2,4 D and Dicamba Resistant Soybeans: Stewardship and Testing
Project No. 18-2019
2019 Final Report

PI: Ashli Brown, Associate Professor/MS State Chemist
abrown@bch.msstate.edu
310 Presidents Circle
Mississippi State, MS 39762
(662) 325 7806

Co-PIs: Darrell Sparks, Assistant Professor
dsparks@bch.msstate.edu
310 Presidents Circle
Mississippi State, MS 39762
(662) 325 7814

Daniel B. Reynolds, Professor of Weed Science
dreynolds@pss.msstate.edu
32 Creelman Street
Mississippi State, MS 39762
(662) 325 2311

Background and Objectives

The Mississippi State Chemical Laboratory (MSCL) routinely analyzes drift complaint samples in the spring. Most of these complaints consist of injured ornamentals or soybeans exposed to the following herbicides: 2,4-D, atrazine, acetochlor, dicamba, glyphosate, and paraquat. The lab currently uses sensitive liquid chromatographic techniques including LC-MS/MS to identify these compounds at residue levels. However, this sensitive method cannot differentiate between the acid, amine, ester, or choline formulations of 2,4-D and dicamba. Therefore, it is imperative new analytical methods are developed to ensure an effective stewardship program. We have begun to investigate the use of Fourier transform infrared spectroscopy (FT-IR) and preliminary data looks promising. Research was conducted in 2017, 2018, and 2019 in Starkville, MS using chemometrics coupled to FT-IR to produce classification models capable of identifying specific 2,4-D and Dicamba formulations present in damaged crop tissue. Our project had two main objectives:

1. Develop and validate analytical testing methods using FT-IR technology to differentiate 2,4-D and dicamba herbicides formulations.

2. Work with the Bureau of Plant Industry and MSU Extension agents to participate in a grower's educational program and design an off-target field sampling program for best practices and fundamental integrated pest management.

Report of Progress/Activity

Application of FTIR spectroscopy and chemometrics for the classification of 2,4-D formulations in damaged cotton and soybean tissue

Increased use of 2,4-D in row crop production may lead to increased cases of damage to susceptible cotton and soybeans following off-target movement (OTM) of 2,4-D. Research was conducted in 2017 and 2018 in Starkville, MS to develop a method using chemometrics and spectroscopy to produce classification models capable of identifying specific 2,4-D formulations present in damaged crop tissue. 2,4-D acid (ACID), dimethylamine salt (DMA), choline salt (CHOLINE), and isooctyl ester (ESTER) were applied to susceptible cotton and soybeans at 33, 17, 8, 4, 2, and 1 g 2,4-D ae ha^{-1}, and samples were analyzed via infrared spectroscopy to generate spectra which were then analyzed by principal component analysis (PCA) and linear discriminant analysis (LDA). Joint PCA-LDA models were only capable of classifying 2,4-D formulation in damaged tissue with up to 36% accuracy, whereas LDA alone produced models with 77 to 80% accuracy. Models performed worst when classifying 2,4-D DMA or ESTER and best when classifying 2,4-D CHOLINE or ACID. Model accuracies were similar regardless of sample media (soybean or cotton tissue) or data format (raw spectral data vs normalized, derived, and smoothed spectra). This research suggests that with further refining, chemometric analysis of spectral data from damaged crop tissue may be an economical, efficient, and promising application to support management of crop injury following OTM of 2,4-D.

Results and Discussion

Raw and Transformed Data Matrices. Baseline- and ATR- corrected spectra (raw spectra) from cotton and soybean samples treated with the different 2,4-D formulations are shown pooled over 2,4-D concentration and sampling timing in Figures 3 and 4, respectively. Significant peaks occurred at 3800 to 3000 cm^{-1} and 1800 to 800 cm^{-1}. The broad peak at 3800 to 3000 cm^{-1} is due to the O-H bend in water found in plant tissue and was ignored in further analysis. The spectral region commonly referred to as the 'fingerprint region' between 1800 to 800 cm^{-1} was included in spectral analysis. Raw cotton and soybean fingerprint spectra from tissue treated with various 2,4-D formulations are shown pooled over concentration and sampling timing in Figures 5 and 6, respectively. Increased resolution of spectral features became observable by narrowing the spectral focus in these Figures. Normalized, derived, and smoothed cotton and soybean fingerprint spectra are shown pooled over concentration and sampling timing in amplified differences in spectral features between samples. In a preliminary analysis of similar data, Reid (2017) used PCA loading plots to identify the most important spectral features in the soybean analyses at 1687 and 1560 cm^{-1}, which most likely represent the aromatic ring of 2,4-D and the primary or secondary amine from its various formulated salts. Similarly, PCA loading plot examination determined peaks between 1633 and 1556 cm^{-1} and 1395 to 1350 cm^{-1} are important for soybean sample classification. The peaks between 1633 and 1556 cm^{-1} likely represent the aromatic ring of the 2,4-D molecule and the primary or secondary amines from the formulated salts, and the features between 1395 and 1350 cm^{-1} indicate the carboxylic acid group present in 2,4-D formulations (Reid 2017). These spectral features provided the basis for determining a spectral range for use in subsequent PCA and LDA.

