|R|$ curve. The *P* values of these wavebands were less than 0.01, which ensures the validity of the waveband selection. Similarly, according to the sensitivity analysis and cross-correlation check, a total of seven sensitive SFs were retained: 403, 719 nm, SD_b, $DEP_{550\text{-}750}$, $WID_{550\text{-}750}$, PRI, MCARI.

Based on the selected SFs for both the recoverable and unrecoverable situations, the classification models were established using NB, GA-SVM and RF on 1 and 3 WAT. In addition to the models solely based on SFs, synthetic models that include both SFs and plant Ht were also established. Table 2 summarizes the OA and Kappa coefficient for the two models under different algorithms and feature combinations. It should be noted that the accuracies of the models for the recoverable situation were significantly greater than for the models for the unrecoverable situation. The models driven by both SFs and Ht (called the SFs + Ht model hereafter) exhibited obvious higher accuracies than the models driven by solely SFs (called the SFs model hereafter). For the three algorithms, NB or RF showed the highest accuracies under different variable combinations.

For models corresponding to the recoverable situation, the OA and Kappa coefficients of the SFs models ranged from 0.69 to 0.75 (OA) and 0.38 to 0.50 (Kappa) at 1 WAT, and from 0.56 to 0.69 (OA) and 0.13 to 0.38 (Kappa) at 3 WAT. The SFs + Ht models exhibited a significantly higher accuracy as OA and Kappa ranged from 0.88 to 0.94 (OA) and 0.75 to 0.88 (Kappa) at 1 WAT and achieved 0.94 (OA) and 0.88 (Kappa) for all three algorithms at 3 WAT. The RF outperformed the other two algorithms for both SFs models and SFs + Ht models, except for the SFs model at 1 WAT, which achieved the highest accuracy under the NB algorithm.

For the models responsible for the unrecoverable situation, the highest OA and Kappa values decreased to 0.53 (OA) and 0.38 (Kappa) for SFs models at 1 WAT and to only 0.38 (OA) and 0.17 (Kappa) for SFs models at 3 WAT accuracies. The SFs + Ht models yielded relatively higher accuracies, with the highest OA and Kappa values reaching 0.63 (OA) and 0.46 (Kappa) at 1 WAT and 0.69 (OA) and 0.58 (Kappa) at 3 WAT. Within these models, NB performed best for the SFs models whereas RF performed best for the SFs + Ht models.

4 DISCUSSION AND CONCLUSIONS

From the spectral variation of the soybean plants receiving dicamba treatments at 1 and 3 WAT, it is interesting to find that the plants receiving less than 0.2X dicamba will have a different fate from those receiving more than 0.2X (including 0.2X), resulting in recoverable and unrecoverable situations for the dicamba-treated plants. It is encouraging that the plants' spectra of the two situations started to show a clear difference at 1 WAT (Fig. 5(A)), which thus permitted the prediction of recoverability at an early stage. Figure 9 demonstrates the relationships between the dicamba rates and corresponding yields, as well as between dicamba rates and several biophysical parameters. It can be seen that the yield loss generally conforms a logarithmic function with the dicamba dosage, which shows a sharp increase at the beginning and then levels off (i.e. almost total crop failure) at 0.2X rate. In addition, despite the soybean plants receiving 0.05X and 0.1X rates recovered at 3 WAT, the yield loss for the 0.05X samples still surpassed 50% of the maximal yield, indicating that the misuse of dicamba on those non-DT soybean plants will cause severe damage and significant yield loss. The variation of the plant Ht, wet Wt and dry Wt over different dicamba rates showed a generally consistent pattern similar to the relationship between yield and rates, with these biophysical parameters declining sharply with rates over 0.05–0.1X. Therefore, for future research, further

Table 2. Classification accuracies of models determining different dicamba rates

		Recoverable (0.05–0.1X)				Unrecoverable (0.2–1.0X)			
		1 WAT		3 WAT		1 WAT		3 WAT	
Algorithms	Variables	OA	Kappa	OA	Kappa	OA	Kappa	OA	Kappa
NB	SFs	0.75	0.50	0.63	0.25	0.53	0.38	0.38	0.17
	SFs + Ht	0.88	0.75	0.94	0.88	0.53	0.38	0.44	0.25
GA-SVM	SFs	0.69	0.38	0.56	0.13	0.31	0.08	0.34	0.13
	SFs + Ht	0.88	0.75	0.94	0.88	0.59	0.46	0.66	0.50
RF	SFs	0.69	0.38	0.69	0.38	0.50	0.33	0.34	0.13
	SFs + Ht	0.94	0.88	0.94	0.88	0.63	0.46	0.69	0.58