PCA, LDA, and Joint PCA-LDA on Raw Data

Cotton. PCA performed on the raw data pooled across concentrations and evaluation timings resulted in PC1 and PC2 accounting for 89 and 7% of the explained variation, respectively, and 100% total explained variation contained in the first 4 PC (Table 2). A 3D PCA scores plot of the first three PC demonstrates little clustering by 2,4-D formulation, despite the high amount of variation contained in the first three PC. LDA of the raw data pooled across concentrations and evaluation timings and using the eigenvectors generated by dimensional reduction via PCA produced a classification model with 33.38% accuracy (Table 3). The discrimination plot for this model is shown in Figure 10, where there is some linearization

of samples by formulation. This linear pattern of the discrimination plot is in contrast to the distinct clustering by formulation reported by Reid (2017) and other previous research in different media (Deng et al. 2016; Lee et al. 2009; Lehmann et al. 2015). Lack of distinct clustering is likely due to the nature of these models as classifying formulation across multiple concentrations and evaluation timings. The corresponding confusion matrix displaying the PCA-LDA model's prediction of 2,4-D formulation from a given sample of crop tissue plotted against the actual value is shown in Table 4. The classification model performed best identifying 2,4-D CHOLINE (44% accuracy), and worst identifying 2,4-D DMA (18% accuracy). LDA conducted alone (without PCA) on the raw cotton spectra produced a classification model with 77.16% accuracy, a significant improvement over the joint PCA-LDA model (Table 3). The discrimination plot of this model there is noticeable linear clustering of each formulation. The level of accuracy produced by this model is more consistent with previous research, although the clustering pattern remains irregular (Deng et al. 2016; Lee et al. 2009; Lehmann et al. 2015; Reid 2017). The corresponding confusion matrix following LDA alone is shown in Table 6 which demonstrates the most accuracy when identifying 2,4-D CHOLINE (89%) and least accurate when identifying 2,4-D DMA (71%) although even the poorest accuracy of this model (71%, Table 6) was a significant improvement over the joint PCA-LDA model and is closer to the accuracy reported by Reid (2017).

Soybeans. PCA performed on the raw soybean spectral data across concentrations and evaluation timings generated five PCs accounting for 69, 23, 5, 1, and 1% variation in PCs 1, 2, 3, 4, and 5, respectively, for a total of 99% variation contained in the first 5 PCs (Table 2). Despite a high proportion of variation explained by PC1 and PC2, the 3D score plot again reflects poor clustering by formulation. Joint PCA-LDA of resulted in a classification model with 36.26% accuracy (Table 3) and the discrimination plot for this model. Linearization of samples by variation is present but overall clustering remains poor relative to previous work in other sample media (Deng et al. 2016; Lee et al. 2009). The corresponding confusion matrix from this model is shown in Table 7. This model was most accurate (44%, 45% accuracy) when classifying tissue containing 2,4-D CHOLINE and ACID (respectively), and least accurate (16% accuracy) when classifying tissue containing 2,4-D ESTER (Table 7), which is in contrast to the cotton models that were most accurate classifying 2,4-D CHOLINE and least accurate with 2,4-D DMA. LDA conducted alone on the raw soybean spectral data produced a classification model with 79.84% accuracy (Table 3). Figure 18 shows a discrimination plot from this model with distinct linear clustering of each formulation. The accuracy of this model was more similar to Reid (2017) although the clustering pattern was linear as opposed to the bunched patterns reported by previous research (Reid 2017; Lehmann et al. 2015), likely due to analysis over a wide range of variable levels (concentrations, sample timings), as opposed to the fixed levels found in most previous research. This LDA's corresponding confusion matrix is shown in Table 9. Up to 80 to 81% accuracy was achieved by this model when classifying 2,4-D CHOLINE and 2,4-D DMA, respectively, and the poorest accuracy was still reasonable (66%) when classifying 2,4-D ESTER (Table 9).

PCA and PCA-LDA on Transformed Data

Cotton. PCA of cotton spectra normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and smoothed with Savitzky-Golay smoothing produced a 3D PC score plot, which does not reflect distinct clustering by formulation despite the first two PCs accounting for 71 and 12% variation, respectively. Joint PCA-LDA on transformed cotton spectra constructed a classification model with only 34.8% accuracy (Table 3), which is reflected by the lack of clustering by formulation in the LDA discrimination plot. In contrast, Reid (2017) produced discrimination plots with a high degree of clustering, but only utilized one concentration fixed over sampling timing. The corresponding confusion matrix for the classification model produced by PCA-LDA on transformed cotton spectra is presented in Table 5. This model was able to achieve up to 45% accuracy when classifying samples treated with 2,4-D CHOLINE and 29% accuracy when classifying samples treated with 2,4-D DMA, the same trend in classification performance observed in models from analysis of non-transformed (raw) cotton spectra.

Soybeans. A 3D PC scores plot of transformed soybean spectra is shown in Figure 16 which depicts little distinct clustering by formulation, despite 83% of the total explained variation being contained in PC1 (Table 2). Joint PCA-LDA produced a classification model with 32.42% overall accuracy (Table 3). The discrimination plot for this model depicts linearization of formulations with little sandwiching or clustering. Similarly, the confusion matrix constructed by joint PCA-LDA reflects poor accuracy across individual 2,4-D formulations (Table 8). Accuracy of this model ranged from 15% accuracy in classifying 2,4-D ESTER to 40% accuracy in classifying 2,4-D ACID (Table 8). These model accuracies, albeit poor, are largely similar to those produced by analysis of raw spectral data and are markedly less than those reported by Reid (2017).

Table 2. Variation explained by each PC following PCA of fingerprint (1800 to 800 cm^{-1}) spectra from cotton or soybean tissue treated 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a]

Data Matrix	Data Type[b]	PC1	PC2	PC3	PC4	PC5
		----------%----------				
Cotton Spectra	Raw	89	7	2	2	-
Cotton Spectra	Transformed	71	12	3	2	2
Soybean Spectra	Raw	69	23	5	1	1
Soybean Spectra	Transformed	83	4	3	2	1

[a]Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester; PC, principal component; PCA, principal component analysis

[b]Raw spectral data were not normalized, derived or smoothed; Transformed spectral data were normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and Savitzky-Golay smoothed

Table 3. Classification model parameters following LDA alone or joint with PCA of fingerprint spectra (1800 to 800 cm^{-1}) from cotton or soybean tissue treated 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a]

Data Matrix	Data Type[b]	Model Source	Accuracy
			%
Cotton Spectra	Raw	PCA-LDA	33.38
Cotton Spectra	Raw	LDA	77.16
Cotton Spectra	Transformed	PCA-LDA	34.8
Soybean Spectra	Raw	PCA-LDA	36.26
Soybean Spectra	Raw	LDA	79.85
Soybean Spectra	Transformed	PCA-LDA	32.42

[a]Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester; LDA, linear discriminant analysis; PCA, principal component analysis

[b]Raw spectral data were not normalized, derived or smoothed; Transformed spectral data were normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and Savitzky-Golay smoothed

Table 4. Confusion matrix from the classification model generated by LDA joint with PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) from cotton tissue treated with 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a]

	Actual formulation	ACID	DMA	CHOLINE	ESTER	
Predicted formulation		----------%----------				
ACID		**25**	16	18	9	
DMA		19	**18**	12	10	
CHOLINE		38	33	**44**	36	
ESTER		18	33	26	**45**	
Accuracy (%)		*25*	*18*	*44*	*45*	*33.38*[†]

[a]Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester; LDA, linear discriminant analysis; PCA, principal component analysis

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 5. Confusion matrix from the classification model generated by LDA joint with PCA of transformed fingerprint (1800 to 800 cm^{-1}) from cotton tissue treated with 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a,b]

	Actual formulation	ACID	DMA	CHOLINE	ESTER	
Predicted formulation		----------%----------				
ACID		**31**	19	16	13	
DMA		23	**29**	18	12	
CHOLINE		36	32	**45**	43	
ESTER		10	20	21	**32**	
Accuracy (%)		*31*	*29*	*45*	*32*	*34.8*[†]

[a]Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester; LDA, linear discriminant analysis; PCA, principal component analysis

[b]Transformed spectral data were normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and Savitzky-Golay smoothed

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 6. Confusion matrix from the classification model generated by LDA without PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) of cotton tissue treated with 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a]

	Actual formulation	ACID	DMA	CHOLINE	ESTER	
Predicted formulation		----------%----------				
ACID		**73**	13	7	9	
DMA		3	**71**	2	3	
CHOLINE		19	14	**89**	14	
ESTER		5	2	2	**74**	
Accuracy (%)		*73*	*71*	*89*	*74*	*77.16*[†]

[a]Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester; LDA, linear discriminant analysis; PCA, principal component analysis

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 7. Confusion matrix from the classification model generated by LDA joint with PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) from soybean tissue treated with 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a]

	Actual formulation	ACID	DMA	CHOLINE	ESTER	
Predicted formulation		----------%----------				
ACID		**45**	25	25	36	
DMA		16	**35**	18	34	
CHOLINE		22	28	**44**	14	
ESTER		17	12	12	**16**	
Accuracy (%)		*45*	*35*	*44*	*16*	*36.26*[†]

[a]Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester; LDA, linear discriminant analysis; PCA, principal component analysis

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 8. Confusion matrix from the classification model generated by LDA joint with PCA of transformed fingerprint (1800 to 800 cm^{-1}) from soybean tissue treated with 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a,b]

	Actual formulation	ACID	DMA	CHOLINE	ESTER	
Predicted formulation		----------%----------				
ACID		**40**	18	30	27	
DMA		26	**36**	28	28	
CHOLINE		16	24	**35**	30	
ESTER		18	22	7	**15**	
Accuracy (%)		*40*	*36*	*35*	*15*	*32.42*[†]

[a]Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester; LDA, linear discriminant analysis; PCA, principal component analysis

[b]Transformed spectral data were normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and Savitzky-Golay smoothed

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 9. Confusion matrix from the classification model generated by LDA without PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) of soybean tissue treated with 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a]

	Actual formulation	ACID	DMA	CHOLINE	ESTER	
Predicted formulation		----------%----------				
ACID		**78**	7	13	14	
DMA		6	**81**	2	14	
CHOLINE		6	7	**80**	6	
ESTER		10	5	5	**66**	
Accuracy (%)		*78*	*81*	*80*	*66*	*79.85*[†]

[a]Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester; LDA, linear discriminant analysis; PCA, principal component analysis