All the accuracies were derived based on the validation datasets. The underlined cells indicate the highest accuracy achieved in certain algorithm and feature combination. NB, naive Bayes; GA-SVM, support vector machine coupling with genetic algorithm; RF, random forest; SFs, selected spectral features mentioned in Section 3.3; Ht, plant height of soybean.

investigation and comparison of the impact of dicamba on plants at lower dosage levels between 0.0X and 0.1X, such as 0.01X, 0.02X, 0.03X, 0.04X, 0.05X, 0.06X, 0.07X, 0.08X, 0.09X and 0.1X, is recommended.

The strong and clear orderly pattern of the spectral response due to dicamba treatments (Fig. 5(A,B)) suggested the possibility of spectrally quantifying different dicamba rates. From the spectral ratio curves (Fig. 5(C,D)), the two peaks around the green and red bands indicate less absorption of pigments like chlorophyll, carotenoid and anthocyanin. As confirmed by Silva *et al.*[42] and Huang *et al.*,[22] herbicides like glyphosate and dicamba can lead to severe destruction of the pigment systems associated with photosynthesis, and affect the dry matter accumulation, resulting in corresponding changes in biophysical parameters (Fig. 9). The decline of the reflectance in the NIR region of the dicamba-affected plants reflects the impact of the herbicide on the foliar cellular structure of the plants.[40] Such a spectral response for dicamba was also in good agreement with Huang *et al.*[22]

The clear separation of the recoverability of dicamba-treated plants (Fig. 5) permits spectral differentiation between the recoverable and unrecoverable situations (i.e. 0.05–0.1X vs 0.2–1.0X). These spectral differences are induced by changes in biochemical status and biophysical parameters due to dicamba treatments. In addition, the hyperspectral imaging technique enables purification of the spectra (i.e. the non-plant portion was eliminated). The high-quality spectral data facilitated the spectral determination of the recoverability of the dicamba-treated plants at an early stage (1 WAT). Therefore, based on only three wavebands at 495, 679 and 752 nm, both the newly developed indices HDRI and HDNI were able to yield a satisfactory classification accuracy, with OA over 0.9 at 1 and 3 WAT. The two response wavebands (495 and 679 nm) are located at the green and red bands, which are also close to the two peaks of the spectral ratio curves (Fig. 5(C,D)). The ratio and normalization of the band combination enhanced the spectral difference between recoverable and unrecoverable samples, which facilitated their separation. The early determination of the recoverability of the dicamba-treated plants can help producers and herbicide applicators to prevent the plants from further injury, which is of practical significance.

Given that the spectral variation pattern differs significantly between the recoverable and unrecoverable situations, two separate models for quantifying the dicamba rates were established. Despite the compositions of the SFs corresponding to the recoverable and unrecoverable situations being different, both selected SF sets include original wavebands, Der & Con features and VIs. It should be noted that features of blue edges (i.e. 403 nm, 409 nm, λ_b, SD_b) and red edges (i.e. 684 nm, λ_r, $DEP_{550\text{-}750}$, $WID_{550\text{-}750}$) regions were included in both SF sets. These SFs are associated with plant stress. Some VIs, such as PRI TCARI and MCARI, were sensitive to the destruction of chlorophyll and carotenoid pigment systems. The PSRI that was sensitive to the Car/Chl ratio was associated with plant senescence. The GI was sensitive to biochemical constituents and leaf area index (LAI). Therefore, all the VIs mentioned above deliver information about the variation in the biochemical or biophysical status of the dicamba-treated plants. It should also be noted that the addition of Ht significantly enhanced the detecting accuracy of the dicamba dosage levels. This might be because the plant height, plant greenness, and plant biochemical and biophysical status of as reflected by spectral features are mutual complementary information that can reflect the influence of dicamba treatment on plants.

Among the three modeling algorithms, the superior performance of RF might be related to its ability to handle high data dimensionality and multicollinearity.[10] The high classification accuracy for the recoverable situation (maximal OA > 0.9) suggested the possibility of the spectral differentiation of the dicamba treatment levels. However, the relatively low classification accuracy (maximal OA = 0.63 at 1 WAT, maximal OA = 0.69 at 3 WAT) of the unrecoverable situation might be because of the saturation effect of the biophysical changes of the severely damaged samples. This effect can be confirmed by the leveling-off pattern of the biophysical parameters within 0.2–1.0X (Fig. 9). It is is more important from a practical perspective that knowing the dicamba treatment levels under the recoverable situation. Some remedial measures, such as water drip washing or additional fertilization, can be implemented at an early stage.