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

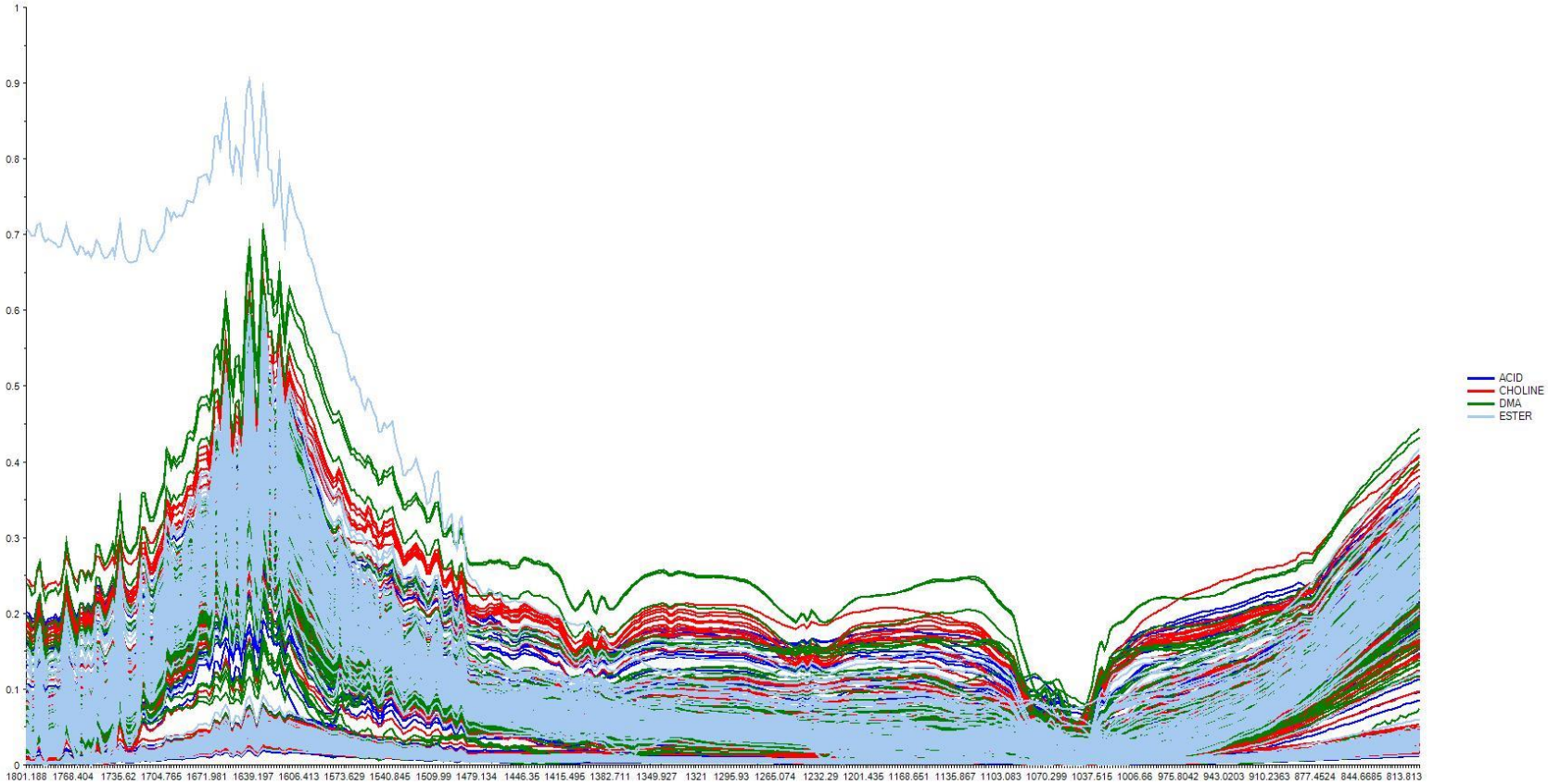

Figure 5. Raw fingerprint spectra (1800 to 800 cm^{-1}) from cotton tissue treated with 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a]
[a]*Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester*

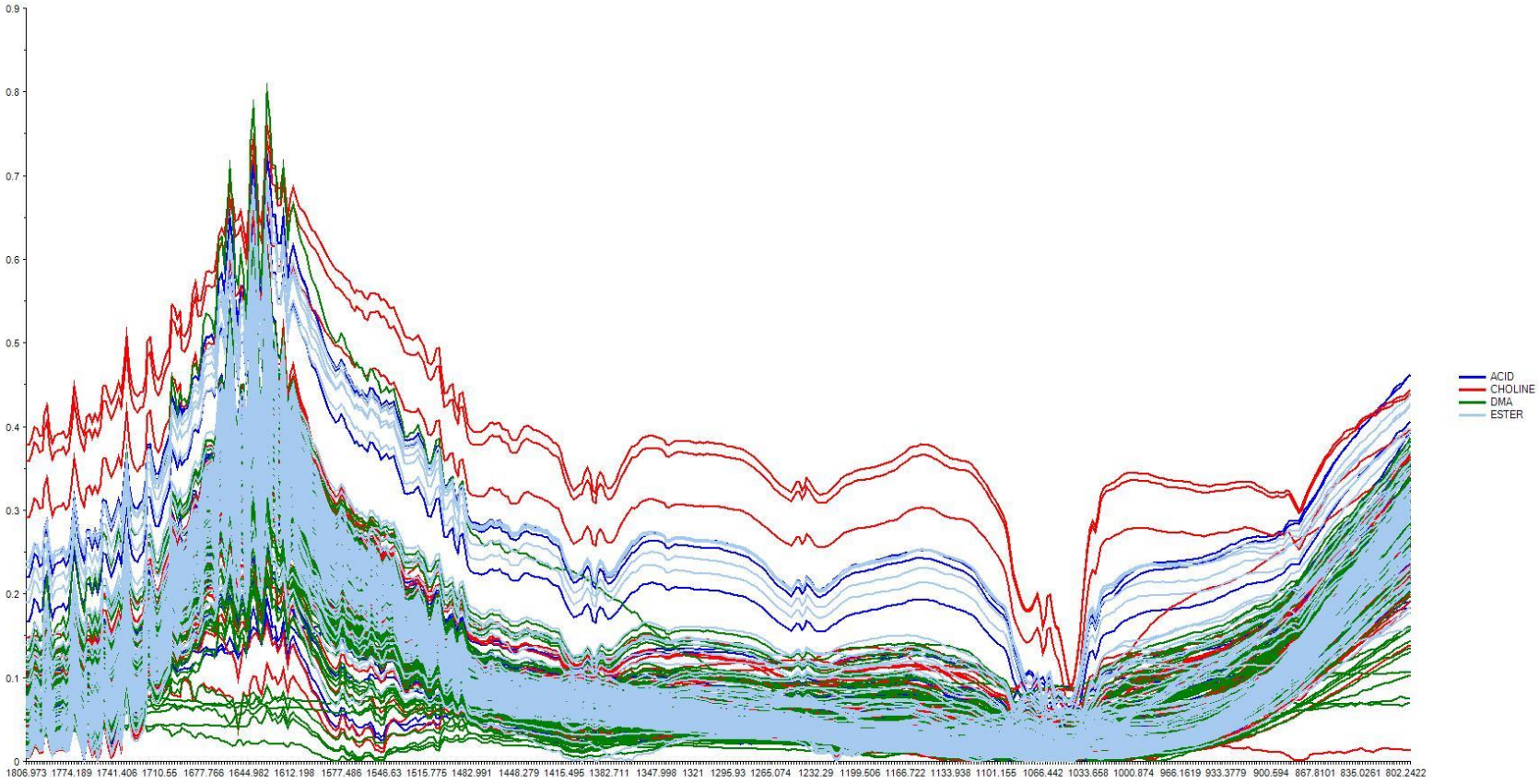

Figure 6. Raw fingerprint spectra (1800 to 800 cm^{-1}) from soybean tissue treated with 2,4-D acid, 2,4-D DMA, 2,4-D CHOLINE, or 2,4-D ESTER pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and 2,4-D concentration (33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1}).[a]
[a]*Abbreviations: DMA, dimethylamine salt; CHOLINE, choline salt; ESTER, isooctyl ester*

Application of FTIR spectroscopy and chemometrics for the classification of dicamba formulations in damaged cotton and soybean tissue

Research was conducted in 2017, 2018, and 2019 in Starkville, MS to develop a chemometrics and spectroscopy method to create a classification model capable of identifying specific dicamba formulations present in damaged crop tissue. Dicamba diglycolamine (DGA), dimethylamine (DMA), N,N-Bis-(3-Aminopropyl) methylamine (BAPMA), and diglycolamine with potassium acetate (DGAKAC) were applied to susceptible cotton and soybeans at 35, 17.5, 8.75, 4.375, 2.1875, and 1.09375 g dicamba ae ha^{-1}, and samples were analyzed with infrared spectroscopy, which were further analyzed using principal component analysis (PCA) and linear discriminant analysis (LDA). Joint PCA-LDA models were only capable of classifying dicamba formulation with 39.82% accuracy, whereas LDA alone was 80 to 85% accurate. Models performed worst when classifying dicamba DMA (27% to 80% accuracy), and best when classifying dicamba DGA/DGAKAC (40 to 85% accuracy). Correct classification of dicamba DGA in the presence of dicamba DGAKAC (and vice-versa) was reduced relative to other formulations, likely due to similarity of the molecular structure of DGA and DGAKAC. This research suggests that with further refining, chemometric analysis of spectral data from damaged crop tissue may be an economical, efficient, and promising application to support management of crop injury cases following OTM of dicamba.

Results and Discussion

Raw and Transformed Data Matrices. Automatic baseline- and ATR- corrected spectra (raw spectra) from cotton and soybean samples treated with the various dicamba formulations are shown pooled over dicamba concentration and sampling timing. The only significant peaks occurred at 3800 to 3000 cm^{-1} and 1800 to 800 cm^{-1}, however, the broad peak at 3800 to 3000 cm^{-1} is due to the O-H bend in water found in plant tissue. As such, only the spectral region commonly referred to as the 'fingerprint region' between 1800 to 800 cm^{-1} was included in spectral analysis. Raw cotton and soybean spectra narrowed to the fingerprint region between 1800 to 800 cm^{-1} from tissue treated with the various dicamba formulations were pooled over dicamba concentration and sampling timing where an increased resolution of spectral features have become observable by narrowing the spectral focus. Normalized, derived, and smoothed cotton and soybean fingerprint spectra from tissue treated with the various dicamba formulations are shown pooled over dicamba concentration and sampling timing, and reflect amplification of differences in spectral features between samples. In a preliminary analysis of similar data, Reid (2017) used PCA loading plots to determine the most important spectral features in the soybean analyses are between 1687 and 1560 cm^{-1} and most likely represent the aromatic ring of dicamba and the primary or secondary amine from the various salts formulated with it. In the cotton analyses, PCA loading plot examination suggests that peaks between 1633 and 1556 cm^{-1} and 1395 to 1350 cm^{-1} are important for sample differentiation. The peaks between 1633 and 1556 cm^{-1} are most likely from the aromatic ring of the 2,4-D molecule and the primary or secondary amines from its formulated salts, and the features between 1395 and 1350 cm^{-1} are typical of a carboxylic acid group, which is present in the majority of the 2,4-D formulations (Reid 2017). These peaks provided the basis for determining a spectral range for use in subsequent PCA and LDA analyses.

PCA, LDA, and Joint PCA-LDA on Raw Data

Cotton. PCA performed on the raw spectral data pooled across concentrations and evaluation timings resulted in the first two PC (Principal Components) accounting for 93% of the explained variance, and 99% total explained variance contained in the first 5 PC. Minor sandwiching/clustering of samples by dicamba formulation can be observed in a 3D PCA graph of the first three PC. LDA of the raw spectral data pooled across concentrations and evaluation timings and using the eigenvectors generated by

dimensional reduction via PCA resulted in a classification model with 39.12% accuracy (Table 3); this discrimination plot is shown in Figure 10, where there is evident linear clustering of samples by formulation. However, this level and pattern of clustering is significantly less structured than those reported in previous research, although said research utilized different sample media (Deng et al. 2016; Lee et al. 2009). Construction of a classification model allows the generation of a confusion matrix displaying the model's prediction of dicamba formulation in a spectrum from a given sample of crop tissue damaged by dicamba plotted against the actual value. The confusion matrix from the classification model generated by joint LDA-PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) from cotton tissue is shown in Table 4. The classification model performed best identifying dicamba DGAKAC (47% accuracy), and worst identifying dicamba DMA (33% accuracy). LDA conducted alone (without PCA) on the raw cotton spectral data resulted in a classification model with 84.78% accuracy (Table 3). The discrimination plot of this classification model, where there is noticeable linear clustering of each formulation. The level of accuracy with this model is more in-line with previous research, although the clustering pattern remains irregular (Deng et al. 2016; Lee et al. 2009). The confusion matrix results of the classification model prediction following LDA alone is shown in Table 6. This classification model was most accurate when identifying dicamba BAPMA (90%) and least accurate when identifying dicamba DMA (80%) although it was no less accurate than 80% for any given formulation (Table 6).