This research concludes that (i) the spectral response of soybean injury caused by dicamba sprays can be clearly captured by HSI, (ii) the recoverable and unrecoverable damage situations due to the herbicide can be accurately differentiated by the recoverability spectral indices HDRI and HDNI developed, and (iii) using optimal spectral feature sets, the dicamba spray rates under recoverable and unrecoverable situations can be determined. The ability of HIS to demonstrate the difference between the recoverable and unrecoverable situations of dicamba-treated soybean

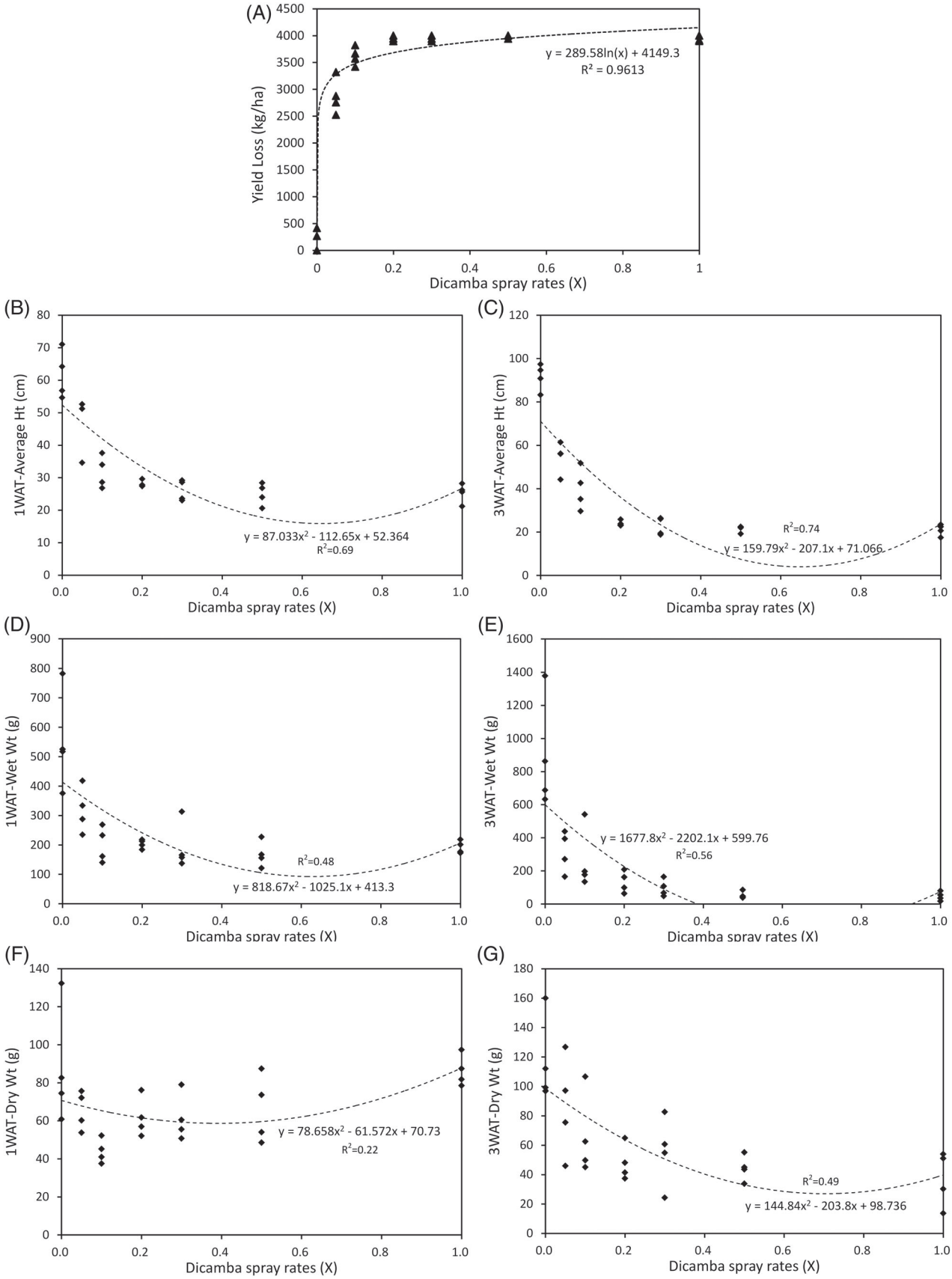

Figure 9. The relationship between dicamba spray rates, soybean yield and biophysical parameters.

plants as well as characterizing the herbicide rates revealed a potential for monitoring and assessing herbicide crop damage via unmanned aerial vehicle, agricultural airplane or high-resolution satellite. More experiments and research on these applications are expected in future.