Soybeans. PCA performed on the raw soybean spectral data pooled across concentrations and evaluation timings resulted in the first PC accounting for 95% of the explained variance, and 100% total explained variance contained in the first 3 PC. Despite a high proportion of variance explained by PC1, the 3D score plot from the PCA on raw soybean spectral data reflects poor clustering of all formulations except dicamba BAPMA. Joint PCA-LDA of raw soybean spectral data resulted in a classification model with 35.15% accuracy (Table 3). A discrimination plot from this model is shown in Figure 15, where there is some linear clustering by formulation visible, but overall clustering appears poor, again in contrast to previous work on other sample media (Deng et al. 2016; Lee et al. 2009). The resulting confusion matrix from this model is shown in Table 7. This classification model performed best (47% accuracy) when classifying tissue containing dicamba DGA, and poorest (27% accuracy) when classifying tissue containing dicamba DMA (Table 7). When LDA was conducted alone on the raw soybean spectral data, a classification model with 79.64% accuracy was created (Table 3). A discrimination plot from this model there is distinct linear clustering of each formulation, reflecting the model's improved classification accuracy. In this case, discrimination plot accuracy was more similar to previous work such as Ami et al. (2010), which used similar methods to classify embryonic stem cell differentiation. However, the clustering pattern was linear as opposed to the bunched patterns reported by Lehmann et al. (2015). The confusion matrix generated by the LDA alone on soybean raw spectral data is shown in Table 9. This model was capable of up to 84% accuracy when classifying dicamba DGA, and only 76% accuracy when classifying dicamba DMA.

PCA and Joint PCA-LDA on Transformed Data

Cotton. PCA on cotton spectra normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and smoothed with Savitzky-Golay smoothing resulted in a 3D PC score plot shown. Poor clustering is present in this score plot, with no samples noticeably clustered by formulation. This trend is reflected in the somewhat reduced amount of total explained variation occurring in the first three PC of the PCA (65, 13, and 6%, respectively). Similarly, a joint PCA-LDA conducted on transformed cotton spectra resulted in a classification model with only 39.82% accuracy (Table 3). The poor accuracy of this model is reflected by the noticeably poor clustering of samples by formulation shown in the LDA

discrimination plot in Figure 12, in stark contrast to the high degree of clustering shown in previous research on other sample media (Ami et al. 2010; Deng et al. 2016; Lee et al. 2009; Lehmann et al. 2015). The poor accuracy of this classification model is further depicted in its corresponding prediction confusion matrix shown in Table 5. This model was only able to achieve 49% accuracy at best (when classifying samples treated with dicamba DGA) and 33% accuracy at worst (when classifying samples treated with dicamba DGAKAC, Table 5).

Soybeans. The 3D PC scores plot of transformed soybean spectra analyzed via PCA. There is some minor clustering of samples treated with dicamba BAPMA or dicamba DMA visible, but overall clustering remains poor. Only 54% of the total explained variance in this classification is contained in PC1, with an additional 21% in PC2, and 7% in PC3, indicating poorly-clustered, highly-variable data (Table 2). Joint PCA-LDA of the transformed soybean spectra resulted in a classification model with 34.59% accuracy (Table 3). The discrimination plot for this model there appears to be some linear clustering of samples by formulation, but most of which is conflated by overlapping formulation clusters. The formulation classification confusion matrix generated from this joint PCA-LDA reflects the poor clustering and accuracy of the model (Table 8). Accuracy of this model ranged from 32 to 38%, was best when classifying dicamba DMA, and worst when classifying dicamba BAPMA (Table 8). These results from transformed soybean spectra are largely similar to those of cotton in that the clustering patterns, model accuracy, and levels of explained variation are poor and dissimilar to those reported in similar research on other sample media (Ami et al. 2010; Deng et al. 2016; Lee et al. 2009; Lehmann et al. 2015).

Table 3. Classification model parameters following LDA alone or joint with PCA of fingerprint spectra (1800 to 800 cm^{-1}) from cotton or soybean tissue treated with dicamba DGA, dicamba DMA, dicamba DGAKAC or dicamba BAPMA, pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and dicamba concentration (35, 17.5, 8.75, 4.375, 2.1875, 1.09375 g dicamba ae ha^{-1}).[a]

Data Matrix	Data Type[b]	Model Source	Accuracy
			%
Cotton Spectra	Raw	PCA-LDA	39.12
Cotton Spectra	Raw	LDA	84.78
Cotton Spectra	Transformed	PCA-LDA	39.82
Soybean Spectra	Raw	PCA-LDA	35.15
Soybean Spectra	Raw	LDA	79.64
Soybean Spectra	Transformed	PCA-LDA	34.59

[a]Abbreviations: BAPMA, N,N-Bis-(3-Aminopropyl) methylamine salt; DGA, diglycolamine salt; DGAKAC, diglycolamine salt with potassium acetate; DMA, dimethylamine salt; LDA, linear discriminant analysis; PCA, principal component analysis

[b]Raw spectral data were not normalized, derived or smoothed; Transformed spectral data were normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and Savitzky-Golay smoothed

Table 4. Confusion matrix from the classification model generated by LDA joint with PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) from cotton tissue treated with dicamba DGA, dicamba DMA, dicamba DGAKAC or dicamba BAPMA, pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and dicamba concentration (35, 17.5, 8.75, 4.375, 2.1875, 1.09375 g dicamba ae ha^{-1}).[a]