ACKNOWLEDGEMENTS

The research presented in this paper was supported by USDA-ARS National Program 305 Crop Production and National Natural Science Foundation of China (Project No. 41671415). The authors thank David Thornton for his field work and dicamba spray, Efren Ford and Earl Gordon for field sampling and laboratory measurement, and Howard Brand and Ryan Poe for hyperspectral image acquisition and processing. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

REFERENCES

1 Pu R, *Hyperspectral Remote Sensing: Fundamentals and Practices*. CRC Press, Boca Raton, FL (2017).
2 Rathod PH, Rossiter DG, Noomen MF and Meer FDVD, Proximal spectral sensing to monitor phytoremediation of metal-contaminated soils. *Int J Phytoremediat* **15**:405–426 (2013).
3 Sytar O, Brestic M, Zivcak M, Olsovska K, Kovar M, Shao H *et al.*, Applying hyperspectral imaging to explore natural plant diversity towards improving salt stress tolerance. *Sci Total Environ* **578**:90–99 (2017).
4 Jan B, Jörg S and Lutz P, Detection of early plant stress responses in hyperspectral images. *ISPRS J Photogramm Remote Sens* **93**:98–111 (2014).
5 Li L, Zhang Q and Huang D, A review of imaging techniques for plant phenotyping. *Sensors-Basel* **14**:20078–20111 (2014).
6 Lowe A, Harrison N and French AP, Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* **13**:80 (2017).
7 Stegmayer G, Milone DH, Garran S and Burdyn L, Automatic recognition of quarantine citrus diseases. *Expert Syst Appl Int J* **40**:3512–3517 (2013).
8 Jingcheng Z, Ruiliang P, Lin Y, Jihua W, Wenjiang H and Guijun Y, Monitoring powdery mildew of winter wheat by using moderate resolution multi-temporal satellite imagery. *Plos One* **9**:e93107 (2014).
9 Potamitis I, Ganchev T and Fakotakis N. Automatic acoustic identification of crickets and cicadas., *International Symposium on Signal Processing & Its Applications*, 2007, pp. 1–4.
10 Belgiu M and Drăguţ L, Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* **114**:24–31 (2016).
11 Calderón MR, Navas-Cortés JA and Zarco-Tejada PJ, Early detection and quantification of verticillium wilt in olive using hyperspectral and thermal imagery over large areas. *Remote Sens-Basel* **7**:5584–5610 (2015).
12 Lin Y, Zhang J, Shi Y, Nie C, Wei L and Wang J, Damage mapping of powdery mildew in winter wheat with high-resolution satellite image. *Remote Sens-Basel* **6**:3611–3623 (2014).
13 Behmann J, Mahlein AK, Rumpf T, Römer C and Plümer L, A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precis Agric* **16**:239–260 (2015).
14 Huang Y, Lan Y, Thomson SJ, Fang A, Hoffmann WC and Lacey RE, Development of soft computing and applications in agricultural and biological engineering. *Comput Electron Agric* **71**:107–127 (2010).
15 Singh A, Ganapathysubramanian B, Singh AK and Sarkar S, Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci* **21**:110–124 (2016).
16 Henry WB, Shaw DR, Reddy KR, Bruce LM and Tamhankar HD, Remote sensing to detect herbicide drift on crops. *Weed Technol* **18**:358–368 (2004).
17 Yao H, Huang Y, Hruska Z, Thomson SJ and Reddy KN, Using vegetation index and modified derivative for early detection of soybean plant injury from glyphosate. *Comput Electron Agric* **89**:145–157 (2012).
18 Feng Z, Guo Y, Huang Y, Reddy KN, Lee MA, Fletcher RS *et al.*, Early detection of crop injury from herbicide glyphosate by leaf biochemical parameter inversion. *Int J Appl Earth Obs Geoinf* **31**:78–85 (2014).
19 Huang Y, Thomson SJ, Ortiz BV, Reddy KN, Ding W, Zablotowicz RM *et al.*, Airborne remote sensing assessment of the damage to cotton caused by spray drift from aerially applied glyphosate through spray deposition measurements. *Biosyst Eng* **107**:212–220 (2010).
20 Ortiz BV, Thomson SJ, Huang Y, Reddy KN and Ding W, Determination of differences in crop injury from aerial application of glyphosate using vegetation indices. *Comput Electron Agric* **77**:204–213 (2011).
21 Yanbo H, Reddy KN, Thomson SJ and Haibo Y, Assessment of soybean injury from glyphosate using airborne multispectral remote sensing. *Pest Manag Sci* **71**:545–552 (2015).
22 Huang Y, Lin Y, Reddy KN and Zhang J, In-situ plant hyperspectral sensing for early detection of soybean injury from dicamba. *Biosyst Eng* **149**:51–59 (2016).
23 Johnson VA, Fisher LR, Jordan DL, Edmisten KE, Stewart AM and York AC, Cotton, peanut, and soybean response to sublethal rates of dicamba, glufosinate, and 2,4-d. *Weed Technol* **26**:195–206 (2012).
24 Collins CA and Seeney FM, *Statistical Experiment Design and Interpretation: An Introduction with Agricultural Examples*. Wiley, Chichester (1998).
25 Ullah S, Groen TA, Schlerf M, Skidmore AK, Nieuwenhuis W and Vaiphasa C, Using a genetic algorithm as an optimal band selector in the mid and thermal infrared (2.5–14 μm) to discriminate vegetation species. *Sensors-Basel* **12**:8755–8769 (2012).
26 Gong P, Pu R and Heald RC, Analysis of in situ hyperspectral data for nutrient estimation of giant sequoia. *Int J Remote Sens* **23**:1827–1850 (2002).
27 Pu R, Foschi L and Gong P, Spectral feature analysis for assessment of water status and health level in coast live oak (*Quercus agrifolia*) leaves. *Int J Remote Sens* **25**:4267–4286 (2004).
28 Pu R, Ge S, Kelly NM and Gong P, Spectral absorption features as indicators of water status in coast live oak (*Quercus agrifolia*) leaves. *Int J Remote Sens* **24**:1799–1810 (2003).
29 Zarco-Tejada PJ, Berjón A, López-Lozano R, Miller JR, Martín P, Cachorro V *et al.*, Assessing vineyard condition with hyperspectral indices: leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens Environ* **99**:271–287 (2005).
30 Kim MS, Daughtry CST, Chappelle EW and McMurtrey JE, The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (a par). In: *Proceedings of the 6th International Symposium on Physical Measurements and Signatures in Remote Sensing*, France: Val d'Isere (1994).
31 Haboudane D, Miller JR, Tremblay N, Zarco-Tejada PJ and Dextraze L, Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens Environ* **81**:416–426 (2002).
32 Daughtry CST, Walthall CL, Kim MS, Colstoun EBD and Iii MM, Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens Environ* **74**:229–239 (2000).
33 Peñuelas J, Gamon JA, Fredeen AL, Merino J and Field CB, Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. *Remote Sens Environ* **48**:135–146 (1994).
34 Gamon JA, Peñuelas J and Field CB, A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens Environ* **41**:35–44 (1992).
35 Chappelle EW, Kim MS and Iii MM, Ratio analysis of reflectance spectra (rars): an algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. *Remote Sens Environ* **39**:239–247 (1992).
36 Gitelson AA, Merzlyak MN and Chivkunova OB, Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochem Photobiol* **74**:38–45 (2001).
37 Merzlyak MN, Gitelson AA, Chivkunova OB and Rakitin VY, Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol Plantarum* **106**:135–141 (1999).
38 Wang S and Chang CI, Linearly constrained band selection for hyperspectral imagery. *IEEE Trans Geosci Remote Sens* **44**:1575–1585 (2006).

39 Yang J, Ye Z, Xu Z, Wei L and Jin H, Attribute weighted naive bayes for remote sensing image classification based on cuckoo search algorithm., *International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)* (2017).
40 Chi M, Rui F and Bruzzone L, Classification of hyperspectral remote-sensing data with primal svm for small-sized training dataset problem. *Adv Space Res* **41**:1793–1799 (2008).
41 Congalton R and Mead RA, A quantitative method to test for consistency and correctness in photointerpretation. *PE&RS* **49**:69–74 (1983).
42 Silva FB, Costa AC, Alves RRP and Megguer CA, Chlorophyll fluorescence as an indicator of cellular damage by glyphosate herbicide in *raphanus sativus* l. plants. *Am J Plant Sci* **5**:11 (2014).