	Actual formulation	BAPMA	DGA	DGAKAC	DMA	
Predicted formulation		----------%----------				
BAPMA		**36** 32	18	19	16	
DGA		27	**40**	26	31	
DGAKAC		22	26	**48**	20	
DMA		15	16	7	**33**	
Accuracy (%)		*36*	*40*	*48*	*33*	*39.12*[†]

[a]Abbreviations: BAPMA, N,N-Bis-(3-Aminopropyl) methylamine salt; DGA, diglycolamine salt; DGAKAC, diglycolamine salt with potassium acetate; DMA, dimethylamine salt; LDA, linear discriminant analysis; PCA, principal component analysis
[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 5. Confusion matrix from the classification model generated by LDA joint with PCA of transformed fingerprint (1800 to 800 cm^{-1}) from cotton tissue treated with dicamba DGA, dicamba DMA, dicamba DGAKAC or dicamba BAPMA, pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and dicamba concentration (35, 17.5, 8.75, 4.375, 2.1875, 1.09375 g dicamba ae ha^{-1}).[a,b]

	Actual formulation	BAPMA	DGA	DGAKAC	DMA	
Predicted formulation		----------%----------				
BAPMA		**42**	15	29	21	
DGA		25	**49**	28	32	
DGAKAC		23	24	**33**	11	
DMA		10	13	10	**35**	
Accuracy (%)		*42*	*49*	*33*	*35*	*39.82*[†]

[a]Abbreviations: BAPMA, N,N-Bis-(3-Aminopropyl) methylamine salt; DGA, diglycolamine salt; DGAKAC, diglycolamine salt with potassium acetate; DMA, dimethylamine salt; LDA, linear discriminant analysis; PCA, principal component analysis

[b]Transformed spectral data were normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and Savitzky-Golay smoothed

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 6. Confusion matrix from the classification model generated by LDA without PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) of cotton tissue treated with dicamba DGA, dicamba DMA, dicamba DGAKAC or dicamba BAPMA, pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and dicamba concentration (35, 17.5, 8.75, 4.375, 2.1875, 1.09375 g dicamba ae ha^{-1}).[a]

	Actual formulation	BAPMA	DGA	DGAKAC	DMA	
Predicted formulation		----------%----------				
BAPMA		**90**	4	5	4	
DGA		1	**84**	5	11	
DGAKAC		6	4	**85**	5	
DMA		3	8	5	**80**	
Accuracy (%)		*90*	*84*	*85*	*80*	*84.78*[†]

[a]Abbreviations: BAPMA, N,N-Bis-(3-Aminopropyl) methylamine salt; DGA, diglycolamine salt; DGAKAC, diglycolamine salt with potassium acetate; DMA, dimethylamine salt; LDA, linear discriminant analysis; PCA, principal component analysis

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 7. Confusion matrix from the classification model generated by LDA joint with PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) from soybean tissue treated with dicamba DGA, dicamba DMA, dicamba DGAKAC or dicamba BAPMA, pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and dicamba concentration (35, 17.5, 8.75, 4.375, 2.1875, 1.09375 g dicamba ae ha^{-1}).[a]

	Actual formulation	BAPMA	DGA	DGAKAC	DMA	
Predicted formulation		----------%----------				
BAPMA		**32**	20	13	15	
DGA		40	**47**	40	38	
DGAKAC		15	18	**33**	20	
DMA		13	15	14	**27**	
Accuracy (%)		*32*	*47*	*33*	*27*	*35.15*[†]

[a]Abbreviations: BAPMA, N,N-Bis-(3-Aminopropyl) methylamine salt; DGA, diglycolamine salt; DGAKAC, diglycolamine salt with potassium acetate; DMA, dimethylamine salt; LDA, linear discriminant analysis; PCA, principal component analysis

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 8. Confusion matrix from the classification model generated by LDA joint with PCA of transformed fingerprint (1800 to 800 cm^{-1}) from soybean tissue treated with dicamba DGA, dicamba DMA, dicamba DGAKAC or dicamba BAPMA, pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and dicamba concentration (35, 17.5, 8.75, 4.375, 2.1875, 1.09375 g dicamba ae ha^{-1}).[a,b]

	Actual formulation	BAPMA	DGA	DGAKAC	DMA	
Predicted formulation		----------%----------				
BAPMA		**32**	29	24	22	
DGA		25	**33**	23	27	
DGAKAC		24	19	**36**	13	
DMA		19	19	17	**38**	
Accuracy (%)		*32*	*33*	*36*	*38*	*34.19*[†]

[a]Abbreviations: BAPMA, N,N-Bis-(3-Aminopropyl) methylamine salt; DGA, diglycolamine salt; DGAKAC, diglycolamine salt with potassium acetate; DMA, dimethylamine salt; LDA, linear discriminant analysis; PCA, principal component analysis

[b]Transformed spectral data were normalized to the area under the curve, derived to the first Savitzky-Golay derivative, and Savitzky-Golay smoothed

[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Table 9. Confusion matrix from the classification model generated by LDA without PCA of raw fingerprint spectra (1800 to 800 cm^{-1}) of soybean tissue treated with dicamba DGA, dicamba DMA, dicamba DGAKAC or dicamba BAPMA, pooled over evaluation timing (7, 14, 21, 28, and 56 d after treatment) and dicamba concentration (35, 17.5, 8.75, 4.375, 2.1875, 1.09375 g dicamba ae ha^{-1}).[a]

	Actual formulation	BAPMA	DGA	DGAKAC	DMA	
Predicted formulation		----------%----------				
BAPMA		**79**	1	3	4	
DGA		11	**84**	11	17	
DGAKAC		3	7	**80**	3	
DMA		7	8	6	**76**	
Accuracy (%)		*79*	*84*	*80*	*76*	*79.64*[†]

[a]Abbreviations: BAPMA, N,N-Bis-(3-Aminopropyl) methylamine salt; DGA, diglycolamine salt; DGAKAC, diglycolamine salt with potassium acetate; DMA, dimethylamine salt; LDA, linear discriminant analysis; PCA, principal component analysis
[†]Weighted average accuracy across all formulation classifications (overall classification model accuracy)

Impacts and Benefits to Mississippi Soybean Producers

This research shows that chemometric analyses of soybean and cotton tissue that have been damaged by various dicamba and 2,4 D formulations and concentrations and collected at a range of evaluation timings may be useful in constructing classification models that can be used to identify the specific formulations. This method for formulation identification provides regulators, industry, and producers with a tool to strengthen stewardship programs; providing effective weed management, improving farm productivity, and maintaining the environmental conservation. This technology could enhance the position of Mississippi as an agricultural leader by exhibiting agricultural responsibility.

End Products

Publications:

Guilherme Sousa Alves, Ashli Brown, Bradley K. Fritz, Daniel Reynolds, Darrin Dodds, Greg Kruger, Jeffrey A. Golus, Kasey Schroeder, Wesley Hoffmann (2020) Dicamba Off-Target Movement From Applications on Soybean at Two Growth Stages Under Different Environmental Conditions. Science of the Total Environment. In Review. Manuscript Number: STOTEN-D-15261.

Soltani, N.; Oliveria, M.O.; Guilherme, S.A.; Werle, R.; Kruger, G.; Norsworthy, J.K.; Spraque, C.L.; Young, B.G.; Reynolds, D.; **Brown, A.**; *Sikkema, P.H. (2020) Large-Scale Off-Target Movement Assessment of Dicamba in North America. *Weed Technology*. DOI:10.1017/wet.2020.17.

Abstracts and Presentations:

Jane Wang, Ben Blackburn, Ashli. Brown, Darrell Sparks, Dan Reynolds (2109) FT-IR Spectroscopy and Chemometrics: An Application for the Identification of Dicamba Formulations; BCH-EPP Student Research Symposium, Dorman Hall MSU. (November 8, 2019)

Ashley Meredith, **Ashli Brown**, Pam Oliver, Leah Ritter, (2019) Transfer of a trace level dicamba method between industry and a state agency to enable assessment of off-target transport. American Chemical Society, San Diego, CA. (August 25-29).

Ashli Brown, Ashley Meredith, Darrell Sparks, John Buol, Greg Kruger, Dan Reynolds (2019) Application of FTIR Spectroscopy and Chemometrics for the Classification of Auxin Herbicide in Cotton and Soybean. American Chemical Society, San Diego, CA. (August 25-29).

Ben Blackburn, John Buol, Dan Reynolds, Cedric Reid, Darrell Sparks, **Ashli Brown** (2019) FT-IT Spectroscopy Coupled to Chemometrics for the classification of auxin herbicides. MSU Undergraduate Research Symposium, Mississippi State University. (April 16, 2019).

John Buol, Dan Reynolds, **Ashli Brown Johnson**, Cedric Reid (2018) Southern Weed Science Society. Developing a Traditional Fourier-Transform Infrared Spectroscopy Method for Detection and High-Fidelity of Various Auxin Herbicide Formulations in Damaged Soybean and Cotton Tissue," SWSS, Atlanta, GA. (January 23, 2018).

Ashli Brown Johnson (2018) Southern Weed Science Society, "Analytical Method Development to Support Ag Research Programs," SWSS, Atlanta, GA. (January 21, 2018).

Magan Green, Angel Phillips, Julie Toler, Ashely Meredith, Zac Carpenter, Dan Reynolds, **Ashli Brown Johnson** (2018) Association of Offical Analytical Chemists (AOAC), Measuring the

Reduction of Dicamba Carryover in Pesticide Application Equipment by LC-MS/MS, Southern Section AOAC International, Atlanta, GA. (April 18, 2018).

Darrell Sparks, Carolyn Russell, Darren Nakamura, Scott Boone, Ashley Meredith, **Ashli Brown Johnson** (2018) 256th ACS National Meeting and Exposition. FT-IR Spectrophotometric Screening of Adulterants in Honey, American Chemical Society, Boston, MA. (August 23, 2018).

Ali Hanson, **Ashli Brown Johnson**, Darrell Sparks, Cedric Reid, Dan Reynolds (2017) Promoting Agricultural Stewardship Through Identification of Synthetic Auxins in Cotton. Spring 2017 Undergraduate Research Symposium, MSU Shackouls Honors College, Mississippi State, MS. (April 13, 2017).

Ashli Brown, Darrell Sparks, Cedric Reid (2017) Method Development for New 2,4 D and Dicamba Formulations. MAIC, Perdido Beach Resort. (July 25, 2017).

Cedric Reid, Gary Cundiff, Daniel Reynolds, Darrell Sparks, **Ashli Brown** (2016) Development of FTIR Method to Identify Herbicides and Their Low Volatile Counterparts. FPRW, St. Pete Beach, FL (July 18, 2016).

Ashli Brown, Darrell Sparks, Cedric Reid, Curtis Atkinson (2015) Using FTIR as a Tool. AOAC 129th Meeting, Los Angeles, CA (September 27-30).

Ashli Brown (2015) Pesticide Residue Analysis. Extension Professional Improvement Conference, MSU Extension, MS (August 12).

Students Mentored:
Graduate Students:
Benjamin Blackburn, MS Biochemistry (In Progress)
John Buol, Ph.D. Agronomy-Weed Science (Degree Conferred, 2019)
Cedric Reid, Ph.D. Biochemistry (Degree Conferred, 2017)
Undergraduate Students:
Ali Hanson
High School Students:
Jane Wang